A Socio-mathematical and Structure-Based Approach to Model Sentiment Dynamics in Event-Based Text

by

Areej Alhothali

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Suzanne Stevenson
Professor, Dept. of Computer Science,
University of Toronto

Supervisor(s): Jesse Hoey
Associate Professor, Cheriton School of Computer Science,
University of Waterloo

Internal Member: Pascal Poupart
Professor, Cheriton School of Computer Science,
University of Waterloo
Peter van Beek
Professor, Cheriton School of Computer Science,
University of Waterloo

Internal-External Member: Olga Vechtomova
Associate Professor, Dept. of Management Sciences,
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Natural language texts are often meant to express or impact the emotions of individuals. Recognizing the underlying emotions expressed in or triggered by textual content is essential if one is to arrive at an understanding of the full meaning that textual content conveys. Sentiment analysis (SA) researchers are becoming increasingly interested in investigating natural language processing techniques as well as emotion theory in order to detect, extract, and classify the sentiments that natural language text expresses. Most SA research is focused on the analysis of subjective documents from the writer's perspective and their classification into categorical labels or sentiment polarity, in which text is associated with a descriptive label or a point on a continuum between two polarities. Researchers often perform sentiment or polarity classification tasks using machine learning (ML) techniques, sentiment lexicons, or hybrid-based approaches. Most ML methods rely on count-based word representations that fail to take word order into account. Despite the successful use of these flat word representations in topic-modelling problems, SA problems require a deeper understanding of sentence structure, since the entire meaning of words can be reversed through negations or word modifiers. On the other hand, approaches based on semantic lexicons are limited by the relatively small number of words they contain, which do not begin to embody the extensive and growing vocabulary on the Internet.

The research presented in this thesis represents an effort to tackle the problem of sentiment analysis from a different viewpoint than those underlying current mainstream studies in this research area. A cross-disciplinary approach is proposed that incorporates affect control theory (ACT) into a structured model for determining the sentiment polarity of event-based articles from the perspectives of readers and interactants. A socio-mathematical theory, ACT provides valuable resources for handling interactions between words (event entities) and for predicting situational sentiments triggered by social events. ACT models human emotions arising from social event terms through the use of multidimensional representations that have been verified both empirically and theoretically. To model human emotions regarding textual content, the first step was to develop a fine-grained event extraction algorithm that extracts events and their entities from event-based textual information using semantic and syntactic parsing techniques. The results of the event extraction method were compared against a supervised learning approach on two human-coded corpora (a grammatically correct and a grammatically incorrect structured

corpus). For both corpora, the semantic-syntactic event extraction method yielded a higher degree of accuracy than the supervised learning approach. The three-dimensional ACT lexicon was also augmented in a semi-supervised fashion using graph-based label propagation built from semantic and neural network word embeddings. The word embeddings were obtained through the training of commonly used count-based and neural-network-based algorithms on a single corpus, and each method was evaluated with respect to the reconstruction of a sentiment lexicon. The results show that, relative to other word embeddings and state-of-the-art methods, combining both semantic and neural word embeddings yielded the highest correlation scores and lowest error rates.

Using the augmented lexicon and ACT mathematical equations, human emotions were modelled according to different levels of granularity (i.e., at the sentence and document levels). The initial stage involved the development of a proposed entity-based SA approach that models reader emotions triggered by event-based sentences. The emotions are modelled in a three-dimensional space based on reader sentiment toward different entities (e.g., subject and object) in the sentence. The new approach was evaluated using a human-annotated news-headline corpus; the results revealed the proposed method to be competitive with benchmark ML techniques. The second phase entailed the creation of a proposed ACT-based model for predicting the temporal progression of the emotions of the interactants and their optimal behaviour over a sequence of interactions. The model was evaluated using three different corpora: fairy tales, news articles, and a handcrafted corpus. The results produced by the proposed model demonstrate that, despite the challenging sentence structure, reasonable agreement was achieved between the estimated emotions and behaviours and the corresponding ground truth.

Acknowledgements

First and foremost, I would like to express my deepest appreciation for the never-ending encouragement, insightful direction, timely feedback, and inspiration so freely offered by my advisor, Jesse Hoey, throughout my Ph.D. studies. It was an enormous privilege and honour to work with him and to study under his guidance. His vision, enthusiasm, sincerity, humility, and wonderful sense of humour have profoundly inspired me. I would especially like to thank him for his understanding, sympathy, and support during hard times.

I wish to convey my sincere gratitude to my dissertation committee members, Peter van Beek, Pascal Poupart, Olga Vechtomova, and Suzanne Stevenson, for contributing their time and valuable comments during the dissertation process. I also very much appreciate the guidance and help I received from Tobias Schroder with respect to numerous aspects of my research.

A very special thank you goes out to all of my present and former colleagues in our research group at the University of Waterloo, in particular, Joshua Jung, Shehroz Khan, Deepak Rishi, Aarti Malhotra, and Dan Wang. Not only did they provide invaluable help and advice with many components of my work, but more importantly, they also played an essential role in the incomparably friendly, supportive atmosphere I found in the Computational Health Informatics Laboratory (CHIL) research group.

I am also grateful for the fast replies to all inquiries and the prompt service offered by current and former staff in the Computer Science Graduate Studies office: Margaret Towell, Paula Roser, and Neoma Power. Additional assistance and quick responses were always available from Computer Science Computing Facility (CSCF) staff members Gordon Boerke and Lori Paniak, both during and after work hours.

Last, but most definitely not least, I would like to say a huge thank you to my father, who always encouraged me to get the highest degree possible; to my stepmother, for being always there to listen, guide, and help; and to my siblings, for their prayers, wisdom, and support during my Ph.D. studies. My most heartfelt gratitude goes to my own small family: my husband Omar for his support and unconditional love, and my daughter Hala for her hugs, laughter, and the toothy grins that light up my days.

Dedication

To the soul of my mother *Mastorah*..

To my father *Matouq Alhothali*..

To my husband *Omar Altalhi*..

To my daughter *Hala* and my newly born son *Anas*..

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In many aspects of life, outside opinions affect our decision-making processes: we seek others' opinions when buying a new product, reading a book, supporting a political campaign, and even when choosing which movie to watch. Humans have a natural desire to express and to seek opinions on products, services, or individuals. Growing interest in this area of human nature over the last decade has led to a corresponding increase in research efforts related to sentiment analysis (SA), as well as intensified investigation of methods for automatically extracting emotional content from natural language texts. The growing popularity of social media and the expanding stream of user-generated content in the cyber world have fuelled this interest in sentiment analysis. The huge amount of opinionated digital content on the Internet and social media (Facebook, Twitter, blogs, and forums) has been particularly valuable for individuals and organizations interested in, for example, timely feedback about a new product, a political movement, or other matters.

The terms "sentiment" and "opinion" are used interchangeably in the literature related to sentiment analysis and natural language processing (NLP), which focus on people's emotional interpretation or feelings about situations or entities. Sentiment analysis, also known as opinion mining, appraisal extraction or attitude analysis, refers to the process of using Natural Language Processing (NLP), machine-learning (ML), and statistical techniques to identify, extract, or classify opinions or sentiments expressed in textual input [152]. SA techniques have been used and examined in a variety of application domains [68] such as public opinion analysis [80, 182, 268];

sentiment analysis in news articles [12, 11]; and customer feedback analysis [277, 152, 184, 89]. Investigations in the SA area are often divided into five main tasks: sentiment classification, detection of opinion spam, measurement of review usefulness, lexicon creation, and aspect extraction. The main task that has received significant attention from SA researchers is sentiment classification, the goal of which is to determine or classify the sentiments expressed in a text selection and to assess whether the text indicates positive or negative sentiment/opinions. Sentiment classification is challenging because it requires a deep understanding of natural language as well as domain and common-sense knowledge. For example, people often express their opinions or sentiments in a complex way: using rhetorical modes such as sarcasm, metaphors, and implication, or using complex styles that allude to other external or internal resources. Effective sentiment classification also necessitates an understanding of human emotions and emotional categories. Understanding and categorizing humans' emotions is another challenging task due to the division among scientists over emotions definitions and emotions categorization.

Sentiment classification can be further categorized into four main subtasks: determining the sentiment orientation of text that expresses an opinion as either positive or negative (opinion mining) [277, 152, 195]; determining the sentiment polarity of an opinionated text as a point in a continuum between two opposing sentiment polarities (polarity classification) [151, 241, 274]; determining the affective state of a writer (writer affect recognition) [306]; or determining the affective state evoked by a text (reader affect recognition) [306, 149, 22]. Despite the wide variety of applications in this field, SA research often is centred around opinion mining, focusing on highly subjective statements (e.g., a movie review) rather than objective data that might still carry sentiment [195]. Most of the methods proposed in the sentiment classification subarea are sensitive to context because they often rely on corpus-based statistics obtained from a domain-specific dataset. They are also generally focused on the use of frequency-based statistics (bag-of-words representations) that ignore word order and sentence structure. Only a handful of attempts have incorporated additional features as a means of accounting for negations, intensities, part-of-speech tags, and contextual valence shifters, and these methods have been found to outperform standard bag-of-words techniques [309, 296]. Several recently proposed compositional models for modelling human sentiments in natural language text include consideration of the dependency/interaction between words. These compositional models offer significant improvement with respect to accurately assigning positive or negative scores to varying levels of granularity (phrases and sentences), yet they are based on word frequencies that are often irrelevant to the sentiment, fail to capture the intensity of the sentiment, and are sensitive to context [183, 247, 246].

The research presented in this thesis tackled the problem of factual/event-based sentiment analysis in a multidisciplinary manner by incorporating sociological theory called affect control theory (ACT) into a compositional sentiment model. The premise of ACT, a socio-mathematical theory that models human sentiments in social events, is that individuals have pre-event, culturally shared feelings about behaviours, identities, or objects, and that social events generate post-event feelings that might differ from the pre-event feelings. The difference between the post-event and pre-event feelings might cause negative or positive emotions. For example, good actions typically cause good emotions and vice versa (e.g., an event such as "*A student helped an elderly lady*" will create positive sentiments for the interactants and any observers). The difference between post-event and pre-event feelings can also be used for estimating the unlikelihood of an event. For example, an event such as "*A teacher teaches a student*" is very likely and results in a very small difference because the behaviour performed is consistent with the identity of the actor, while an event such as "*A teacher abuses a student*" is unlikely since it represents a substantial difference between post-event and pre-event feelings. The action or most probable action that each interactant will take after the event occurs can also be estimated using the difference between post-event and pre-event feelings. For example, if a student *helps* an elderly lady, she will most likely *thank* him, and she will be less likely to *attack* him. ACT represents human sentiments, i.e., the pre-event culturally shared and post-event feelings, in three-dimensional space. ACT lexicons are composed of thousands of terms, i.e., behaviours and identities, that have been collected from numerous languages and cultures.

The use of ACT for SA problems is advantageous for several reasons. The fact that ACT represents human emotions in multidimensional space is thought to provide comprehensive and universal representations of human emotions [191, 102, 224, 33]. These multidimensional representations have a very long history in sociology and psychology and have been verified empirically over many years, and the ACT mathematical equations have also been developed based on extensive empirical studies of impression formation. ACT provides finer-grained emotion analysis because it models human emotions toward different entities involved in social events. ACT takes into account the interaction among social event entities in estimating post-event feelings.

This interaction is compatible with the linguistic principle underlying a compositional semantic model, according to which the meaning of a sentence is a function of the words that comprise it. For this reason, in the work conducted for this thesis, ACT was incorporated into a compositional SA model that analyzes the emotions triggered for readers and interactants by the social events expressed in the text through the incorporation of a three-dimensional sentiment lexicon and a set of ACT impression-formation equations. To pursue and developed the ACT structured-based sentiment analysis model, two main problems need to be addressed.

The first problem was to extract events and their associated components from textual content. Despite the wide interest in extracting events from natural language text, few researchers have performed fine-grained event extraction in order to extract predicates and their associated components from a given sentence. For this work, to extract events and their components (i.e., subject, object, modifiers, and location), three main approaches were explored. First, a dependency-based approach was developed that, for any given dependency parser output, parses a sentence and identifies the event components and their modifiers. Dependency parsers provide a uniform representation of the grammatical relationships among the words in a given sentence [39]. For example, in a sentence such as "*John helped Mary*", the dependency parser will show that *John* is the nominal subject of *help*. Using these uniform representations, the algorithm simply locates the event components. The second approach involved the implementation and extension of a syntactic-based algorithm [225] that performs a breadth and depth search in a syntactic parse tree to extract event triplets (i.e., subject, verb, and object). The triplet algorithm [225] was extended to deal with sentences structured ungrammatically and to extract modifiers (i.e., adjectives or adverbs that modify the subject, object, or verb) and the location of the events. Third, a semantic- and syntactic-based event extraction method was proposed that first performs semantic parsing in order to locate the predicate in the sentence, then executes a breadth and depth search in the syntactic tree to locate the other components of the event. The results from these three approaches were compared against two manually annotated corpora for different syntactic structures: news headlines and news articles. The results were also compared against a syntactic-based event extraction algorithm and an ML-based algorithm. The findings show that the semantic- and syntactic-based approach outperformed the dependency-based, syntactic-based, and ML-based approaches for both grammatically and ungrammatically structured sentences.

The second problem was that ACT utilizes a limited number of human-coded lexicons that contain a relatively small inventory of words. The two most recent English ACT lexicons were collected in 2002 and 2003 [71, 161]. Due to their age, these sentiment ratings may not be up to date and may not reflect the sentiment/attitude of current generations toward behaviours and identities. Overcoming these limitations entailed the development of a semi-supervised graph-based label propagation algorithm designed to extend the ACT sentiment lexicon by building on two modalities: a semantic and a distributed word representation. The graph-based label propagation algorithm propagates the label/scores from labelled nodes to the remainder of the graph based on the geometry of the data (i.e., the similarity between the nodes of the graph). Extensive evaluation was conducted of several unsupervised and supervised methods for inducing the ACT sentiment lexicon. The semi-supervised label propagation algorithm was also assessed in relation to variants of word representations, including semantic, distributed, and count-based word representations/embeddings. The results demonstrate that linearly combining the similarities of the semantic and neural word embedding features in the semi-supervised algorithm generated a multidimensional sentiment lexicon that exhibited higher correlation with the actual sentiment ratings than other word representations and state-of-the-art methods. To the best of the author's knowledge, the work described in this thesis represents the first study that has examined multi-dimensional lexicon expansion and the first that combines both semantic and distributed features in a label propagation algorithm.

The augmented lexicon, ACT impression-formation equations, and the event extraction algorithm were utilized for the development of a structured-based SA approach that models human emotions in event-based text. The model parses event-based sentences or documents, extracts event components and their modifiers, and computes pre-event and post-event feelings. Using the extracted components and based on ACT principles, human sentiments were modelled in two different contexts.

The first proposed model is an entity-based reader affect recognition model that analyzes the sentiment evoked in readers by single events (i.e., news headlines). This model analyzes the sentiment of the readers with respect to both subject and object and then represents these sentiments in a three-dimensional space. These three-dimensional scores can then be mapped to an emotion label (e.g., *"amused"*,*"disappointed"*, and*"delighted"*, etc.). For example, readers of a sentence such as *"John killed Mary"* will experience negative emotions from reading about such an event, yet they will experience different emotions toward the subject and the object.

They might, for example, be *"angry at"* the subject and *"sorry for"* the object. This model was evaluated on a manually annotated dataset of news headlines collected from a variety of news resources and annotated with the emotions they evoke. The results of the evaluation show that the proposed model outperformed the standard supervised learning approach based on bag-of-words models.

The second modelling was of the interactants' emotions and their optimal behaviour with respect to a sequence of events (e.g., fairy tales and news articles). With consideration of the sentiment generated during a previous event, coupled with the pre-event culturally shared sentiment associated with the current event, the emotions that each interactant in the story would feel after interacting in multiple events were computed. The next optimal behaviour that each participant in the event would exhibit after several interactions was also estimated based on computation of the behaviour that minimizes the difference between post-event and pre-event feelings. As with ACT [99], by computing the partial derivative of the difference between the post-event and pre-events feelings and solving for the behaviour. In a manner similar to that for determining optimal behaviour, the optimal identities or profiles that characterize each participant during social events were also estimated. This model was evaluated using two real-world human-annotated corpora: a fairy tale dataset with human ratings of interactant sentiment [4] and news articles annotated according to writer sentiments with respect to the interactants [254]. The model was also assessed relative to a handcrafted story that was created for evaluating how the model would treat simple events that do not require any highly advanced parsing techniques. The results of the evaluation of the proposed model against the user manual annotation showed reasonable agreement as well as interesting patterns of emotional and behavioural responses that were deemed worthy of further investigation.

The model proposed in this thesis addresses the problem of sentiment analysis from different and new perspectives. Unlike most of the methods proposed by researchers in the SA field, which model human sentiment in one-dimensional space, the model proposed in this thesis represents human emotions in a three-dimensional space. This three-dimensional representation was validated both theoretically and empirically to ensure that it accurately represents human emotions [191, 70]. The proposed model developed through this research analyzes the sentiment triggered by factual/objective natural language, an area that has received less attention from SA researchers. The new model is also focused on the sentiment of the readers of and interactants in an event, not on the sentiment of the writers or on the opinion mining problems that are typically addressed in SA research. Further innovations are that the model analyzes the sentiments for different levels of granularity (sentence level and document level) and that it performs an entity-based sentiment analysis by assigning different emotions to different entities involved in the events. It also models the sentiment triggered by a sequence of events by taking into account the previously generated sentiment. The following section provides additional details about the contributions of this research.

1.1 Summary of Research Contributions

The contributions of the research presented in this thesis can be divided into four main categories:

– The new model enables fine-grained event extraction that uses semantic and syntactic features to extract the predicate of a given sentence and its associated components (i.e., participants, modifiers, and location of an event) and provides an evaluation of other syntactic-based, dependency-based, and ML-based methods of extracting events from grammatically and ungrammatically structured sentences.

– Three-dimensional sentiment lexicons have been regenerated and extended in a semi-supervised manner using a label propagation algorithm built on semantic and neural word embedding models, enabling extensive assessment of the label propagation algorithm using a variant of semantic, distributed, and count-based similarity measures, and the results have been compared against unsupervised, supervised, and state-of-the-art methods.

– Readers' emotional responses to social events and their associated participants (actors and objects) in event-based textual information have been modelled, the proposed method has been evaluated against newly collected and annotated news headlines, and the similarities between the true sentiments of the readers and the estimated emotions have been demonstrated.

– The emotions of interactants have been modelled, along with their optimal behaviour and optimal profile in event-based articles; a recursive linear combination of multiple identities and behaviours has been performed; and the performance of the model has been assessed with respect to a handcrafted story, a fairy tale dataset, and a news article corpus.

1.2 Thesis Outline

The structure of the thesis is as follows:

- Chapter 2 provides background regarding emotion theory, ACT, statistical language-modelling methods, and a review of the literature related to sentiment analysis and sentiment lexicon expansion.

- Chapter 3 explains semantic, syntactic, and dependency parsing techniques, reviews some of the state-of-art methods developed for addressing the event extraction research problem, introduces the proposed semantic and syntactic approach developed through the research presented in this thesis, and provides the results of an evaluation based on two human-coded datasets.

- Chapter 4 describes the semi-supervised label propagation algorithm and the multigraph label propagation approach and presents the multimodality lexicon augmentation technique as well as an extensive evaluation of the lexicon induction methods using different word features.

- Chapter 5 includes a summary of ACT and emotion formation equations, presents the proposed model for single-event entity-based SA, and details both the dataset used for evaluating the proposed method and the results obtained.

- Chapter 6 outlines emotion dynamics and optimal behaviour estimation in ACT, presents an approach for modelling the emotions of the interactants and their optimal behaviour in event-based textual contents, and outlines the evaluation of the proposed approach based on three datasets.

- Chapter 7 offers conclusions about the contributions of the research and suggests possible avenues for future investigation.

Chapter 2

Background

This thesis proposes a multidisciplinary approach that incorporates affect control theory (ACT) into a structure-based sentiment analysis (SA) model. A three-dimensional sentiment lexicon has been augmented in a semi-supervised fashion using semantic and neutral word vector representation, with social events and their entities being extracted from event-based textual input. The following subsections thus provide background about 1) emotion theory and the theories most commonly used in sentiment analysis and other related research areas, 2) the main principles and mathematical background underlying ACT, 3) the statistical language modelling methods and vector space models used in lexicon augmentation methods, and 4) published work related to sentiment classification and lexicon creation.

2.1 Emotions Theory and Sentiment Analysis

Human emotions have been the subject of extensive study in disciplines such as psychology, anthropology, sociology, and, more recently, computer science. Researchers in fields that include affective computing (AC) [201]; human-computer interaction (HCI) [31, 113]; and sentiment analysis (SA) [68, 210] have also recently begun contributing to these knowledge areas by computationally modelling and evaluating existing theories. The variations in methods of conceptualizing emotions has led computer scientists to develop a number of computational models of emotions that can be grouped into two main types: categorical approaches that are based on mostly cognitive appraisal theories, and dimensional approaches based on dimensional theories [235]. Cognitive appraisal theory, the dominant theory in psychology as well as in AC and SA, posits that emotions arise from individual judgements about situations or events. Most SA and AC computational appraisal models represent emotions using discrete labels known as the basic emotions [33, 89]. The discrete approach most widely adopted is the model developed by Ekman et al., which classifies emotions according to facial expressions into six semantically distinct categories: sadness, fear, anger, happiness, surprise, and disgust [60]. In spite of their popularity, the basic emotions fail to represent the complexity of human feelings since evidence has shown that the basic emotions rarely occur alone and cannot be generalized across different cultures or languages [222].

Dimensional models, on the other hand, typically represent emotions in two- or three-dimensional space: arousal, valence, and sometimes dominance/control [221]. Five dimensional models are commonly known, which are those by Wundt [299], Schlosberg [237], Osgood [190], Davitz [55], and Sokolov et al. [249]. Each uses different dimensions to describe emotions. Pleasure-displeasure (valence/evaluation) and arousal-calm (arousal/activity) are components of these models, while the control/dominance dimension is included only in the models devised by Osgood and Davitz. Plutchik offered a new representation of emotions called the "wheel of emotions", which contains eight primary bipolar emotions: joy versus sadness, anger versus fear, trust versus disgust, and surprise versus anticipation. This three-dimensional model combines both basic and complex categories with dimensional theories in which emotions are arranged in concentric circles, with the complexity of the emotions increasing from the inner to the outer circles [202]. Although psychological evidence shows that people perceive words or any other events emotionally in more than one dimension [70, 15, 221], multidimensional models have been used less frequently in SA and AC research. For the research presented in this thesis, a multidimensional approach was adopted for modelling human emotions in natural language text.

A related point for consideration is that most multidimensional theories are linked to the "core affect" concept [221], the counterpoint to the appraisal theory, which describes human emotions in their simplest and most primitive state. Core affect is defined as a neurophysiological (mental, non-cognitive, and non-reflective) state that is consciously accessible as a simple raw feeling evident in an individual's moods/emotions. This raw feeling changes subconsciously

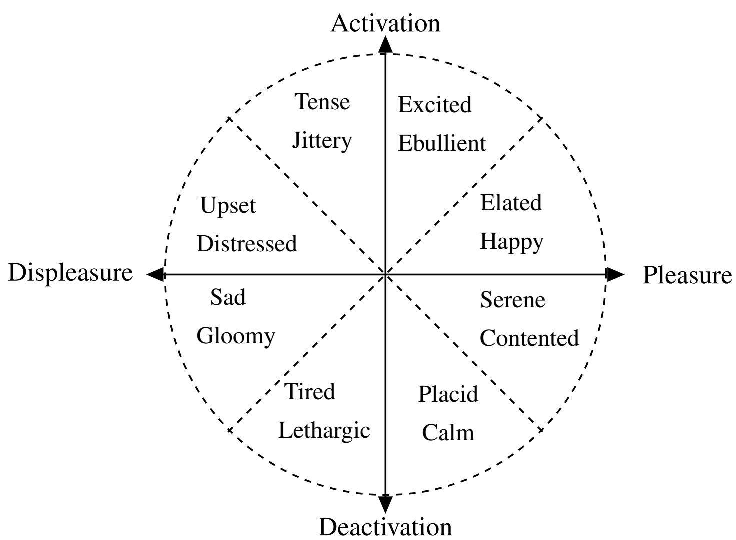


Figure 2.1: Affective (valence-arousal) circumplex

based on events and without any degree of personal control or socially learned structure. These feelings can be represented as a point in two-dimensional space: valence (pleasure-displeasure) and arousal (active-quiet), as shown in Figure 2.1. ACT is a cognitive appraisal theory based on the premise that people appraise words using culturally shared meanings and that these meanings can be represented in three-dimensional space.

2.2 Affect Control Theory

Affect Control Theory (ACT) is a new version of symbolic interactionism developed by Mead, who stated that individuals process gestures, words, or behaviour as symbols or meanings shared among people within the same culture [165]. These shared meanings help individuals to anticipate and understand how others will act in a particular situation. ACT was built upon Osgood et al.'s [191] semantic differential ratings (*affective meaning*) and empirical studies of impression formation [99, 102, 215]. The *affective meaning* was introduced by Osgood and his colleagues [191]. In a large set of cross-cultural studies in the 1950s, Osgood showed that concepts

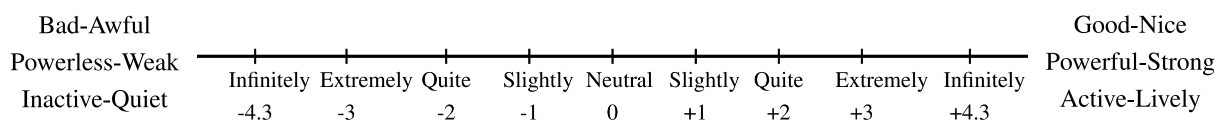


Figure 2.2: Semantic differential of *affective meaning*

carried a culturally dependent, shared affective meaning that could be characterized to a great extent using three simple dimensions: evaluation (good versus bad), potency (powerful versus powerless), and activity (lively versus quiet) [191]. This *semantic differential* scale of evaluation, potency, and activity (EPA) is thought to represent universal and cross-cultural dimensions of affective meaning for words. Each affective meaning, as shown in Figure 2.2, is measured on a scale from -4.3 (infinitely bad, powerless, or inactive) to $+4.3$ (infinitely good, powerful, or lively). [1] Impression formation research is conducted by obtaining the EPA values of identities and behaviours both in general and when combined in various descriptive behavioural situations. Then, a regression analysis is performed to estimate the impression created by an event [215, 96].

Studies show that most of the culturally shared sentiments, *affective meanings*, are stable and are only changed gradually over many years. Comparing the stability of the different aspects of sentiment, researchers have found that evaluation (E) is very stable and does not significantly vary from year to year. The change in sentiment ratings can only occur when there is a recognized cultural movement in the society. For example, identities relating to homosexuality changed drastically in North America at the end of the 20th century, from a very negative to a positive sentiment. Potency (P), on the other hand, is less stable than evaluation, and the potency scores for behaviours can change more rapidly than for identities. A good example of an identity that shifted from powerless to powerful is that of young women in North America. A good example of an identity that shifted from powerless to powerful is young women in North American. Activity (A) was found to be the least stable factor in comparison with the other sentiment factors; for example, the evaluation ratings for identities that describe leadership and authority changed substantially in the 20th century, from being not active to very active. Studies have also shown that cross-cultural differences are not significant. Different cultures often share the same perspectives about identities and behaviours; however, sentiment scores might vary from

[1]The $[-4.3, +4.3]$ scale is a historical convention

one culture to another. The variation in sentiment ratings was found to be more prominent in the activity ratings of different cultures than the other aspects of sentiment, and sentiment ratings also might differ across different gender and sub-cultural groups. However, evidence has shown that the variance among different sub-cultures or gender groups is relatively small [99, 101].

Before reviewing the mathematical model and the main principles of affect control theory, some of the terminology used in the ACT literature needs to be defined. First, we clarify the terms used in both the emotions literature and the ACT literature. In the emotion literature, and sometimes in affective computing and SA, writers interchangeably use the terms "affect", "emotion", "sentiment", and "mood" on a regular basis. "Affect" in ACT is a general term that refers to the culturally-shared *affective meaning*. Sentiment or *fundamental sentiment* refers to the affective response measuring the goodness, powerfulness, and activeness of a variety of concepts in a culture. For example, males in North America generally perceive teachers as good, powerful, and quite active. *Transient impression* in ACT, on the other hand, refers to the situated meanings that an event generates. For example, when a teacher exhibits a negative behaviour, they will be perceived as bad, quite powerful, and quite active. " Emotions" in ACT refer to emotional labels associated with *affective meaning* scores that translate the impression created by an event. For example, people associate being afraid with a negative evaluation (i.e., a bad feeling), weakness, and being passive. "Moods" and "traits" refer to more enduring emotions that become a part of an individual's personality and affect their future interactions [145].

Second, we define some of the terms that are used in ACT to describe the main components of an event. *Identity* is a term that is used to describe an individual's role in a society which might refer to their occupation, socio-economic status, or aspects related to religion or politics. People can maintain different identities in a given situation (e.g., a mother and a wife who works as a teacher will alternate between these three identities depending on the situation). *Actor*, the identity or noun in an event, is an agent that enacts a social event. *Behaviour*, the verb in an event-based sentence, describes the act that the actor performs. *Object* is the target of the action that was enacted by the actor. *Setting* is the place where the event took place. *Modifier* is an attribute (adjective) that modifies the actor's or object's identity. *Social event* is any situation that describes a behavioural event when an *actor* who exhibits a *behaviour* towards an *object* in a location or *Setting*. The *actor* and *object* both have role identities (e.g., a teacher). Each behaviour and identity has a sentiment rating or *affective meaning*. *Self-directed event* is any interactional event that the actor performs toward themselves (e.g., self grooming).

ACT postulates that individuals interpret social events using both cognitive and affect processing. During the affect processing stage, individuals define a situation by unconsciously choosing a specific sentiment profile for each element in the event (i.e., actor, behaviour, object, setting). Individuals share the same *fundamental sentiments* about concepts (i.e., identities and behaviours) [20, 99], and these *fundamental sentiments*, defined in three-dimensional representations (*affective meaning*), govern individuals' social interactions. A post-event feeling, *transient impression*, will be produced after an event occurs and this temporary feeling is toward the elements in that specific event and context. The post-event feeling is influenced by the behaviour that the actor performed. For example, good behaviour will produce a positive sentiment toward the actor, object, and setting of an event. If another event occurs which involves the same individuals as in the previous event, people often use their previous feelings as the pre-event feelings of the current event. This process of how sentiments change after an event takes place is called *impression-formation* in ACT literature.

The *deflection*, i.e., the difference between the *fundamental sentiment* and *transient impression*, plays an important role in forming individuals' emotions toward an event's entities. Individuals try to maintain their *transient* states near their *fundamental sentiments* through their actions and interpretations of events. Maintaining or failing to maintain the affective meaning associated with self-identity tends to result in positive or negative emotions respectively. If the *deflection* of transient impression from a fundamental sentiment is high, people tend to re-conceptualize the associated entities in the event to lower the *deflections* [251]. This thesis focuses on the sentiment part of ACT, and therefore, the rest of this section discusses *fundamental sentiments*, *transient impressions*, the *deflection*, and emotions formation. More details about the emotion dynamic and optimal behaviour can be found in Chapters 5 and 6. More details about ACT can be found in [20, 99, 251].

2.2.1 The ACT Mathematical Model

In ACT, each event is typically defined by three or four elements: actor (A), behaviour (B), object (O), and sometimes settings (S). In this thesis, the term event elements or entities refers to the actor, the behaviour, the object, and the setting. The three EPA values represent each of these elements that capture the fundamental sentiments they evoke in terms of evaluation, potency, and activity. According to ACT grammar, the *fundamental sentiment* is represented as follows:

$$f = \{\bar{A}_e\ \bar{A}_p\ \bar{A}_e\ \bar{B}_e\ \bar{B}_p\ \bar{B}_a\ \bar{O}_e\ \bar{O}_p\ \bar{O}_a\},$$

and the *ransient impression* evoked by an event is equal to:

$$\tau = \{\hat{A}_e\ \hat{A}_p\ \hat{A}_e\ \hat{B}_e\ \hat{B}_p\ \hat{B}_a\ \hat{O}_e\ \hat{O}_p\ \hat{O}_a\},$$

where the A, B, and O represent the *actor*, *behaviour*, and *object* respectively, and the subscripts *e*, *p*, and *a* respectively represent the evaluation, potency, and activity. The *fundamental sentiment* or EPA values are denoted by an over-bar and the post-event sentiment or EPA values are denoted by a caret. For example, $\bar{A}_e$ is the *fundamental sentiment* associated with the actor (good vs. bad) and $\hat{A}_e$ is the post-event sentiment associated with the actor or *transient impression*. The *transient impression* is calculated by multiplying a matrix M by t, a vector of interaction terms composed of the fundamental pre-event sentiments. M comprises a set of prediction coefficients that are estimated through the regression analysis of empirically measured impressions:

$$\tau = Mt \tag{2.1}$$

$$\begin{aligned} t \quad = \quad &\{1\ \bar{A}_e\ \bar{A}_p\ \bar{A}_a\ \ \bar{B}_e\ \bar{B}_p\ \bar{B}_a\ \bar{O}_e\ \bar{O}_p\ \bar{O}_a \\ &\bar{A}_e\bar{B}_e\ \ \bar{A}_e\bar{B}_p\ \ \bar{A}_e\bar{B}_a\ \ \bar{A}_p\bar{B}_e\ \ \bar{A}_p\bar{B}_p\ \ \bar{A}_p\bar{O}_a\ \ \bar{A}_a\bar{B}_a \\ &\bar{B}_e\bar{O}_e\ \ \bar{B}_e\bar{O}_p\ \ \bar{B}_p\bar{O}_e\ \ \bar{B}_p\bar{O}_p\ \ \bar{B}_p\bar{O}_a\ \ \bar{B}_a\bar{O}_e\ \ \bar{B}_a\bar{O}_p \\ &\bar{A}_e\bar{B}_e\bar{O}_e\ \ \ \bar{A}_e\bar{B}_p\bar{O}_p\ \ \ \bar{A}_p\bar{B}_p\bar{O}_p\ \ \bar{A}_p\bar{B}_p\bar{O}_a\ \ \bar{A}_a\bar{B}_a\bar{O}_a\} \end{aligned} \tag{2.2}$$

A different M matrix is constructed for each grammar (AB), (ABO), or (ABOS). Each column in the matrix represents the coefficient scores used for computing one of the post-event entities. For example, if the AB (actor-behaviour) grammar is used, the columns are as follows: $A_e, A_p, A_a, B_e, B_p, B_a$. To give an idea of how the M matrix is constructed and how the coefficients affect the calculation of τ, the following equation shows the actor post-event evaluation $\hat{A}_e$ estimated using the impression equations [71]:

$$\hat{A}_e = -0.26 + \mathbf{0.41}\ \bar{\mathbf{A}}_\mathbf{e} + \mathbf{0.42}\ \bar{\mathbf{B}}_\mathbf{e} - 0.02\ \bar{B}_p - 0.10\ \bar{B}_a + \mathbf{0.03}\ \bar{\mathbf{O}}_\mathbf{e} + 0.06\ \bar{O}_p$$

$$+\mathbf{0.05}\ \bar{\mathbf{A}}_{\mathbf{e}}\bar{\mathbf{B}}_{\mathbf{e}} + 0.03\ \bar{A}_e\bar{O}_p + \mathbf{0.12}\ \bar{\mathbf{B}}_{\mathbf{e}}\bar{\mathbf{O}}_{\mathbf{e}} - 0.05\ \bar{B}_e\bar{O}_p - 0.05\ \bar{B}_p\bar{O}_e$$
$$+0.03\ \bar{A}_e\bar{B}_e\bar{O}_e - 0.02\ \bar{A}_e\bar{B}_e\bar{O}_p.$$

The coefficients in the above equations show that the largest factors that indicate the degree to which the post-event evaluation of the actor $\hat{\mathbf{A}}_{\mathbf{e}}$ is positive or negative are the pre-event evaluation of the actor $\bar{\mathbf{A}}_{\mathbf{e}}$ (i.e., how good or bad the actor is) and the pre-event evaluation of the behaviour $\bar{\mathbf{B}}_{\mathbf{e}}$. This means that if a bad person performs a positive action, he/she will still be perceived as a negative person, and if a good person performs a bad behaviour, he/she will have a negative evaluation. The post-event evaluation of the actor is slightly negatively affected by the potency of the behaviour $\bar{\mathbf{B}}_{\mathbf{p}}$ that he/she is performing. This effect indicates that powerful behaviours tend to be perceived negatively by people in the culture. The positive coefficient $\bar{\mathbf{O}}_{\mathbf{e}}$ indicates that the post-event identity associated with the actor is slightly affected by the pre-event identity of the object (i.e., how good or bad the object of the event is). Interacting with those whose identities are considered bad will make us seem bad, and interacting with those whose are considered to be good will make us seem good. This concept is called guilt by association. The combination of the pre-event behaviour and object evaluation $\bar{\mathbf{B}}_{\mathbf{e}}\bar{\mathbf{O}}_{\mathbf{e}}$ also has an effect on how the actor seems. If actors perform a negative behaviour directed at a bad person or a positive behaviour directed at a good person, they will be seen/perceived as a good actor, while they will be perceived as less good if they perform a negative behaviour directed at a good person or a good behaviour directed at a bad person. This interaction term $\bar{\mathbf{B}}_{\mathbf{e}}\bar{\mathbf{O}}_{\mathbf{e}}$ is called the balance term.

These impression-formation equations have been estimated for different cultures and in numerous languages. The five most recent sets of impression-formation equations were developed for the United States [71], Canada [161], Japan [243], Germany [238], and China [243]. The coefficients of the matrix M vary slightly from one culture to another. For example, in an impression-formation equation for Canada, the $\bar{\mathbf{O}}_{\mathbf{e}}$ is associated with a negative coefficient, which indicates that if an actor interacts with a good person, the post-event evaluation of their identity $\hat{\mathbf{A}}_{\mathbf{e}}$ will be slightly negatively affected, and vice versa. In contrast, the $\bar{\mathbf{O}}_{\mathbf{e}}$ is associated with positive coefficient scores in the set of equations for Germany and the United States and with zeros in the equations for Japan and China. The impression equations used in this thesis are the ones for the United States and Canada.

The incorporation of *settings* that indicate where the event took place, such as a school or a library, can be achieved by adding the EPA values of the setting and using the coefficient values of the actor-behaviour-object-setting (ABOS) grammar rather than the actor-behaviour-object (ABO) grammar:

$$f = \{\bar{A}_e\ \bar{A}_p\ \ \bar{A}_e\ \bar{B}_e\ \bar{B}_p\ \bar{B}_a\ \bar{O}_e\ \bar{O}_p\ \bar{O}_a\ \bar{S}_e\ \bar{S}_p\bar{S}_a\}$$

$$\tau = \{\hat{A}_e\ \hat{A}_p\ \ \hat{A}_e\ \hat{B}_e\ \hat{B}_p\ \hat{B}_a\ \hat{O}_e\ \hat{O}_p\ \hat{O}_a\ \hat{S}_e\ \hat{S}_p\ \hat{S}_a\}$$

$$\begin{aligned} t = \{1\ \bar{A}_e\ \bar{A}_p\ \bar{A}_a\ \ \bar{B}_e\ \bar{B}_p\ \bar{B}_a\ \bar{O}_e\ \bar{O}_p\ \bar{O}_a\ \bar{S}_e\ \bar{S}_p\ \bar{S}_a \\ \bar{A}_e\bar{B}_e\ \bar{A}_e\bar{B}_p\ \bar{A}_e\bar{O}_e\ \bar{A}_p\bar{B}_e\ \bar{A}_p\bar{B}_p\ \bar{A}_p\bar{B}_a\ \bar{A}_p\bar{O}_e\ \bar{A}_p\bar{O}_p \\ \bar{A}_p\bar{O}_a\ \bar{A}_a\bar{B}_a\ \bar{B}_e\bar{O}_e\ \bar{B}_e\bar{O}_p\ \bar{B}_p\bar{O}_e\ \bar{B}_p\bar{O}_p\ \bar{B}_p\bar{O}_a\ \bar{B}_a\bar{O}_p \\ \bar{A}_e\bar{B}_e\bar{O}_e\ \ \bar{A}_e\bar{B}_p\bar{O}_p\ \ \bar{A}_p\bar{B}_p\bar{O}_p\ \bar{A}_p\bar{B}_p\bar{O}_a\ \} \end{aligned} \tag{2.3}$$

The *fundamental sentiment* f and *transient impression* t of the self-directed event (e.g., a school girl laughs at herself) are computed using only two elements, the actor and behaviour of the event, as follows:

$$f = \{\bar{A}_e\ \bar{A}_p\ \bar{A}_e\ \bar{B}_e\ \bar{B}_p\ \bar{B}_a\ \}$$

$$t = \{\ 1\ A_e\ A_p\ A_a\ \ B_e\ B_p\ B_a\ A_eB_e\ A_eB_p\ A_pB_e\ A_pB_a\ A_aB_e\}.$$

The addition of attributes or adjectives that modify the identities (e.g., "good friend" or "abusive father") is calculated from the EPA values of both the identity and the modifiers. The EPA profiles of particular modifiers are symbolized as $P = \{P_e, P_p, P_a\}$ and identities as $R = \{R_e, R_p, R_a\}$, and the profile for identity-modifiers amalgamation is $C = \{C_e, C_p, C_a\}$. The modifier-identity profile is computed by summing over the multiplication of the coefficient times the modifier and the identity:

$$C = p\,P + r\,R + a \tag{2.4}$$

where p and r are coefficients estimated from empirical studies of the modifiers and identity, respectively and a is a constant. For example, For example, using the Ontario affective dictionary [161], the affective rating of "father" is $[1.84, 1.78, 0.02]$, "abusive" is equal to $[-2.23, 0.34, -0.02]$, and "abusive father" is equal to $[-1.51, 1.37, -0.21]$.

the *deflection*, which is defined as the discrepancy between the *fundamental sentiment* and the *transient impression*, is calculated from the squared Euclidean distance between the sentiments and impressions, as given in the following equation. The *deflection* does not essentially indicate positive emotions or negative emotions, but instead signifies whether or not the event met someone's expectation

$$d = \sum_i (f_i - \tau_i)^2 \tag{2.5}$$

$$d = (\bar{A}_e - \hat{A}_e)^2 + (\bar{A}_p - \hat{A}_p)^2 + (\bar{A}_a - \hat{A}_a)^2 + (\bar{B}_e - \hat{B}_e)^2$$
$$+(\bar{B}_p - \hat{B}_p)^2 + (\bar{B}_a - \hat{B}_a)^2 + (\bar{O}_e - \hat{O}_e)^2 + (\bar{O}_p - \hat{O}p)^2 + (\bar{O}_a - \hat{O}_a)^2$$

ACT covers two different types of emotions: characteristic and structural. Characteristic emotions are emotions that actors experience when they perform a behaviour that confirms their identity. Structural emotions are emotions that individuals experience while interacting with others. In ACT, emotions are also indexed according to the three values of evaluation, potency, and activity, and emotions triggered by an event are a function of the fundamental identity (for the actor or the object) and the transient identity (for the actor or the object). More details about emotion dynamics can be found in section 5.2.3 and 6.2.

The ACT dictionary contains words for labelling identities, behaviours, modifiers, settings, and emotions. Each word is defined by an average EPA rating for males and females separately (e.g., the rating of father in the Ontario dataset= [1.84,1.78,0.02] for a male and [2.01,2.29,-0.23] for a female, suggesting that females have a more positive evaluation of their father than do males). The EPAs were derived empirically from large-scale studies that took place in different cultures/languages over varying periods of time. Researchers conducted these studies in the United States, Canada, Germany, Japan, China, and Northern Ireland. For example, the Ontario 2001 dataset was gathered between 2001 and 2003 in Guelph, Ontario, and contains 2,294 affective ratings for 993 identities, 601 behaviours, 500 modifiers, and 200 settings. Section 2.4.2 provides additional details about the semantic differential method used for gathering affective ratings.

2.3 Statistical Language Modelling Background

One component of this thesis involved the implementation of several lexicon induction algorithms using variant word representations and similarity metrics as a means of expanding a multidimensional sentiment lexicon. This section provides a short introduction to word representations and word similarity metrics.

2.3.1 Vector Space Model and Word Representations

Statistical language modelling (SLM) is an essential component in many research areas, including natural language processing (NLP), speech recognition, and information retrieval. SLM techniques are designed to estimate the distributions of various language phenomena by applying statistical estimation techniques to real-world training data. Representing language phenomena in terms of parameters has proven effective in a number of NLP tasks, and researchers have proposed several computational models of the semantic or syntactic relationships between words. The vector space model (VSM) is the most traditional statistical language-modelling approach, with a long and rich history in NLP. Based on the distributional hypothesis, which states that words that appear in a similar context have similar meaning [93, 69], a VSM represents similarity between words as proximity in an n-dimensional space induced using unsupervised techniques to capture syntactic or semantic properties. VSM denotes words as vectors, with each dimension corresponding to a feature that might refer to a semantic or syntactic interpretation [278]. VSMs are especially appealing because they can be trained from unlabelled data and used for supervised learning problems, a feature found to be effective for a variety of applications such as name entity recognition (NER), parsing, or sentiment analysis [272]. The earliest effort to induce word meanings or to represent words in terms of vectors was one-hot representation, in which words are regarded as atomic symbols of their co-occurrence with other words in each vocabulary. The feature vector of one-hot representation denotes the Boolean co-occurrence of a given word with other words in the vocabulary (e.g. Cat = [0,0,0,0,0,1,0,0,0,0,....]) that are located along the length of the vocabulary. This sparse and high-dimensional representation is not helpful for measuring the similarity between words, and it is unable to estimate rare and complex words. These limitations led researchers in the NLP field to investigate other unsupervised approaches for inducing word meaning using a variety of methods that can be classified [272] as belonging to four main categories: distributional (count-based), clustering-based, distributed, and hyperparameter-tuned count-based word representations.

Distributional Word Representation

Distributional representation signifies words based on the co-occurrence matrix F of size $v \times c$ where v is the vocabulary size and c is some context. Matrix F computes the co-occurrence frequency of certain items (terms, word, bi-gram) in a particular situation (context, document, pattern). The co-occurrence matrix can be a word-by-word matrix $w \times w$ that estimates the co-occurrence frequency f_{ij} of the w_i and w_j within a specific boundary (context window), or a word-by-document matrix $w \times d$ where d is the number of documents in the data. One might also construct it to capture any other properties, such as modifier$\times$adverb, person$\times$ product, or subject$\times$verb$\times$object, etc. The word-by-word approach estimates the co-occurrence frequency based on the maximum likelihood probability given by:

$$P(w_i, w_j) = \frac{f_{w_i,w_j}}{N} \tag{2.6}$$

where f_{w_i,w_j} is the co-occurrence frequency of w_i and w_j in N, which is the size of a window (or corpus). In the word-by-document approach, the word co-occurrence frequency is denoted by $f_{w,d}$, which is the number of times the word w appears in document d in a set of documents D:

$$P(w, d) = \frac{f_{w,d}}{D} \tag{2.7}$$

Each row in the co-occurrence matrix has a dimensionality of c that may be too large for use. To solve this high-dimensionality problem, NLP and information retrieval (IR) researchers have investigated numerous techniques for mapping this sparse matrix F to a lower dimensional space of size $v \times s$ where $s \leq c$ using some function g where $f = g(F)$. The methods most commonly used for computing this function are latent semantic analysis (LSA) [139], principal component analysis (PCA), or clustering-based techniques. To solve zero-frequency problems and to obtain better estimates of infrequent words, researchers have also employed several smoothing techniques in distributional representation, such as term frequency-inverse document frequency (tf-idf) weighting [228], pointwise mutual information (PMI) [276], positive PMI [189], and a t-test.

Clustering-Based Word Representations

Clustering-based approaches induce word meanings by performing unsupervised clustering over words and then mapping a vocabulary of words to different clusters. To overcome the spareness and dimensionality problem inherent in computing the co-occurrence frequency of n-gram models, a number of researchers have investigated methods for inducing word meaning in a cluster-based fashion [148, 272]. The Brown hierarchical clustering algorithm [32], the most commonly used word-clustering algorithm, induces clustering over words by maximizing the mutual information in bigrams. Based on the distributional hypothesis, Brown's algorithm relies on the assumption that similar words have a similar distribution of words to their immediate left and right. The goal of this algorithm is to partition a vocabulary V of a set of words $\{w_1, w_2, ..., w_n\}$ into k clusters so that each word is assigned to a cluster $c \in k$. The quality of clustering c is measured by computing the logarithm of the probability p normalized by the length of the text:

$$P(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} p(w_i|c(w_i))\, p(c(w_i)|c(w_{i-1}))$$

$$\begin{aligned} Quality(c) &= \sum_{i=1}^{n} \log\ p(w_i|c(w_i))\, p(c(w_i)|c(w_{i-1})) \\ &= \sum_{c=1}^{k} \sum_{c'=1}^{k} p(c, c') log \frac{p(c, c')}{p(c)p(c')} + \sum_{w}^{n} p(w)\, log\, p(w) \end{aligned}$$

$$p(c, c') = \frac{n(c, c')}{\sum_{c,c'} n(c, c')}\ ,\ p(c) = \frac{n(c)}{\sum_{c} n(c)},\ p(w) = \frac{n(w)}{n}$$

where $n(w)$ is the number of times word w appears in a text, $n(c)$ is equal to the number of times a word in cluster c appears in the text, and $n(c, c')$ is equal to the number of times a word in cluster c appears with a word in a cluster c'. Brown's clustering algorithm has been evaluated using a variety of contexts, such as NER [172], dependency parsing [132], and other NLP applications.

Distributed Word Representations

Distributed word representations, or neural word embeddings, represent words in a low-dimensional, continuous representation whereby each dimension corresponds to semantic or latent syntactic features. Like distributional word representation, distributed representation is usually based on co-occurrence statistics. The difference is that it is dense, more compact, and less sensitive to data sparsity and can implicitly represent an exponential number of word clusters. Distributed word representation has a long history in NLP [104, 220, 61], and it is typically induced using a neural language model such as a feed-forward network or a recurrent neural network. The idea of using a neural network language model to induce a distributed word representation was introduced by Bengio et al. [19], and several subsequent variants have been proposed. The Bengio et al. model is based on the n-gram model, in which the probability of a sequence of words is determined by the $n-1$ preceding words:

$$p(w_1, w_2, ..., w_t) = p(w_1)p(w_2|w_1)....p(w_t|w_{t-1}, ..., w_{t-n+1}) \tag{2.8}$$

The model developed by Bengio et al. [19] first maps each w_{t-i} to a d-dimensional feature vector $c(w_{t-i})$ and concatenates each feature vectors in the n-gram example $x = c(w_{t-n+1}) \oplus c(w_{t-n}) \oplus ...c(w_{t-1})$, where c is a lookup table and $\oplus$ is a concatenation. A standard multilayer neural network is then trained by maximizing the log-likelihood. The output of the hidden layer is computed using hyperbolic tangent function $tanh$, and the output of the output layer is computed using a $softmax$ activation function:

$$P(w_t|w_{t-n+1}, ...w_{t-1}) = \frac{e^{z_m}}{\sum_k e^{z_k}}, \tag{2.9}$$

$$z_m = b_m + \sum_i w_{mi}\, tanh(h_i + \sum_j v_{ij}\, x_j), \tag{2.10}$$

where b and h are the bias weight, and w and v are the weight of the corresponding output and hidden layers. The model is trained by minimizing the error in the training data with respect to the model parameters, which can be achieved by back-propagating the error from the output to the input layer and updating the weight and bias accordingly.

Mikolov et al. [168, 169] proposed a new variation of the neural language model using a recurrent neural network (RNN) to induce word meanings from their contexts. This approach overcomes the limitation of the earlier approach by Bengio et al. [19], in which the dependencies between words were restricted to a fixed text size (n-gram). RNN differs from standard neural language models in that it operates on the internal state spaces rather than the input space, allowing the model to extend the dependencies over unspecified intervals. More recently, Mikolov et al. [170] proposed skip-gram negative-sampling (SGNS) neural language models that yielded superior performance with respect to many NLP tasks. The SGNS model is designed to learn word representations that are useful for predicting nearby words (context words). The architecture of the neural model is composed of two layers: a linear hidden layer and a soft-max output layer. The objective of the skip-gram model is to maximize the averaged log probability:

$$p(w_{t+j}|w_t) = \frac{\exp({v'_{w_{t+j}}}^T v_{w_t})}{\sum_{w=1}^{W} \exp({v'_w}^T v_{w_t})} \tag{2.11}$$

where w_{t+j} signifies the context words, w_t is the target word, v'_w and v_w are the respective output and input vectors of a word w, and W is the number of the words in the vocabulary.

In a recent study, Pennington et al. [199] proposed Global Vectors for Word Representation (GloVe), which factorizes the word-context co-occurrence matrix in order to reformulate the optimizations of the SGNS model. This proposed model is designed to minimize a cost function (weighted least squares) equal to the difference between the dot product of the vectors of two words and the logarithm of the number of their co-occurrences:

$$J = \sum_{j}^{V} \sum_{i}^{V} f(X_{ij}) w_i^T w_j + b_i + b_j - \log X_{ij} \tag{2.12}$$

where w_i and b_i are the word vector and the bias, respectively, of the main word i; b_i and b_j are the context word vector and bias, respectively, of word j, and f is a weighting function that assigns a relatively lower weight to rare and infrequent co-occurrences.

Word representations (distributional, cluster-based, and distributed) have been employed in a variety of NLP applications. Turian et al. [272] compared these word space vectors to word features for chunking (syntactic sequence labelling), and NER tasks. Pennington et al. [199] assessed GloVe vis-a-vis other word representations, including the SGNS model, with respect to word similarity, word analogy, and NER tasks. Their results showed that GloVe performs better than other word representation for all tasks. This claim was later refuted by subsequent studies that revealed that the SGNS model often outperforms the GloVe model in regard to the above tasks [144].

Hyperparameter-Tuned Count-Based Word Representation

Levy et al. [143] demonstrated that the SGNS model implicitly factorizes a word-context PMI matrix. They analyzed the objective function of the skip-gram using negative sampling, which often provided state-of-the-art accuracy in many NLP contexts. They derived a simple matrix factorization that applies shifted positive pointwise mutual information (SPPMI) and singular value decomposition (SVD) on the co-occurrence statistics (the word-context matrix):

$$SPPMI(w_i, w_j) = max\{\frac{p(w_i, w_j)}{p(w_i)p(w_j)} - log_2(k), 0.0\}, \tag{2.13}$$

where $p(w_i, w_j)$ is the empirical co-occurrence probability of a pair of words w_i, w_j; $p(w_i)$ and $p(w_j)$ are the marginal probabilities of w_i and w_j, respectively; and k is the number of negative samples in word2vec (e.g., $k = 10$). They found that this simple technique can produce word vectors that have the same quality as the skip-gram word vector, and they demonstrated that using typical count-based features with some hyperparameter tuning can achieve state-of-the-art accuracy with respect to many NLP tasks, including word similarity and analogy tasks. In another recent study, Levy et al. [144] found that the negative sampling approach used in the SGNS model samples words according to a smoothed unigram distribution. This process ensures that the model is sampling frequent words relatively less often than their actual co-occurrences. This context distribution smoothing (CDS) is transferable to the PMI through the raising of the frequency of the context word by a smoothing factor α, as follows:

$$PPMI_\alpha(w, c) = max\{log_2 \frac{p(w, c)}{p(w)p_\alpha(c)}, 0.0\}, \tag{2.14}$$

where $p_\alpha(c)$ is the smoothed marginal probability of word context c, which is equal to $\frac{f_c^\alpha}{\sum_c f_c^\alpha}$ where f_c is the frequency of context c. This variant of the PMI was found to alleviate the bias toward rare words and to provide consistent improvement in the results. Levy et al. [144] also proposed that the SVD be calculated and that some weight p be added to the eigenvalue matrix Σ_d,

$$W_{SVD} = U_k \Sigma_k^p, \tag{2.15}$$

where k represents the top rows of U and Σ, and p can at best be equal to 0.5 or 0.0. Levy et al. also presented an extensive evaluation of the proposed factorization approach, the SGNS model, and GloVe with respect to word similarity and analogy tasks. They showed that the results of using CDS and an SVD word vector (W_{SVD}) are comparable to those obtained by the SGNS model and that both the factorization approach and the SGNS model perform better than GloVe [144].

Improving Word Embedding Using Lexical Features

Several recent studies have proposed models that incorporate semantic features to improve the performance of word embeddings in many NLP tasks [301, 308, 66, 27]. Yu et al. [308] proposed a joint relation constrained model (RCM) designed to improve the continuous-bag-of-words (CBOW) [170] word embeddings using similarity information obtained from semantic lexicons. The algorithm learns word representations from both a corpus and a semantic lexicon by combining the CBOW and RCM objective functions. CBOW computes the probability of a target word w_t, given context words w_{t-c}^{t+c} as follows:

$$p(w_t|w_{t-c}^{t+c}) = \frac{\exp({v'_{w_{t+j}}}^T \sum_{-c\leq j\leq c, j\neq 0} v_{w_t+j})}{\sum_{w=1}^{W} \exp({v'_w}^T \sum_{-c\leq j\leq c, j\neq 0} v_{w_t+j})} \tag{2.16}$$

where v_w and v'_w are the input and output word vector, respectively, and c is a window size. The RCM defines an objective function that maximizes the log probability of a word w being related to words in the R_w subset of the vocabulary as follows:

$$p(w|w_i) = \frac{1}{N} \sum_{i=1}^{N} \sum_{w \in R_{w_i}} \log p(w|w_i) \tag{2.17}$$

And the joint RCM and CBOW as follows:

$$\frac{1}{T} \sum_{t=1}^{T} \log p(w_t|w_{t-c}^{t+c}) + \frac{C}{N} \sum_{i=1}^{N} \sum_{w \in R_{w_i}} \log p(w|w_i) \tag{2.18}$$

This model was evaluated with respect to semantic similarity tasks and has demonstrated enhanced performance compared to the CBOW and skip-gram models. Farugi et al. [66], on the other hand, tackled the problem of incorporating semantic knowledge in word embeddings differently by proposing a retrofitting model, a post-processing method that fits a pre-trained word embedding using a relational graph that encodes semantic similarity between words. The goal of a retrofitting model is to learn new word representations such that they are close (based on similarity metrics) to their neighbours in the semantic lexicon graph. The objective function of this model, as shown below, is thus to minimize the Euclidean distance between a word vector $\hat{q}_i$ and its neighbours q_j in a semantic graph E:

$$\sum_{i=1}^{n} \alpha||q_i - \hat{q}_i||^2 + \sum_{(i,j)\in E} \beta_{ij}||q_i - q_j||^2 \tag{2.19}$$

where n is the number of words in the vocabulary, α and β are the weights for controlling the strength of the association and $\hat{q}_i$ and q_i are the original word vector and the retrofitted word vector, respectively. The retrofitting approach outperformed several pre-trained word vectors in addressing a variety of natural language problems (e.g., word similarity, syntactic relationship, and sentiment analysis). In an approach similar to the joint learning method proposed by Yu et al. [308], Bollegala et al. [27] combined the GloVe objective function (Equation 2.12) with a regularizer J_s that encodes the semantic relationship between a word w and its context words in a vocabulary v:

$$J_s = \frac{1}{2} \sum_{i\in v} \sum_{j\in v} R(i,j)(w_i - w_j), \tag{2.20}$$

where $R(i,j)$ is a binary function that returns 1 if the two words are related and 0 otherwise. The method developed by Bollegala et al. was evaluated against other proposed methods for different semantic relation tasks, and it outperformed the approaches presented by Yu et al.'s [308], Faruqui et al. [66], and Xu et al. [301] approaches. Similarly, Bian et al. [23] proposed a model that incorporates morphological, syntactic, and semantic knowledge in order to enhance word embedding learning. Lui et al. [155] created a technique that incorporates semantic knowledge into a corpus-based word embedding learning method, in which this semantic knowledge is represented in the form of the ordinal ranking inequalities of the words. This learning framework acquires distributed word embeddings by solving a constrained optimization problem. This approach was evaluated using several NLP tasks and produced better results than either the distributed word embedding approaches (SGNS and GloVe) or the method proposed by Bian et al.

2.3.2 Word Similarities

Two types of word similarities are used in NLP and related research areas: relational and attributional. Attributional similarity measures the degree of correspondence between attributes of words. If A and B have a high degree of attributional similarity, they are called synonyms (e.g., car-vehicle). In contrast, relational similarity measures the correspondence between the internal relationships of two word pairs. If two word pairs a : b and c : d have a high degree of relational similarity, they are labelled analogous (e.g. $tree : leaves :: book : papers$). Verbal analogies often use this format $A : B :: C : D$ where ":" means "is to" and "::" means "as" so it reads A is to B as C is to D.

As a term, semantic similarity could be confusing because it might refer to either type of similarity. The work presented in this thesis was directed at measuring the attributional similarity between words in order to extend the affective meaning lexicon. For this reason, the terms "semantic similarities" and "word similarities" are both used to refer to attributional similarities. In general, the similarity between two words (vectors x and y of dimension n) can be calculated using different metrics such as the dot product, the Euclidean distance, and the cosine similarity.

Attributional similarity can be measured in several ways using lexicon-based methods, corpus-based (frequency-based) methods, or combination of two. The corpus-based, or count-based, approach is based on co-occurrence statistics, which generally means that words in similar contexts are considered to have similar meanings. Lexicon-based, or thesaurus-based, methods are based on Resnik's [211] claims that a natural way of calculating semantic similarity in a taxonomy is to measure the distance between the nodes that correspond to the items being compared: the shorter the path, the more similar the items. Only a handful of studies have examined the viability of word lexicons in computing word similarities. Turney et al. [275] compared different algorithms for computing word similarities. They examined the methods most commonly used, including LSA [139], PMI for information retrieval (PMI-IR) [276], a thesaurus-based approach based on Roget's Thesaurus [116], and the Terra and Clarke algorithm [265], as well as combinations of all of these algorithms. Different combinations have been evaluated, and the results have shown that the combinations yielded greater accuracy, than any single approach.

2.4 Sentiment Analysis Literature Review

Sentiment analysis, or opinion mining, is a general term that refers to the research area that involves examination of the problems related to the detection, extraction, and classification of the sentiment or opinion expressed in textual input. SA research can be grouped into five main categories: *subjectivity classification* , which is the task of employing machine-learning (ML) algorithms and/or a sentiment lexicon to determine the subjectivity of a given text (e.g., a news article or movie review) [194, 302, 213, 285]; *polarity classification*, which is the task directed at determining the orientation of the sentiment in a given text as either positive or negative or as belonging to other predefined classes; *review usefulness measurement*, whose goal is to establish the quality of a review (e.g., good vs. bad) [153], *opinion spam detection*, which is a subtask aimed at detecting spam reviews; and *lexicon creation*, which is focused on creating a sentiment dictionary using labelled words. The work presented in this thesis was concentrated on *polarity classification* and *lexicon creation*. The following sections thus provide background about the research that has been conducted with respect to these two categories. Additional detail about the other categories and about SA and opinion mining research can be found in [195, 210, 153].

The problem of *polarity classification* has been tackled at different levels of granularity: at the document, sentence, phrase, and word levels. One of the most common research problems in sentiment or polarity classification is to determine the sentiment polarity of a document (e.g., product review, comment, news article, or on-line discussion). SA researchers have performed *polarity classification* using either ML-, lexicon-, or hybrid-based methods. The latest technique combines both ML- and lexicon-based methods. Researchers have also looked at the identification of sentiment in natural language text from both the readers' and the writer's perspectives. However, the mainstream of SA research has been focused on the writers' sentiments or opinions.

The work conducted for this thesis included the investigation of the viability of incorporating ACT in a structured sentiment classification model that identifies sentiment from the reader and interactant perspectives by taking into account the dependency/interaction between words. The problem was also tackled at the sentence and document/article levels. The following subsection thus discusses previous research efforts related to *polarity classification* using ML-based, lexicon-based, hybrid-based, and compositional-based model methods in addition to outlining studies that have addressed the problem of detecting reader and interactant sentiment. Subsequent subsections outline research efforts with respect to the subarea of polarity lexicon induction as well as previous attempts to use an affective meaning dictionary or ACT principles.

2.4.1 Polarity Classification

Machine Learning-based Methods

There have been two main ML methods in SA research: supervised learning and unsupervised learning. Supervised learning methods generally use the occurrence frequencies of words that indicate opinion/sentiment and then classify the word as positive or negative, with three-star ratings or five-star ratings using one of the common classification methods such as naive Bayes [287, 310]; maximum entropy; support vector machine (SVM) [52, 287, 227, 310, 179]; logistic regression; artificial neural network (ANN) [179]; or K-nearest neighbour (KNN) [287]. Researchers have tried many combinations of features such as part-of-speech (POS) tagging, n-grams, a term weighting scheme, a sentiment lexicon, and term presence [196, 77]. Other studies also added other features to the bag-of-words model as a means of accounting for negations, intensities, part-of-speech tags, and contextual valence shifters, and these methods have been found to outperform standard bag-of-words techniques [309, 296, 128, 58, 112, 279, 146, 300].

Unsupervised learning methods, on the other hand, classify the sentiment of a phrase or text by computing the semantic orientation (SO) of the textual input with respect to positive and negative paradigm words. Goldberg et al. [81] presented a semi-supervised graph-based algorithm for classifying the rating of unlabelled reviews by assessing the similarity between the reviews and a set of labelled reviews. After constructing a similarity graph, the algorithm labels the unlabelled reviews (nodes) by solving an optimization problem that forces ratings to be smooth throughout the graph.

Other approaches have considered the syntactic structure of the sentences in determining the sentiment polarity of a given document. Nasukawa et al. [184] proposed a method that identifies the relationship between the target subject (aspect or feature) and the sentiment expression associated with it as a means of facilitating the identification of the sentiment orientation and polarity of a document. To identify subjects in sentences and their related expressions, their study used several techniques including part-of-speech (POS) tagger and a syntactic parser in order to determine local dependencies and fragment boundaries. This method was applied on several corpora in different domains, yielding precision values ranging from 75% to 95% on a benchmark corpus and real-word corpus, respectively. The authors demonstrated that identifying the relationships of sentiment expressions to subjects improves the overall accuracy of SA detection. However, they reported failure with respect to discovering the sentiment polarity of some sentences due to the complexity of their structure. MacDonald et al. [163] proposed a structured model that jointly predicts sentiment orientation at the sentence and document levels. Their proposed model represents documents and sentences in the form of a clique. The model was trained using a margin-infused relaxed algorithm (MIRA) learning algorithm [49] and the inference was performed using the Viterbi algorithm [283]. This approach was evaluated using online product reviews from Amazon.com and yielded a 62.6% degree of accuracy at the sentence level and 80% at the document level.

Whitelaw et al. [293] proposed a method that can handle complex statements, such as “not very happy” and “very good,” that the baseline technique (bag-of-words) might fail to classify accurately. They augmented the standard bag-of-words technique with appraisal groups represented as sets of attribute values in a semantic taxonomy based on appraisal theory. The study adopts a semi-automated process to build a lexicon of appraising adjectives and other modifiers. The accuracy reported for the use of appraisal groups reached 90% when an SVM technique was employed on a standard movie review corpus [194]. Joshi et al. [122] proposed an algorithm that combines uni-gram features with dependency triplets that link words and their modifiers. This model was evaluated on an opinion mining dataset and found to perform better than the unigram model and other dependency features. Lia et al. [147] presented a probabilistic modelling framework for simultaneously detecting sentiment and topic by incorporating the distribution over topics and using them as features for each document. Their study employed latent Dirichlet allocation (LDA) to infer the distribution over topics, then evaluated the approach on several datasets including the Pang and Lee [194] movie review and multi-domain dataset.

Lexicon-Based Methods

Turney and Littman [274] applied an unsupervised method that determined the semantic orientation at the word and phrase levels by calculating the PMI between a phrase and a set of positive and negative paradigm words, such as "good" and "bad". Turney et al. employed an AltaVista NEAR operator to search a document that contains paradigm words near the words of the phrase and then computes the average sentiment orientation of all of the phrases. This approach yielded an accuracy level between 65% and 85% with respect to predicting the sentiments in reviews from Epinions.com. Lu et al. [158] computed the sentiment polarity or strength of a given review by multiplying the strength of the adjectives and adverbs used in the review. The strength of the adjectives and adverbs was calculated in a semi-automated fashion using graph-based label propagation that built on progressive rules and search engine results. When this approach was evaluated using a Chinese hotel review lexicon, it yielded a precision score of 71.65%. Kontopoulos et al. [131] proposed an analogy-based approach for Twitter SA. Their proposed model breaks down a textual input into aspects and assigns a sentiment grade for each aspect/feature. A domain-specific ontology was first populated and then used for classifying the sentiment of textual input. Sentiment classification was performed using a web service called *OpenDover* [2] which was used for detecting the sentiment in an opinionated text.

Only one previous attempt has been made to use the affective meaning dictionary sentiment classification. Yelena Mejova [167] presents an approach for using the ACT semantic dictionary as a means of determining the sentiment polarity of text. Her study evaluated the ACT lexicon using a simple classification method (voting strategy) that computes the average difference between the scores of positive and negative words in a document. With an accuracy level of 59% on the movie review dataset [194], this method outperformed classifiers based on SentiWord [53] and WordNet-Affect [256]. Despite the fairly encouraging nature of these results, they were obtained using a limited-sized lexicon and only an evaluation rating. In a similar vein, Mullen et al. [181] presented a new technique for incorporating the three-dimensional semantic differential values developed by Osgood et al. [191] into a text classification tool. They derived the values of the semantic differential based on a relative minimal path length (MPL) using WordNet [171], which generated 5,410 adjectives. Combining the three text-wide values from Osgood et al. [191, 190]

[2] OpenDover sentiment tagging web service: http://opendover.nl/

with Turney's values [277] and unigram words [196]] yielded an accuracy level of 84.6% with respect to classifying movie reviews from Epinions.com into positive or negative sentiments. However, the use of semantic differential values alone yielded poor results. The study also failed to clarify whether the findings were due to the theory or the implementation.

Compositional-Based Methods

One limitation of bag-of words models is their inability to embody the syntactic dependency between words. This flat representation ignores word order and sentence structure and fails to account for negations and word modifiers that might reverse the meaning of the sentence. As well, different combinations of the same words might have different meanings. There have been only a handful of attempts that use the composition of sentences to infer sentiment at the phrase and sentence levels as a means of addressing this problem. Wilson et al. [296] proposed a classification approach that uses bag-of-words representation and incorporates dependency along with head-modifier relationships. They implemented a two-step classification approach that first determines the subjectivity of the sentences (neutral versus polar) and then classifies them into more finely grained classes (positive, neutral, and negative). The proposed model achieved significantly better results than the baseline: 75.9% accuracy on subjectivity classification and 65.7% on fine-grained classification. Nakagawa et al. [183] proposed an approach that includes consideration of the interaction between words in English and Japanese sentiment classification. They used a dependency tree to capture compositionality in sentences and then employed the L-BFGS algorithm to apply conditional random field (CRF) and belief propagation techniques for the estimation of the model parameters. The algorithm achieved a degree of accuracy of between 77% and 86% on a variety of corpora, with these end values representing the results for movie reviews and the Multi-Perspective Question Answering (MPQA) opinion corpus, respectively.

To overcome the limitations of bag-of-words models, several approaches have employed compositional formal semantic models based on Frege's principle, which is that the meaning of a sentence is a function of the meaning of its words [72]. The semantic meaning of a sentence $S = \{w_1, w_2, ...w_n\}$ consisting of n words with a grammatical structure $t_1t_2..t_n \rightarrow^{\alpha} S$, where t_i is the grammatical type of w_i and $\rightarrow^{\alpha}$ is the pre-grouped reduction (typological ordering) of a sentence S, can be represented as follows [43]:

$$\overrightarrow{w_1\, w_2\, ..w_n} = F(\alpha)(\overrightarrow{w_1}, \overrightarrow{w_2}, ...\overrightarrow{w_n})$$

where w_i is the distributional meaning of a word, and $F(\alpha)$ is a function that maps a sentence S with syntactic structure α to a vector of meaning. For example, the logical formula for a sentence such as "Cats like milk" will be $like(cats, milk)$ where the meanings are represented as vectors. Mitchell at al. [173] investigated several common operations such as vector addition and pointwise multiplications to represent the compositionality of a sentence. These operations are nevertheless undesirable because they lead to erroneous equality as a result of their cumulative nature where $\overrightarrow{w} + \overrightarrow{v} = \overrightarrow{v} + \overrightarrow{w}$ and $\overrightarrow{v} \odot \overrightarrow{w} = \overrightarrow{w} \odot \overrightarrow{v}$, in which, according to this property, both sentences "John bites a dog" will be equal to "A dog bites John". To deal with this problem, Coecke et al. [44] developed a cross-disciplinary framework based on logic, category theory, and physics. In their approach, a sentence vector is a function of the Kronecker product of its words. For example, the meaning of a subject, a verb, and an object in a sentence are equal to the pointwise multiplications of the meaning of the verb by the Kronecker product of its subject and object. Researchers employed distributional word representation (Tf-Idf) [87] to evaluate this framework based on a toy handcrafted corpus [88] and on sentence similarity. Both experiments showed promising results and improved performance compared to non-compositional approaches and other compositional models [173].

Several recent studies involved the investigation of the viability of neural network models with respect to learning a continuous representation of text for sentiment analysis. Several neural network models were implemented for the learning of generic word embeddings from different corpora [19, 170, 14] or sentiment-specific word embeddings based on consideration of the sentiment of the text [137, 264]. Other studies presented an approach to learning sentence or document representation using neural network models. Yessenalina et al. [307] presented a sentiment analysis approach that represents words as matrices and uses iterated matrix multiplication as phrase-level representation. This model outperformed a model based on bag-of-words with respect to classifying the MPQA sentiment dataset [295], which contains 535 news articles that are manually annotated with phrase-level subjectivity and intensity. Socher et al. [248] introduced an approach that uses recursive neural networks (RNNs) to construct sentence-level semantic representations that then are used as features for classifying each phrase and then the entire document. The RNN model was designed in a bottom-top fashion over a parse tree. The sentiment of each phrase was determined based on the sentiment expressed by its constituant parts. No feature engineering is needed, nor are any semantic or linguistic resources utilized, but a large set of manually annotated corpora with fine-grained sentiment scores is required. This model outperformed state-of-the-art methods on movie review corpora achieving an accuracy level of 85.4%. Tang et al. [263] proposed a neural network approach that learns document representation for sentiment classification. The proposed model consists of a convolutional neural network (CNN) or long short-term memory (LSTM) to construct sentence embeddings from word embeddings. These sentence representations are then fed into a gated recurrent neural network that uses these semantic features to induce document representations. The model was trained in a supervised fashion using stochastic gradient descent with a cross-entropy loss function. This approach was evaluated with respect to a number of opinion mining corpora, and the model outperformed the RNN, exhibiting degrees of accuracy of 67.6% and 45.3% for LSTM on Yelp14 and IMDB datasets, respectively. Despite the success of neural network methods in capturing the semantic and syntactic latent features of an input text, these methods are computationally intensive, and they require a substantial amount of data from a hand-coded corpus.

Reader Sentiment Prediction

SA work has generally been focused on the extraction of emotions from the writer's perspective, with only a few recent studies tackling the problem of predicting readers' emotions, a deficiency that is due to the lack of reader-perspective opinionated corpora. Lin et al. [149] present an approach in which they used features such as bigrams of Chinese characters, news article metadata, affix similarities, and emotional expressions to classify Chinese news articles (e.g., Yahoo!-Kimo news) into user emotional ratings (happy, sad, surprise, etc.). Using an SVM, their model achieved 77% accuracy. Similarly, Yang et al. [306] compared reader emotions derived from Yahoo!-Kimo news with writer emotions in Chinese blog posts. They plotted the results according to the Russell two-dimensional graph (valence-arousal) [233], which showed that reader and writer emotions are associated with specific quadrants of the graph. It is worth mentioning that each corpus was annotated with totally different sets of emotions: eight emotions were associated with the news corpus, and 14 with the blogs corpus. Bhowmick et al. [22] presented a multi-label classification approach (RAKEL) using words, polarity, and semantic frame features to classify news articles into multiple emotion classes: disgust, fear, happiness, and sadness. Evaluation of 1,306 sentences from a news archive revealed the superiority of their approach, which resulted in an accuracy level of 77%.

In 2007, Task 14: Affective Text, in SemEval-2007, was to classify Google and CNN news headlines as either positive or negative expressions, as well as to identify distinct emotions such as anger, sadness, happiness, and surprise [255]. The study proposed a number of approaches for completing this task. Katz et al. presented a supervised approach to the classification of news headlines using Roget's New Millennium Thesaurus and a feature employed for emotion-word-mapping procedures. The next step in their method was the computation of the average emotional score, which was then assigned to the entire document [127]. The reported results were 88.58%, 19.46 %, and 8.62% for accuracy, precision, and recall, respectively. Kozareva et al. [133] presented an unsupervised learning approach that computes PMI scores for words in headlines vis-a-vis sentiment words by using the headline to query search engines (MyWay, AllWeb, and Yahoo). The results for accuracy, precision, and recall using this approach were 85.72%, 17.83%, and 11.27%, respectively. Employing lexical resources such as WordNet, SentiWordNet, and WordNet-Affect, a linguistic rule-based approach [38] achieved enhanced accuracy by means of identifying the main subject in a headline and then assigning it a higher weight than other words. This approach also includes consideration of the extraction of the names of entities and negations. The reported levels of accuracy, precision, and recall of this method were 89.43%, 27.56%, and 5.69%, respectively.

Interactant Sentiment Prediction

This section discusses work related to modelling human emotions that exist in or arise from textual content. More specifically, the studies tackled the problem of modelling emotions that arise from event-based text such as fairy tales or news articles. The modelling of emotions that develop from fairy tales has been a topic of investigation on the part of several researchers. Alm et al. [4] annotated and analyzed the pattern and distribution of emotions in a corpus composed of three sets of children's stories by Beatrix Potter, H. C. Anderson, and the Brothers Grimm. They annotated 22 fairy tales with the feelers' (main characters') emotions based on Ekman's basic emotions [60], and they analyzed the pattern of emotional development contained in these fairy tales. They found that the emotions in fairy tales often follow a specific pattern: they start with neutral emotions and end with happy ones. They also discovered that, in fairy tales, neutral is the most common emotional state, while anger and sadness are the most prolonged.

Another study used the NRC Emotion Lexicon to analyze the distribution of emotions and the density of emotion words in the most popular novels and fairy tales [177] [178]. The researchers examined the polarity and emotion density of 292 novels and 453 stories as well as the associations between some emotions and the targets. For example, they showed how fear was more present in German literature during World War I and World War II. Compared with novels, fairy tales were also found to contain a greater range of emotions such as anticipation, disgust, and surprise. Similarly, the authors of [284] looked at fairy tales written in German and then computed the inter-rater agreement between human annotators. The human annotators were asked to annotate the emotions perceived from entire stories and from different units of the story text. This technique of manually annotating fairy tale text has been the primary method used by most researchers in the field. Only one study, which was based on international political interactions, involved the exploration of the development of sentiment produced by a sequence of events [98]. Using ACT equations, the model analyzes sentiment associated with the subject and object by first considering that the initial impression and sentiment equal zero, next computing the re-identified identity of both interactants, and then calculating the ideal behaviour on the basis of the current and previous events. The model was evaluated using four interactants and 3,077 events describing real interactions among Israel, Egypt, Syria, and the PLO from 1971 to 1978 [98].

Joseph et al. [120] recently proposed a probabilistic graphical model that integrates an ACT-based Gaussian mixture model (ACT-GMM) with a language model that extracts the semantic features between words. The event components, or entities (i.e., actor, behaviour, and object), are drawn from a simple language model (categorical distribution with a Dirichlet prior). The fundamental sentiments, or EPA profiles, are sampled from a Gaussian distribution governed by latent variables drawn from an inverse chi-squared distribution, which is taken from the categorical distribution of the entities. The transient impression is computed using the ACT impression-formation feature, and the deflection is calculated in a stochastic manner. The model inference is performed in two steps: maximum a posteriori (MAP) for learning the parameters of the language model and expectation maximization (EM) for learning the parameters of the ACT-GMM.

The main goal of the model is to estimate the EPA profiles of identities and behaviours in an event-based article. The researchers evaluated the proposed model with respect to predicting the behaviour of a particular actor and object in event-based news articles, and they compared the results against unigram and bigram language models in addition to the outcome from a model that uses only ACT. The results of the evaluation showed that the proposed model outperforms the unigram and ACT-only models, but the performance of the bigram language model was significantly better. Joseph et al. also included a case study of the way news media perceived the identities and behaviours involved in the Arab spring events.

It is worth mentioning that the method proposed in this thesis is the first study that incorporates ACT for modelling human emotions in natural language text [2]. The proposed model differs from the approach adopted by Joseph et al. in that it models reader and interactant emotions/sentiment and predicts the optimal behaviour of the interactants with respect to the next event in an entirely deterministic manner. In contrast, the Joseph et al. method predicts the current behaviour of a particular actor and object using a probabilistic language model that incorporates ACT impression-formation equations. Their approach neither addresses the sentiment aspect of the theory nor aims to estimate human emotions triggered by events.

The main limitations of the proposed polarity classifications methods that this thesis addressed are threefold. Most of the proposed approaches in the sentiment analysis field aimed at predicting the sentiment expressed in opinionated documents and fewer studies have looked at event-based documents. The majority of the research efforts in the sentiment analysis research area used n-gram models combines with some features such as POS tags or dependency features and only a few studies considered the syntactic dependency between all the words in the sentence. They also often classify emotion into a score between two polarities or some predefined classes while the multidimensional model was found to represent human emotions more adequately [70].

2.4.2 Sentiment Lexicon Induction

One of the most popular methods of performing SA tasks is to use sentiment lexicons that associate words with their polarity (e.g., positive or negative). Such lexicons might be compiled manually (e.g., the Harvard General Inquirer [253] and Micro-WNOp [37]) or semiautomatically using some labelled words (e.g., SentiWordNet [8] and MPQA [296]). These lexicons are composed of a relatively small set of words associated with their semantic orientation (valence): either positive, negative, or neutral. Several researchers have generated multidimensional (affective meaning) lexicons that associate words/concepts with three-dimensional real-value vectors of evaluation, potency, activity or with valence, dominance, and arousal. The first *affective meaning* dictionary was introduced by Osgood et al. [191]. To determine the factors that can describe a variety of terms, they asked participants to rate concepts in relation to a large number of scales of bipolar adjectives. They found that three abstract dimensions account for 68% of the variations in the users' ratings: evaluation, potency, and activity.

The Affective Norms for English Words (ANEW) is another multidimensional lexicon developed by Bradley et al. [28], for which they created a set of normative emotional ratings corresponding to valence, arousal, and dominance for collections of objects that elicit emotions, in this case, English words. Each dimension is measured on a nine-point rating scale. Warriner et al. [291] put forward the most recent multidimensional hand-coded corpus, which maps words to a three-dimensional scale of valence, arousal, and dominance, much like the ANEW nine-point scale for measuring each dimension.

Researchers have conducted extensive studies of the semi-automated acquisition of the polarity of words, but it remains an open problem. Much of the work has been targeted at associating words with their prior polarity (semantic orientation) or with their contextual polarity [152, 176, 281]. The work conducted for this thesis was focused on the prior polarity of words, so the following subsections describe the published literature related to prior polarity acquisition. Two primary approaches have been applied for tackling the problem: corpus-based and thesaurus-based methods. Other studies have used graph-based label propagation models for obtaining polarity.

Corpus-Based Approaches

Corpus-based methods rely on fixed seed sets (labelled examples) and co-occurrence statistics from unlabelled corpora in order to induce a sentiment lexicon. Several studies have utilized general co-occurrence features that apply techniques for inducing the semantic orientation of words, such as term frequency-inverse document frequency (TF-IDF) weighting; SVD [275]; a bootstrap mechanism on raw text [212, 111, 266]; the use of syntactic information [94, 117, 216, 200, 294, 41, 262]. Other researchers have gathered and used co-occurrence statistics from the Internet [276, 111, 41] for computing semantic orientation by establishing the similarity of words and phrases based on Internet queries. At the same time, additional studies have involved the examination of domain-specific corpora (e.g., social media) [117, 90]. A number of recent studies have utilized the distributional features obtained from word embeddings to generate semantic lexicons. Astudillo et al. proposed a regression-based model that uses structured skip-gram 600-word embedding to create a Twitter-based sentiment lexicon. Rothe et al. transformed word embedding vectors into a lower dimensional (ultra-dense) representation by training a gradient descent algorithm with two objective functions on lexicon resources [218]. A recent study induced a domain-specific sentiment lexicon using a proposed model, which was based on label propagation utilizing word embeddings created with SVD and PMI [90]. Fast et al. combined both skip-gram word vector features and crowd-sourcing to map words to 200 predefined categories [67].

Thesaurus-Based Approaches

Thesaurus-based methods, on the other hand, use the lexical relationships in WordNet [171], General Inquirer [253], the OpenOffice thesaurus, or other human-coded resources to build sentiment lexicons [109, 124, 259, 64, 208, 258, 229]. Hu et al. [109] used the semantic similarity between adjectives to compute the semantic orientation of a set of words. Kamps et al. [124] developed an algorithm to build a lexical network of synonyms and antonyms, computing sentiment orientation by measuring the relative distance between words and based on seeds such as “good” and “bad”. Esuli et al. [64] proposed a random-walk-based algorithm that uses semantic similarity between the words listed in WordNet. Takamura et al. [259] proposed a model that regards words as electrons and the semantic orientation of words as a spine of electrons. Based on a glossary, lexical network electrons propagate their spine to their neighbours until the system reaches a “minimum energy” configuration. Their model then determines the semantic orientation of the word based on the spine of the electron. Rao et al. [208] presented an extensive evaluation of a dictionary-, corpus-, and graph-based label propagation model that employs different lexical resources to induce the polarity of words.

A recent study conducted by Joseph et al. [119] involved the investigation of how semantic similarity and semantic association affect the way we label others (e.g., student, doctor, etc.). They used Amazon's Mechanical Turk to conduct a survey that asked participants to rate 88 pairs of words drawn from the Simlex-999 dataset created by Hill et al. [103] according to whether these words are equivalent and whether they can be seen together (i.e., in the same context), They compared the participants' annotations to the semantic similarities between words and found that greater semantic similarities and semantic association between words can increase the likelihood that the two identities can be applied to each other.

Graph-based Label Propagation Methods

A number of studies have used *graph-based learning* methods to induce sentiment polarity. Graph-based Label propagation is a transductive approach designed to use a few already labelled nodes in order to label a set of unlabelled nodes. In general, this method employs a fully connected graph in order to encode the similarity between nodes, with the weight edges between nodes representing the relationships between them. The weight could be computed using cosine similarity, a Gaussian kernel, or any other metric. The graph propagates the labels or scores of the known nodes to the unlabelled nodes by repeatedly multiplying the weight matrix (affinity matrix) against the labels or scores vector. Many versions of graph-based label propagation have been proposed based on this general idea [314, 315, 312, 297, 304, 305]. These algorithms follow the main principle underlying a graph-based label propagation algorithm, but each kind of algorithm uses different types of adjacency matrices and clamping assumptions. More details about the label propagation algorithm can be found in section 4.2.

Several studies have used *graph-based models* to induce semantic lexicons. A polarity-propagation or sense-propagation algorithm induces the sentiment polarity of unlabelled words based on some given seed words (positive, negative) and the lexical relationships between them (e.g., Word-net synonym, antonym) [255, 208, 63, 258]. Some researchers have developed a weighted label propagation algorithm that propagates a continuous sentiment score from seed words to lexically related words [80, 110]. Velikovich et al. [282] proposed web-based graph propagation as a means of eliciting polarity lexicons. Their graph is built on a co-occurrence frequency matrix and the cosine similarity (edges) between the words and the seed words (nodes). The model then computes both a positive and a negative polarity magnitude for each node in the graph, each of which is equal to the sum over the maximum weighted path from every seed word (either positive or negative).

Multidimensional Sentiment Lexicon Induction

Only a handful of studies have explored the expansion of multidimensional sentiment lexicons. Kamps et al. [124] use a WordNet-based metric to elicit the semantic orientation of adjectives. They evaluated the lexicon generated against the manually constructed list from the Harvard IV-4 General Inquirer dictionary [253]. The work by Kamps et al. was focused only on adjectives and involved assigning them a binary value (either good or bad, potent or impotent, etc.). They extended a three-dimensional sentiment lexicon using a thesaurus-based label propagation algorithm based upon WordNet similarities [2], and then compared their results against the Ontario dataset [161].

Chapter 3

Fine-Grained Event Extraction from Unstructured English Language Text

3.1 Introduction

Various kinds of events are expressed and discussed in textual documents such as books, news articles, tweets, and blogs. At their simplest, events include any situation that describes an act performed by an actor toward an object, or in other words, any situation that happens or occurs. Extracting events is a task that involves identifying the verb (predicate) and its arguments (e.g., the participants and geolocation information) from textual documents. The information extracted about events can be crucial for a number of natural language processing (NLP) and data mining applications (e.g., text summarization [226], fraud detection, event monitoring/tracking in social media, and question answering applications [51]). The goal of the work presented in this thesis was to extract events and their entities from natural language text in order to build an event-based sentiment analysis (SA) model. This chapter therefore discusses fine-grained event extraction that identifies/detects from English sentences an event (predicate) and its entities, which include the subject, object, modifiers (i.e., adjectives and adverbs that modify the subject, verb, or objects), and settings (i.e., the location where the event took place). Consideration of this sentence from a real world news article [254] provides an example of event extraction:

> “ *The legislation works* ***to make*** *prescription drugs* ***more affordable****,* ***improves access*** *to primary and preventive care and* ***broadens access*** *to benefits for people with Medicare who have low incomes.*”

This example demonstrates the challenge involved in extracting events automatically because the sentence might have multiple events, each of which can be expressed using finite verbs, non-finite verbs, adjectives, or prepositional phrases. The problem of extracting and classifying events from unstructured text has been tackled previously in many linguistics and computational linguistics studies. Two main research streams are related to the problem of extracting events: *coarse-grained* and *fine-grained* event extraction. In *coarse-grained*, the goal of the studies is to identify the events and their attributes, which include predefined event classes as well as temporal and spatial information. For example, using *coarse-grained* extraction approaches, the extracted event and attributes of a sentence such as “*No injuries were reported over the weekend*” using TimeML annotation schemes [231] are EVENT: “reported”, CLASS: REPORTING, TENSE:PAST, TIMEX3: “over the weekend”. Previously proposed *coarse-grained* event extraction methods typically utilize syntactic and lexical information (e.g., WordNet [171] and FrameNet [9]) to extract events [231, 6], or they have employed machine-learning (ML) methods (e.g., conditional random field (CRF) and support vector machine (SVM)) that entail training on a set of engineered features and annotated corpora [21, 157, 280, 1, 185]. Additional details about research efforts directed at extracting events and their attributes can be found in the literature survey provided in [123].

Fine-grained event extraction methods, on the other hand, are intended to extract additional details about an event, such as the participants in the event and their modifiers. A moderate number of studies have addressed the fine-grained event extraction problem, the primary emphasis of which is the extraction from English language text of sentence triplets (i.e., subject, object, and predicate). A sentence triplet is defined as a relationship that links a subject and an object, with the relationship being the predicate. For example, the triplet of “*a woman was helped by a police officer*” is represented as *officer: helped: woman* which is equal to *subject: predicate: object*. Although extracting triplets from sentences is important for many natural language tasks, only a handful of studies have tackled this particular problem. These studies achieve event extraction through the implementation of ML techniques trained on manually annotated data, or they involved the development of rule-based algorithms that utilize syntactic and dependency parsing techniques. All of the proposed fine-grained event extraction approaches, however, are based on the assumption that the sentences are structured grammatically and that each sentence describes only one event (i.e., a single subject, predicate, and object). Most fine-grained approaches have also not been evaluated using real-world datasets.

Related research efforts have focused on the task of semantic role labelling (SRL). SRL systems aim to identify the arguments (i.e., the agent and recipient) of a designated predicate in a given sentence, in other words, link a given predicate to their arguments. Several semantic role labelling systems have been successfully developed by employing machine-learning techniques that utilize a wide range of syntactic and semantic features [204, 79, 270, 205, 303, 78]. Since the focus of this thesis is on extracting sentence triplets, this chapter only discusses research studies that aim to identify the event (predicate) and their participants from English text. More details about the semantic role labelling studies can be found in [193, 36].

This chapter introduces a proposed rule-based fine-grained event extraction algorithm that extracts events and their entities from unstructured English text. The proposed rule-based approach combines lexical resources with syntactic parsing techniques in order to extract multiple events from both grammatical and ungrammatical sentences. The proposal and evaluation of several syntactic-based and dependency-parsing-based methods for extracting events from any textual documents are also described. In particular, this research has resulted in the development and evaluation of three rule-based algorithms. First is a proposed syntactic-based technique that locates event entities by performing a breadth and depth search in the parse tree. The second technique involves the implementation a dependency parser-based algorithm that locates the event entities and their modifiers based on the dependency parsing output. The third proposal is a model that incorporates both semantic and syntactic features to identify and extract fine-grained information about events from English sentences. The proposed algorithms (syntactic-based, dependency-based, and semantic-syntactic) have also been evaluated against state-of-the-art methods for extracting triplets from real-world human-annotated corpora (i.e., news headlines and news articles). The results show that the rule-based algorithms generated results that are comparable to the inter-rater annotation scores that represent agreement between human participants and that are relatively similar to the results obtained using ML-based approaches. The incorporation of semantic features significantly improved the performance of the syntactic-based approach, and the accuracy levels achieved by the semantic-syntactic method were 79% and 82% with respect to extracting the predicate from news headlines and news articles, respectively.

3.2 Background

This section first presents general background information about *coarse-grained*, followed by details related to *fine-grained* event extraction and, more specifically, methods for extracting triplets from English sentences.

3.2.1 Coarse-Grained Event Extraction Methods

Coarse-grained event extraction methods are designed to identify from an input text the event (predicate) and properties related to the event type, verb tense, and temporal and locative information. Previously proposed coarse-grained methods have been based on either the use of ML methods trained on annotated data or the building of a rule-based algorithm that employs syntactic information and semantic lexicons to extract events. Sauri et al. [230] implemented a rule-based algorithm that uses linguistic-based and statistical-based knowledge to extract events. Their proposed algorithm first performs part-of-speech (POS) tagging and lemmatization to determine event candidates and their grammatical features. The algorithm then identifies events and their classes by performing a lexical lookup in WordNet [171]. The Sauri et al. algorithm was evaluated using the TimeBank 1.2 corpus [206], and a corpus annotated with TimeML [231] tags consisting of 300 annotated articles; the evaluation yielded a precision level of 74.03% and a recall success rate of 87.31% for event extraction. Bertha et al. [21] considered the event extraction problem as a classification task. To extract events from an input text, they trained an SVM model on a wide range of engineered features, including affix, morphological, POS, and temporal features. They also evaluated their approach using the TimeBank 1.2 corpus [206] obtaining 82% and 71% precision and recall levels, respectively. In a similar vein, Llorens et al. [157] employed a TempEval-2 dataset [207] for training a CRF model on linguistic and syntactic features and testing the results. The model yielded an 81% precision level and an 86% recall success rate with respect to identifying events from English text. All coarse-grained methods are focused on event extraction and classification (i.e., only the predicates). However, in the research for this thesis, the goal was to perform fine-grained event extraction and to obtain greater detail about the events.

3.2.2 Fine-Grained Event Extraction Methods

Several studies and model implementations have previously been carried out in the area of the extraction of triplets from textual documents. These studies typically use methods based on ML, syntactic parsing, or dependency parsing to extract events and their participants from English sentences.

Machine-Learning Methods

Dali et al. [50] presented an ML approach that identifies sentence triplets in a supervised manner using an SVM model. The proposed method first performs tokenization and data cleaning (removing punctuation and stop words). A list of all possible triplet candidates is then given to the SVM along with their labels, which are either positive (correct triplet) or negative (incorrect triplet). The model was trained on a different set of context-related, POS, and syntactic features so that a positive score would be assigned to the correct triplet candidates and negative scores to other candidates. If multiple candidates have a different object but the same subject and predicate, the algorithm merges the objects that have the highest positive scores. The SVM algorithm was trained and tested on a manually annotated new article corpus of 800 sentences (700 sentences for training and 100 for testing). Using only this limited number of features and labelled data, the model provided a precision level of 38.6% and a recall success rate of 46.80% with respect to the extraction of triplets.

Methods Based on a Syntactic Parser

An understanding of the background information about syntactic-based event extraction methods requires knowledge of the definitions of sentence structure terminology. In English, declarative sentences (i.e., sentences that make a statement) are often composed of a noun phrase (NP) and a verb phrase (VP). The NP usually precedes the VP and represents the subject in the sentence, and the VP represents the verb in the sentence which is referred to as the predicate. The VP might also be followed by a NP that represents the object of the sentence and might be followed by a prepositional phrase (PP) or an adverb phrase (ADVP) [65]. The syntax and meaning of a sentence can be depicted in the form of a syntactic tree made up of a root, nodes, and lines. The root, which is the topmost node in the tree, indicates the type of parsed text (e.g., S - sentence). The other nodes in the tree represent the words or the phrases in the sentence, and the lines indicate containment relationships (i.e., a node is contained in or contains another node).

Several syntactic tree structures are used in the linguistics and computational linguistics fields; however, the standard structure used in many syntactic parsing techniques is the Treebank syntactic structure. Treebank is a parsed text corpus in which each sentence in the corpus is annotated with its syntactic, POS tag, and semantic structure. The Stanford syntactic parser is one of the most widely used parsing methods in the NLP field [129]. This parser is a probabilistic natural language parser that analyzes the grammatical structure of a sentence and generates a Treebank parsed text. The Stanford parser was trained on a substantial number of manually annotated parsed sentences so that it could predict the syntactic structure of a new text. Figure 3.1 depicts the parse tree for the sentence "*A woman was helped by a police officer*". As shown in the figure, the parse tree for the sentence has a sentence root (S) and two children: a noun phrase (NP) and a verb phrase (VP).

To extract sentence triplets using a Treebank-based parsing representation, Rusu et al. proposed a purely syntactic-based approach that performs depth and breadth searches in the Stanford Treebank parser [129] output in order to locate the subject, verb, and object [225]. Their proposed algorithm searches a parse tree and locates the NP and VP in the tree's children. The algorithm then performs a breadth search in the NP to locate the first noun (NN) in the subtree (the subject of the sentence), followed by a depth search in the VP to extract the deepest verb (predicate of the sentence). To extract the object, the algorithm searches for any NP in the VP's children. As an example of how the Rusu et al. algorithm functions, the triplet extracted from the above parsed sentence, shown in Figure 3.1, is (*woman: helped: police*).

Another related study integrated POS tagging with syntactic and lexicon-driven methods to extract relation and events from textual documents. Trained and tested on a news corpus, the model achieved an F-score of 64.75% for event extraction and 75.35% for relation extraction [6].

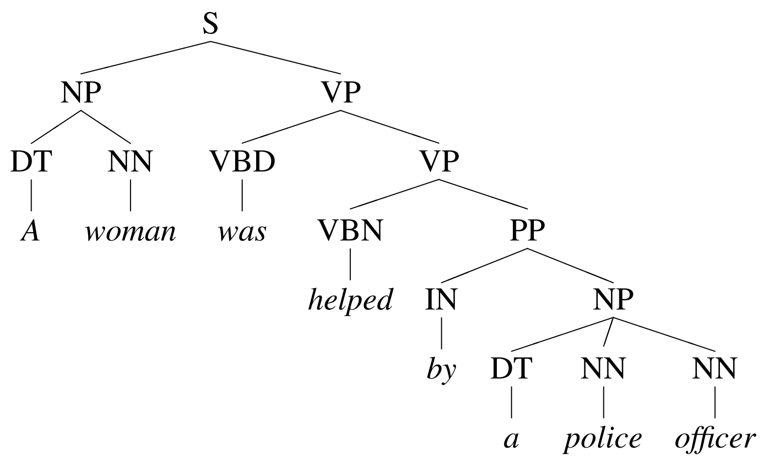


Figure 3.1: Treebank parse tree

Methods Based on Dependency Parsing

Dependency parsing is the task of analyzing the grammatical structure of sentences/phrases and providing a uniform representation that links the "head" words to their modifiers. For example, in a sentence such as "*It prevents infection and diseases*", as shown in Figure 3.2, "*it*" is the head word of " *prevents*" and they are linked by a nominal subject relationship. Several studies have proposed dependency parsing methods, which can be categorized as two types: tree-oriented parsers, and neural-network-based parsers. Nivre et al. [188] proposed MaltParser, a greedy transition-based dependency parser. A first-order graph-based parser was proposed by Mcdonald et al. [164]. Many recent approaches have involved the use of neural networks. For example, Garg et al. proposed transition-based dependency parsers based on a Temporal Restricted Boltzmann Machine (TRBM) [76], and Tenetorp et al. built recursive neural networks for transition-based dependency parsing [250]. More recently, Chen et al. proposed a neural-network-based transition dependency parser that relies on POS tags, dense features that represent words, and arcs as dense vectors. They modelled the interactions among these features using a new cube activation function [39].

Only a few studies have utilized dependency parsing output in order to extract sentence triplets [118, 114, 120]. Jivani et al. [118] presented an approach that utilizes both a Treebank

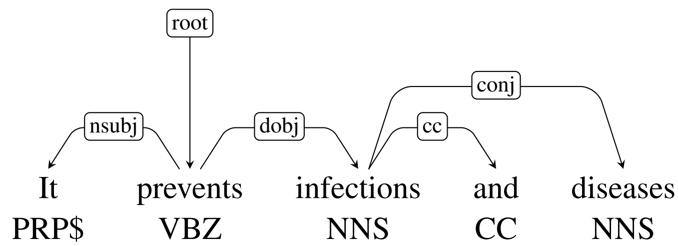


Figure 3.2: Dependency parsing results using the Stanford parser

parser and a dependency parser to extract multiple triplets from English sentences. Their method was targeted at solving the problem of multiple subjects, objects, and predicates in English sentences. Their model searches a syntactic parse tree to extract all of the nouns and verbs in the sentences, and then utilizes the output of the dependency parser to link the subjects and objects according to their relationship (predicate) [39].

3.3 Methods

3.3.1 Event Extraction Based on Syntactic Parsing

To extract events and their entities from unstructured event-based text, this research involved the development of an algorithm that parses a sentence and returns the event quintet (entities). The algorithm begins by dividing the sentence into phrases in an order based on the type of delimiters (e.g., during, while, after, before, comma, and semicolon) and then parses each phrase using the Stanford syntactic parsing method [130]. English language syntactic parsers typically return a parse tree that contains two children: an NP and a VP, as shown in Figure 3.3 and Figure 3.1. The NP often contains the subject of the sentence while the verb and the object are located in the VP. The syntax of many English sentences, however, might be depicted in a different structure or order of phrases. For example, the syntactic parse tree of an ungrammatical sentence might look entirely different from a typical parse tree because it might start with a phrase, a fragment, or a clause root, and the children might not contain a direct noun phrase (NP) or verb phrase (VP) (Figure 3.4).

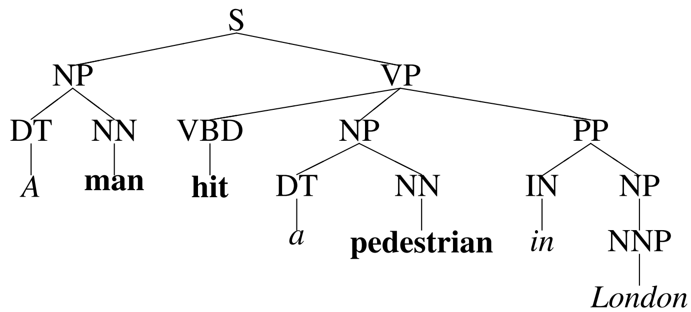


Figure 3.3: A parse tree of a grammatically correct sentence and the extracted triplet in **boldface**

To enable consideration of other English syntactic structures, a linguistic rule-based algorithm inspired by Rusu et al. [225] was implemented. As discussed in section 3.2, the Rusu et al. algorithm is based on the assumption that the syntactic structure of a parsed sentence follows the standard structure of English language sentences (Figure 3.3). The algorithm will thus succeed in extracting triplets only from sentences that follow a standard syntactic structure and will fail to extract triplets from sentences with different syntactic structures. For example, the parse tree for the news headline indicated in Figure 3.4 shows that the first VP in the tree does not have a verb within its children and that the actual predicate is located in its descendant.

To overcome this problematic structure and to include consideration of ungrammatically structured sentences, the new proposed algorithm performs a depth search (also known as depth first traversal) for the NP and VP in any type of roots, not just the one that starts with a sentence root (S.). The algorithm performs a depth search from the root to locate the first NP and deepest verb phrase VP. After extracting the NP and VP, the algorithm performs a breadth search within the NP in order to locate the subject (S) (the last noun (NN), proper noun (NNP), or adjective (JJ) in the NP) and the subject modifiers (SM) (all of the adjacent adjectives (JJ), nouns (NN), or proper nouns (NNP)). If the NP contains another NP, the algorithm uses the second NP to extract the subject. To extract the predicate or the verb, the algorithm performs a depth search in the VP, returns the deepest VP, and then performs a breadth search to extract the adjacent adverbs (RB) (verb modifiers). The algorithm next performs a depth search in the VP to locate the object (O), followed by a search for the object modifiers (OM) in the adjective (JJ) or nouns (NN) adjacent to the object. To extract locations from sentences, the algorithm also searches the VP for any

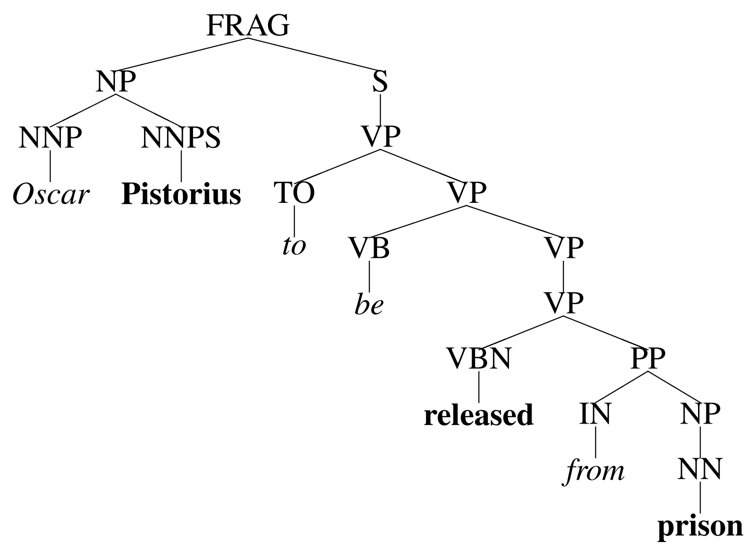


Figure 3.4: Parse tree for a grammatically correct sentence, with the extracted triplet in **boldface**

prepositional phrases (PP) and return the nouns of the NP as the location (L) and any adjacent adjectives as modifiers of the location (LM). The algorithm looks for all the prepositions that designate places (e.g., above, across, on top of, on). If the sentence is in a passive voice format, the algorithm swaps the object and subject, as detailed in Algorithm 1. For example, the quintet extracted from the headline example in Figure 3.4 is *Blank: released: Pistorius: prison:Blank* which is *subject: predicate: object: setting: modifiers*.

The proposed syntactic-based event extraction algorithm presented in this section differs from the Rusu et al. approach in the following respects: (1) It searches for the deepest VP in the tree to locate the predicate not the first VP in the tree; (2) It considers the last child in the NP as the subject and all prior siblings as subject modifiers whereas the Rusu et al. algorithm extracts the first noun in the NP as the subject; (3) It considers different POS tags (i.e., proper nouns, nouns, and adjectives) when searching for the subject of a sentence, while the Rusu et al. algorithm looks only for nouns; (4) It performs a deep search in order to locate the object and the location rather than searching only the first children of the VP subtree; (5) It considers the passive voiced sentence when extracting the subject and object; (6) It extracts multiple modifiers of the subject, verb, and object; (7) It extracts the locations (settings) of the event and their modifiers; (8)

It considers the possibility of multiple performers of an event by looking for a coordinating conjunction (cc) tag in the subject or object subtree.

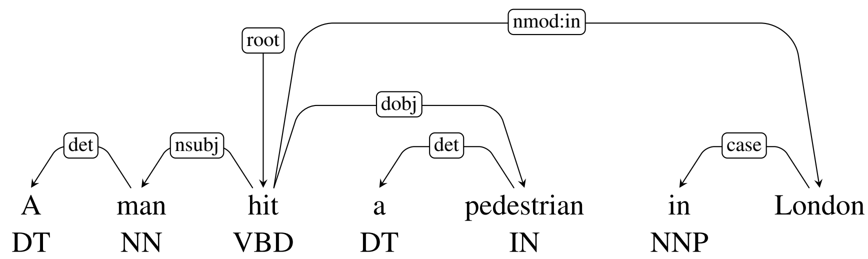
Algorithm 1 Quintet parsing and extractor

```
function QUINTET-EXTRACTION(Treebank parse tree: tree)
    NP= first NP in tree's children
    VP= deepest VP in tree's children
    Event(S)=extract-Noun(NP)
    Event(V), Event(O), Event(L)=extract-verb(VP)
     return Event
end function

function EXTRACT-NOUN(Parse tree NP)
    N =The last child in NP of type (NN.,NNP., JJ.)
    NM= siblings of Noun of type (JJ., NN.)
    if NP contains coordinating conjunction (CC.) then
        NP= search for the next NP.
        add another noun=Extract-Noun(NP)
    end if
     return (N, NM)
end function

function EXTRACT-VERB(Parse tree VP)
    V =Perform a breadth search on the VP and return the last verb
    VM= Extract all siblings of V of type (RB., RBR., RBS.)
    NP= Perform a depth search for NP on the VP
    O, OM=Extract-Noun(NP)
    PP= preposition phrase in VP
    if PP's children ∈ {from, of, in, at, on, to} then
        NP= perform a breadth search and locate the first NP
        L, LM=Extract-Noun(NP)
    end ifreturn (V,VM), (O,OM),(L,LM)
end function
```

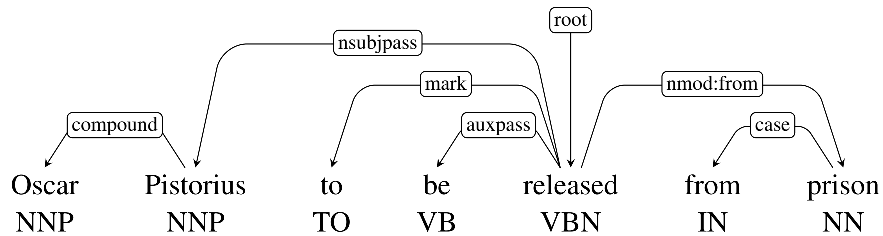
3.3.2 Event Extraction Based on Dependency Parsing

The dependency parsing method is very simple and straightforward, relying heavily on the precision of parsing techniques and the POS tagger. It first parses a sentence and returns the output of the dependency parser and POS tagger. The Stanford dependency parser [39] and POS tagger [271] were used in this algorithm. Based on the output of the POS tagger and the dependency parser, the new approach presented here searches first for all dependency relationships that indicate object-verb (e.g., "*obj*"= object, and " *dobj*"=direct object) or subject-verb relationships (e.g., " *nsubj*"=nominal subject, and "*subj*"=subject) in order to extract the predicate (verb) of the sentence. If the POS tag of the extracted predicate indicates that it is a verb, the algorithm extracts the associated object or subject. The setting is then extracted from the dependency relationship of a prepositional clause "*prepc*" and the modifiers of any associated subject, predicate, or object are extracted from any dependency triplets of the type modifier (e.g.,"*amod*"=adjectival modifier). The algorithm also considers sentences that are in the passive voice. Algorithm 2 provides an abstract of the algorithm.

An example of how this rule-based approach performs on both grammatically and ungrammatically structured sentences is given in Figure 3.5, which shows the dependency representations for the same examples referred to in the previous section. In Figure 3.5a, the dependency parse tree indicates that there is one nominal subject (*nsubj*) dependency relationship that links the verb "*hit*" and "*man*", a direct object (*dobj*) relationship that links "*hit*" with "*president*", and a prepositional modifier (*nmod:in*) that links "*hit*" with "*London*". Based on this representation, the extracted quintet for "*A man hit a pedestrian in London*" is *man: hit: pedestrian: London: BLANK* while the extracted quintet of the dependency tree shown in Figure 3.5b is *BLANK: released: Pistorius: prison: BLANK*. The use of this simple and straightforward approach has shown promising results for different grammatical structures, but it failed to recognize and detect events expressed in terms of a noun, adjectives, or multi-term expressions. For example, in a sentence such as "*Guyana and US in dispute over democracy project*," the algorithm will fail to extract the predicate " dispute" because the dependency parser output will contain neither a subject-verb nor a verb-object relationship. As described in the next section, this particular limitation was addressed through the incorporation of semantic knowledge.



(a) Dependency parsing tree for a grammatical sentence



(b) Dependency parsing tree for a news headline

Figure 3.5: Dependency parsing results using the Stanford parser

Algorithm 2 Dependency-parsing-based quintet extractions

function DEPENDENCY-QUINTET-EXTRACTION(dependency output : $output$, POS tagger output :pos)
  w_1,w_2,$relationship$ =search $output$ and return dependency triplet of object or subject relationship
  if the pos of $w_2 \in$ {VB,VBG,VBD,VBN} **then**
    if $relationship$ of type passive subject or object **then**
      $Event(O) \leftarrow w_1, Event(V) \leftarrow w_2$
    else $Event(S) \leftarrow w_1, Event(V) \leftarrow w_2$
    end if
  end if
  w_1,w_2,$relationship$ =search $output$ and return dependency triplet of type prepositional modifier
  if w_1 ==V **then**
    $Event(L) \leftarrow w_2$
  end if
  w_1,w_2,$relationship$ =search $output$ and return dependency triplet of type modifier
  if $w_1 \in \{V, S, O, L\}$ **then**
    $Modifiers(w_1) \leftarrow w_2$
  end if
  return $Event$, $Modifiers$
end function

3.3.3 Semantic and Syntactic Event Extraction

The previously described algorithms are characterized by a number of limitations that are due to the complex linguistic structure of event-based text and the shortcomings of syntactic-based approaches for extracting events from event-denoting expressions. For example, in a sentence such as "*Several anti-war* **protests** *were held in Spain*" the main event is expressed in terms of the nominalization expression "**protests**" while the above algorithms will extract "**held**" as the predicate of the sentence. This section presents the proposed approach developed for this thesis, which takes into account all types of event-denoting expressions through the incorporation of lexical resources and linguistic cues for the extraction of predicates and their arguments from English text. The new algorithm focuses first on identifying all possible predicates, based on consideration of event-denoting expressions, followed by recognition of other arguments related to the event (i.e., participants and modifiers), based on the extracted predicates. Linguistic evidence demonstrates that events (predicate) in English language sentences can be expressed in terms of a wide range of syntactic expressions, such as finite verbs (e.g., *was managed*), noninfinite verbs (e.g., *to investigate*), noun (e.g., *war*), adjectives (e.g., *dead, pregnant*), nominalizations (e.g., *demonstration*), predicative clauses (e.g., *be prepared*), or prepositional phrases (e.g., *on board*).

The extraction of events from text with simultaneous consideration of event-referring expressions first involved POS tagging and syntactic parsing to extract all of the predicate candidates. The only focus at this stage was on all events that are expressed in terms of verb, noun, and adjectives. The algorithm then performs a lexical lookup to filter out all candidates that are not identified as an event in the lexical resources. Next, in a manner similar to that of the syntactic-based approach described above, the algorithm extracts the event arguments based on the output from the syntactic parser. Using the syntactic parse tree, the algorithm locates the subtrees that contain the predicates and then searches each subtree for any object, modifiers, and location, followed by a search of the preceding NP for any subject and adjectives (modifiers). The POS tagging and syntactic parsing were achieved through the implementation of Stanford POS tagging [271] and the Stanford syntactic parser [39], respectively, while the lexical lookup was performed using a WordNet lexical resource [171]. Employing semantic parsing for the extraction of events (predicates) from English sentences is a technique that has been pursued previously [230]; however, this algorithm integrates semantic parsing with a syntactic-based algorithm in order to extract

```
(ROOT
  (NP
    (NP
      (NP (NNP Guyana) (CC and) (NNP US))
      (PP (IN in) (NP (NN dispute))))
      (PP (IN over) (NP (NN democracy) (NN project)))))
```

```
('conj_and', 'Guyana', 'US'),
('prep_in', 'Guyana', 'dispute')
('nn', 'project', 'democracy')
('prep_over', 'Guyana', 'project')
```

Figure 3.6: Output of the syntactic and dependency parsers for *Guyana and US in dispute over democracy project*

events and their entities (quintet). Algorithm 3 provides an abstract of the semantic and syntactic extraction method.

To illustrate the above procedure with an example, Figure 3.6 shows the POS tags and syntactic sentence structure as well as the dependency parser output for "*Guyana and US in dispute over democracy project*". As can be seen, the dependency- and syntactic-based approaches fail to identify the event (predicate) in the sentence since the syntactic parse tree has no VP and the dependency output contains no subject-verb or object-verb relationship. Applying the semantic and syntactic extraction method enables the extraction of the event candidates (i.e., all the adjectives, verbs, and nouns in the sentence), which are "*dispute, democracy, and project*". All of the non-event predicates are then filtered out from the event candidates, which gives "*dispute*". Next, the relevant arguments are extracted based on the location of the VP in the syntactic parse tree and a subsequent search for the subject in the preceding NP and the object in the following NP. This semantic and syntactic rule-based approach accurately extracts the event and its arguments from this example which is *Guyana and US: dispute: democracy project*.

Algorithm 3 Semantic and syntactic-based event extraction

function PREDICATE-EXTRACTION(syntactic parse tree: $tree$, dependency output parse tree: $output$, POS tagger output : pos)
    for $e \in pos$ **do**
        if e of type V*, N*, or JJ* **then**
            $PredicateCandidates \leftarrow e$
        end if
    end for
    for $c \in PredicateCandidates$ **do**
        if c not of type events in $lexicalResource$ **then**
            remove c from $PredicateCandidates$
        end if
    end for
    for $c \in PredicateCandidates$ **do**
        $Event$, $Modifiers$=Argument-extraction($output$,c)
    end for
    return $Event$, $Modifiers$
end function
function ARGUMENT-EXTRACTION(syntactic parsing output: $tree$, $PredicateCandidate$)
    CS=locate-candidates-subtree ($tree$)
    NP=locate-noun-phrase-subtree ($tree$)
    $Event(S)$, $Modifiers(SM)$=Extract-Noun(NP)
    $Event(O)$, $Modifiers(OM)$=Extract-Noun(CS)
    PP=locate-prepositional-phrase-subtree (CS)
    $Event(L)$, $Modifiers(LM)$=Extract-Noun(PP)
    return $Event$, $Modifiers$
end function

3.4 Dataset

The three algorithms described in section 3.3 were evaluated with respect to extracting events along with their entities from grammatical and ungrammatical sentences. Two real-world datasets were used for the evaluation: a news headline dataset and a news article corpus. The news headline dataset consisted of 1,658 news headlines collected from different newswire resources. It was manually annotated through Mechanical Turk, with participants being asked to choose a subject, a predicate, and an object for each headline from a list of all the words in a sentence. Additional choices were *implicit* and *blank*. *Implicit* indicates that the subject, verb, or object are implied and not expressed directly in the sentence, while *blank* means that the sentence is lacking an explicit or an implicit subject, verb, or object. Three participants annotated each headline, and the agreement among the annotators was considered to be substantial to moderate (Table 3.2). The average length of the all of the news headlines is eight words. After the dataset was cleaned and any unrelated entries removed, all of the annotations provided were considered. For example, if one participant annotated the subject of a headline stating that "*Super Bowl-winning quarterback Russell Wilson divorces wife*" as "*Russell* " and another participant chose "*Wilson*", both "*Russell*" and "*Wilson*" were deemed to be the subject of the sentence, as in *Russell Wilson: divorces:wife*. Further details about the dataset can be found section 5.4.

The proposed methods developed for this thesis were also evaluated using grammatically correct sentences from a news article corpus. A Multi-Perspective Question Answering (MPQA) dataset, the news article dataset [254], contains 134 news articles, including a total of 8,069 sentences from political articles that have been manually annotated for SA purposes. This corpus also includes annotations for the agent (the subject), the behaviour, and the object in each of the sentences that were of interest for this research. Some of the annotations, however, might be a phrase or a clause. The testing of the proposed algorithms was therefore conducted only on examples that have no more than three words in the subject, verb, or object. This constraint reduced the size of the dataset to 1,106 sentences whose average length was 24 words. For example, the annotated version of the sentence "*Our reform will prohibit insurance companies from denying coverage because of your medical history* " is *reform: prohibit: insurance companies*, with both " *insurance*" and " *companies*" to be considered accurate annotations of the object. Details about the corpus and the annotation scheme are provided in section 6.3.

3.5 Evaluation

The effectiveness of the new proposed syntactic- and semantic-based algorithm was compared against some of the state-of-the-art methods, a rule-based [225] and an ML-based [50] approach, as well as a syntactic- and a dependency-based approaches that were examined earlier in this

Event Extraction Algorithm	Subject	Predicate	Object	Triplet
Semantic and syntactic-based method	**85**	**79**	**57**	**48**
SVM-based method by Dali et al. [50]	–	–	–	38
Syntactic-based method	83	49	55	36
Dependency-based method	70	70	41	30
Method by Rusu et al. [225]	64	48	21	15

(a) Results for the news headline dataset

Event Extraction Algorithm	Subject	Predicate	Object	Triplet
Semantic and syntactic-based method	**59**	**82**	**66**	**43**
SVM-based method by Dali et al. [50]	–	–	–	32
Dependency-based method	47	42	38	29
Syntactic-based method	31	41	38	10
Method by Rusu et al. [225]	13	2.5	2.8	0.6

(b) Results for the news article dataset

Table 3.1: The percentage accuracy of subject, verb, object sentence triplets extracted from news headline and news article datasets using different rule-based algorithms against the results obtained from state-of-the-art rule-based and ML methods, with results in ascending order (bottom to top) according to performance, and the highest scores in **boldface**

research. The algorithms were evaluated for extracting sentence triplets (subject, predicate, and object) from two real-world datasets. The same syntactic and dependency parsing techniques as well as POS tagging methods were used in all the algorithms. The same parsing techniques were also employed for the reimplementation of the Rusu et al. approach [225]. The ML method [50] was also trained and tested on the same datasets and using the Stanford POS tagger. As with the model proposed by Dali et al., the SVM model was trained on context-based and POS features (i.e., the word, first word on the right, second-next word on the right, first word on the left, the three last letters of the verb, POS tag for the word). The datasets were divided into a training-testing proportion that is similar to the one used by Dali et al.: 80% for training and 20% for testing. The SVM was trained on $1,325$ and tested on 333 examples from the news headlines,

Event Extraction Algorithm	Subject	Predicate	Object
Kappa score P1 and P2	0.65	0.55	0.36
Kappa score P1 and P3	0.62	0.60	0.39
Kappa score P2 and P3	0.69	0.57	0.36

Table 3.2: Results of the Kappa Cohen inter-rater agreement among the three judges (P1, P2, P3) of the news headlines dataset

and trained on 884 and tested on 222 examples from the news article dataset. The performance of the proposed models was evaluated based on computations of the accuracy: the number of correctly extracted words divided by the total number of examples in the test set. For the rule-based algorithms, the accuracy of the prediction of each item in the triplet was calculated as well as that of the extraction of the three elements together. The ML method, however, was evaluated with respect to the identification of all of the items in the triplet together.

3.6 Results and Discussion

As shown in Table 3.1b, the results for all the methods evaluated were comparable and consistent with respect to both the inter-rater agreement among the human annotators and the results produced by the ML approach. The inter-rater agreement values listed in Table 3.2, are ~ 0.65, 0.60, and 0.39 for the extraction of the subject, predicate, and object, respectively. This level of agreement is considered to be substantial for extracting the subject, moderate for identifying the predicate, and fair for determining the object. The same pattern is also observable for the extraction of the subject, predicate, and object using the rule-based methods, for which the accuracy of the subject extraction was the highest, followed by the accuracy of the predicate identification and then the object extraction from short and ungrammatically structured sentences. The results also demonstrate that the accuracy level of event extraction from short sentences (news headlines) was generally higher than from long sentences (sentences from news articles) as most of the algorithms performed better on the news headline dataset.

The linguistically motivated rule-based methods, in general, produced competitive results compared with the ML approach, which is trained on a huge number of annotated corpora. The results reveal that the lowest accuracy level for both datasets was obtained using the Rusu et al. algorithm, and that the syntactic-based method performed better on both the grammatical and ungrammatical text. These findings demonstrate that just including consideration of different syntactic sentence structures can drastically improve the performance of the syntactic-based approach. The performance of the methods based on dependency parsing was better than that of the syntactic-based approach for grammatically correct sentences. The results also indicate that the integration of semantic features with the syntactic-based model significantly enhanced the performance of the syntactic-based approach, especially with regard to extracting the predicate, as the semantic and syntactic-based approach outperformed the ML and other rule-based approaches.

3.7 Conclusion

Researchers have investigated and proposed a moderate number of methods for recognizing events and their arguments, such as the participants in the event and their modifiers. Most of the studies conducted in this area have led to the proposal of techniques for extracting events that either use ML approaches or use linguistic rule-based methods that utilize the semantic or the syntactic structure of sentences. In this thesis, a structure-based SA system is proposed. The proposed system uses ACT to model human emotions arising from different entities in a text. To build a structure-based model and to predict human emotions resulting from different items in an input text, several syntactic-based rule-based methods for event extraction were implemented and evaluated, and a linguistically motivated rule-based approach has been proposed that incorporates semantic- and syntactic-based features for extracting events from a piece of text. The new approach is the first one that includes consideration of both grammatically and ungrammatically structured sentences, uses both semantic and syntactic parsing features to obtain fine-grained events, and extracts sentence quintets (entities): the subject, predicate, object, settings, and modifiers.

Due to the lack of an annotated corpus that contains all of the entities, the proposed approaches have been evaluated with respect to their success in extracting triplets (subject, verb, an object). The proposed algorithms were assessed against state-of-the-art approaches (rule-based and ML approaches) with regard to two real-world types of English language text having different grammatical structures. In general, the results demonstrate the considerable challenges associated with the extraction of events and their arguments from unstructured English text because successful completion of this task requires a syntactic and semantic understanding of words and their role in a sentence. The highest overall level of accuracy achieved for triplet identification was between 43% and 48%.

The inter-rater agreement and accuracy of the methods show that recognizing the predicate and the subject of a sentence is simpler than extracting the object. The results of the evaluation also demonstrate that extracting events from short sentences such as news headlines (eight words, on average) is, in general, easier than from longer sentences (24 words, on average). With respect to the extraction of triplets from both grammatical and ungrammatical text, the best performance was obtained with the use of the semantic- and syntactic-based rule-based algorithm, while the worst was produced by the syntactic-based approach proposed by Rusu et al. [225].

The ML approach performed better than some of the rule-based approaches. However, it was highly unbalanced since the number of positive examples is very small, and it also required significant running time because the training set contains all possible triplet candidates for each sentence. For these reasons, the SVM was implemented with the number of candidates limited to half of all possible candidates, and the algorithm, similar to [50], was trained on a huge amount of annotated data (80% of the data). The results for both datasets revealed that the semantic-syntactic approach significantly outperformed the ML and the other rule-based methods.

Chapter 4

Semi-Supervised Affective Meaning Induction Expansion Using Semantic and Distributed Word Representations

4.1 Introduction

Most sentiment analysis (SA) systems rely heavily on sentiment lexicons that play an important role in improving their performance. Sentiment lexicons, or sentiment dictionaries, typically map words or phrases to a rating that indicates how individuals perceive these words/terms emotionally. The model proposed in this thesis relies on affect control theory (ACT) and its associated dictionaries that were originally composed of a relatively small number of words associated with three-dimensional evaluation, potency, and activity (EPA) scores. To augment the human-coded lexicon, graph-based learning algorithms built on semantic, distributional, and distributed word representations were implemented. Extensive evaluation of a variety of unsupervised and supervised learning algorithms for expanding and regenerating three-dimensional sentiment lexicons was also conducted.

Semi-supervised learning algorithms are advantageous for many classification tasks as they learn from both unlabelled and labelled data and require only a small set of labelled instances to enable them to label the rest of the data points. Semi-supervised learning algorithms are particularly valuable for lexicon expansion since they permit the labelling of a set of given words based on their semantic and distributed similarities with other words and also enable the re-estimation of manually labelled words (seed words) given the structure of a similarity graph. The semi-supervised algorithm presented in this thesis was implemented using two different graph structures: a simple, or single, graph that has at most only one edge between its vertices and a multigraph that has parallel edges between its vertices, with each edge representing a feature from a different modality. The single-graph semi-supervised approach was implemented using semantic (dictionary-based) word features, count-based word embeddings, neural network word embeddings, and a linear combination of both neural and semantic word embeddings. The multigraph semi-supervised learning approach was constructed and implemented using semantic and neural word embeddings. The label propagation algorithms were trained on both pre-trained word embeddings and on word embeddings that the author obtained by using a Wikipedia dataset to train neural network models and a count-based model (singular value decomposition).

The results of a comparison of the induced sentiment ratings with their corresponding manually annotated values reveal that combining both semantic and distributed word embeddings generated the highest-ranked correlation score of $\tau = 0.51$ with respect to recreating two sentiment lexicons: the one by Warriner et al. [291] and the General Inquirer [253]. The results also show that of the three dimensions, evaluation (E) was associated with the highest ranked correlation score, while the lowest score was for potency (P). The results also demonstrate no significant difference between the pre-training fusion approach that combines multimodal features linearly and an approach that performs the fusion simultaneously during the learning algorithm. The results using the trained word embeddings were consistent with the results obtained using pre-trained word embeddings, with the best performance being obtained using skip-gram negative-sampling (SGNS) word embeddings, followed by Global Vectors for Word Representation (GloVe) and count-based word embeddings.c

To compare the results of the semi-supervised learning algorithm with some of the standard unsupervised and supervised learning methods, an unsupervised learning algorithm that computes the EPA scores of the current words in the affective meaning dictionary [274] was implemented. The unsupervised learning algorithm uses word co-occurrence statistics aggregated from search engine results in order to estimate the sentiment orientation of a given word by computing the difference between the strength of its associations with positive paradigm words (e.g., good, powerful, active) and with negative paradigm words (e.g., bad, powerless, quiet).

Compared with the other algorithms, the unsupervised learning approach yielded the lowest accuracy level and the lowest-ranked correlation scores. A supervised learning algorithm was also implemented: a support vector regression (SVR) model trained on skip-gram word embedding and the full training set (one-third of the dictionary). The results from the supervised learning algorithm were slightly better than those obtained with the semi-supervised and unsupervised learning algorithms, yet the supervised learning algorithm does not generate the lexicon independently from the labelled words. The EPA scores induced for this study were also evaluated against some of the state-of-the-art methods used for lexicon expansion, and the new method demonstrated improvements in the τ correlation and F1-score compared to the currently used algorithms.

This thesis presents the first extensive evaluation of semi-supervised graph-based learning methods for multidimensional lexicon expansion. The research involved the computation of evaluation, potency, and activity (e.g., valence, dominance, and arousal) scores rather than only an evaluative factor (i.e., valence), the development of a proposed novel approach to the linear combination of semantic and distributed word representation in label propagation algorithms, the comprehensive evaluation of four algorithms against a manually annotated dataset as well as supervised and unsupervised learning algorithms, the sampling of seed words from the corpus or dictionary instead of the common use of fixed seed words (e.g., good, bad, happy, sad etc.), and the creation of a significantly large three-dimensional lexicon of $\sim$3words that can be leveraged by researchers in the SA and social science fields. The approach proposed in this thesis will reduce the cost of manually annotating sentiment lexicons, help integrate the affective meaning of today's growing vocabulary, and assist with the estimation of variations in attitudes toward words (e.g., same-sex marriage, abortion).

4.2 Background

Graph-based semi-supervised learning methods are transductive models that use the geometry of the data to spread the label from the labelled node to the remainder of the data. The key to semi-supervised or transductive problems is the prior consistency assumption (i.e., the smoothness assumption or the cluster assumption) which posits that similar points (nearby points) or points that exist in the same structure (cluster or manifold) will have a similar label. The primary advantage of transductive semi-supervised learning methods over an inductive supervised learning algorithm is that they utilize unlabelled instances to improve the decision boundaries of the model. Semi-supervised learning algorithms have also been found to be parallelizable, scalable to large data, and very useful in practice, especially for problems in which the data are represented in the form of a graph (e.g., social networks, websites, and semantic lexicons). Transductive semi-supervised algorithms, however, cannot be generalized to unseen data, a process that requires the algorithm to be reimplemented for labelling any new instances.

To find graph labelling that is consistent with both labelled instances and the geometry of the dataset, graph-based semi-supervised algorithms are executed in two stages: graph construction and label inference. The graph construction stage has a significant effect on the performance of graph-based learning methods and is often performed using a neighbourhood method (e.g., k-nearest neighbour graph construction methods that add edges between an instance and its k-nearest neighbours [286]) or using metric learning methods (e.g., IDML [57]). The label inference stage, on the other hand, refers to the task that performs inference on the graph in order to assign labels to unlabelled nodes. Researchers have used either iterative algorithms for labelling unlabelled instances (e.g., label propagation [314, 315]; label spreading [312]; modified adsorption [260]; measure propagation [257]; and sparse label propagation [54]), or they use graph regularization methods (e.g., manifold regularization [16]).

Based on the consistency assumption, proposals have included a number of label propagation algorithms that rely on the same idea of building a similarity graph with labelled (seed/paradigm words) and unlabelled nodes (words). The labels or scores of the known nodes (words) are then propagated through the graph to the unlabelled nodes by repeatedly multiplying the weight matrix (affinity matrix) by the labels or scores vector. Proposed by Zhu et al. [314, 315], the first label propagation algorithm was given a set of labelled $L = (X_l, Y_l)$ and unlabelled $U = (X_u, Y_u)$ data points, where x_i is a data point or feature vector, and y_i is the label of the labelled nodes or zeros. The algorithm build an affinity matrix W by computing a Gaussian kernel $w_{ij} = e^{-||x_i - x_j||^2/2\sigma^2}$ and computes the random walk normalized Laplacian matrix $\Delta = D^{-1}W$ where D is the degree matrix.

The algorithm then propagates the label to the neighbour nodes by computing $\hat{Y} \leftarrow \Delta Y$ and clamps the labelled nodes Y_l to their initial values L after each iteration, since the values of the labelled instances $\hat{Y}_l$ are constrained to be equal to Y_l (their initial values).

The label propagation algorithm proposed by Zhou et al. [312] uses the same idea, but they relaxed the clamping assumption by using a clamping factor α which takes a value between 0 and 1 and uses the normalized graph Laplacian $\mathscr{L} = D^{-1/2} W D^{-1/2}$ instead of Δ. The algorithm then propagates the label to neighbour nodes by computing $\hat{Y} \leftarrow \alpha \mathscr{L} Y + (1-\alpha) \mathscr{L} Y^0$, where Y^0 represents the initial label and the unlabelled scores. Several other graph-based learning algorithms have been proposed [304, 305]; however, Zhu et al.'s [314] and Zhou et al.'s [312] algorithms are the most widely used label propagation methods, especially in lexicon induction research. In the work for this thesis, the Zhu et al. [314, 315] algorithm was implemented and different clamping assumptions were investigated.

The graph-based learning methods mentioned above are used for a single-graph structure in which the features or data points are from one modality. Many studies have demonstrated that, compared with using only one modality, employing multiple modalities (e.g., textual and visual) can provide an enhanced understanding of the feature space and lead to superior learning model performance. To deal with a multimodality learning problem in supervised or unsupervised learning methods, researchers have adopted two different fusion strategies: pre-training feature-level fusion, in which features from different modalities are concatenated into a single feature vector [136], or or post-training fusion, in which a linear combination, min-max aggregation, or voting strategy is performed at the output level [59] [75].

Only a few studies have addressed the problem of multigraph label propagation or the multimodality graph-based learning problem in which the graph has multiple edges between its vertices, with each edge coming from a different modality. To handle a multigraph structure in graph-based semi-supervised methods, researchers have either performed a pre-propagation fusion method with the goal of converting the multigraph into a single graph by linearly combining the weight of its multiple edges [252], or they have adopted a post-propagation fusion approach in which a confidence voting strategy is performed at the output level after each graph has been trained separately [252]. Other researchers have proposed a regularization framework for learning from two or more graphs generated from different modalities [269, 288]. Tong et al. [269] proposed two optimization strategies that combine the features of two graphs based on two different fusion schemes: linear form and sequential form.

The work presented in this thesis involved the implementation of both single-graph and multigraph label propagation using different semantic, distributional, and distributed word embeddings. Background related to semantic and distributional word embeddings and sentiment lexicon expansion is provided in Chapter 2.

4.3 Method

Previous attempts to expand sentiment lexicons using graph-based semi-supervised algorithms have achieved greater accuracy than other standard methods [109, 5, 208]. To evaluate the effectiveness of semi-supervised graph-based approaches with respect to expanding multidimensional sentiment lexicons, this thesis work included the implementation of a label propagation algorithm [314, 312] combined with four methods for computing the affinity matrix using different word representations and word similarity metrics. Also implemented was a multigraph label propagation [269] technique, an approach that extends a single-graph label propagation algorithm [312] for multimodality learning problems, with the goal of inducing a sentiment lexicon using different word embeddings.

4.3.1 Single-Graph Label Propagation

The lexicons used in affect control theory (ACT) consist of a group of concepts (words) associated with EPA vectors. The task of augmenting ACT lexicons given a set of seed words was formulated as follows: Given a set of words $X = \{x_1, \ldots, x_l, x_{\ell+1}, \ldots, x_n\}$ each represented in terms of embeddings $Z = \{z_1, \ldots, z_\ell, z_{\ell+1}, \ldots, z_n\} \subset \mathbb{R}^m$, where m is the dimension of each word vector; $x_l (l \leq \ell)$ is the set of the labelled words; and $x_u (\ell + 1 \leq u \leq n)$ are the unlabelled words, where $n = |V|$ is all the words in vocabulary V. Let Y be a three-dimensional label set $Y = \{y_1, \ldots, y_\ell, y_{\ell+1}, \ldots, y_n\} \in \mathbb{R}^3$, where $-4.3 \leq y_{l,i} (l \leq \ell) \leq +4.3$ and $y_{u,i} (\ell + 1 \leq u \leq n) = 0$. The goal is to define a function that estimates Y_u (the labels of the unlabelled words) from $z_i \in z_{i=1}^n$ (the labelled and unlabelled data points) and $y_i \in y_{i=1}^\ell$ (the labelled words), where $|U| + |L| = |V|$ and V represents all the words in the vocabulary set.

To label the remainder of the unlabelled instances, the label propagation algorithm first constructs an undirected weighted graph $G = \{E, V\}$, where V is a set of vertices in the graph and E is a set of weighted edges. The weight of the edges is represented by a weight matrix, or affinity matrix, W, which is an $n \times n$ symmetric positive semi-definite matrix (where $w_{ij} \geq 0$). The graph adjacency, or affinity matrix, encodes the similarity between the vertices/nodes (words), which are often computed using distance metrics (e.g., Gaussian kernel, point wise mutual information, cosine similarity, or Euclidean distance).

The labels then propagate the label to adjacent nodes by computing $Y \leftarrow \Delta Y$, where Δ is the random walk normalized Laplacian matrix $\Delta = D^{-1}W$, where is the degree matrix $D_{ii} = \sum_j w_{ij}$ and $y^n \leftarrow y^{n-1}..y^1$. After each iteration, the labelled nodes Y_l will clamp to their initial value with clamping factor α, which is a trade-off parameter between consistency with the initial labels (hard clamping $\alpha = 1$) and consistency with the graph (relaxed clamping $\alpha = 0.8$), as follows:

$$Y(t+1) = \Delta \alpha Y(t) + (1-\alpha)Y^0 \tag{4.1}$$

This procedure is repeated until Y converges, and the label Y_u with the highest score is chosen [314]. Using the graph smoothness assumption, researchers have further simplified the semi-supervised label propagation transductive approach into an inductive approach by minimizing the quadratic energy function to find a labelling that is consistent with the initial values of the seed data points as well the geometry of the data. The inductive closed-form solution was found to be equivalent to the iterative algorithm in Equation 4.1, but it reduces the complexity of the iterative algorithm from $O(kn^2)$, to $O(n)$, where n is the number of data points (i.e., nodes in the graph or the vocabulary), and k represents the neighbours of each data point. More details about the inductive approaches can be found in [18].

4.3.2 Multi-Graph Label Propagation

To formulate the multi-graph or multi-modality learning problem, the same notations used in section 4.3.1 were followed. Given a set of words $X = \{x_1, \ldots, x_\ell x_{\ell+1}, \ldots, x_n\}$, each represented in terms of two word embeddings $Z^a = \{z^a_1, \ldots, z^a_\ell, z_{\ell+1}, \ldots, z^a_n\} \subset \mathbb{R}^{m^a}$ and $Z^b = \{z^b_1, \ldots, z^b_\ell, z^b_{\ell+1}, \ldots, z^b_n\} \subset \mathbb{R}^{m^b}$, where m^a and m^b are the dimensions of the word embeddings obtained from modality a and modality b, respectively. Let W^a and W^b be an $n \times n$ affinity matrix constructed from Z^a and Z^b, respectively, where for example, W_{ij}^a is the similarity between z_i^a, and z_j^a and n is the number of words in the vocabulary V. Let Δ^a and Δ^b be the unnormalized random walk Laplacian for W^a and W^b, respectively.

In a manner similar to that for single-graph label propagation, let Y be defined as a three-dimensional label set, where Y_ℓ is the label set of the seed words, and Y_u is the label set of the unlabelled words. The learning task defines a function that estimates Y_u from W^a, W^b, and Y_ℓ. In the multigraph setting, to spread the label from the labelled nodes to the rest of the graph, the weight of the edges is simultaneously combined in the iterative algorithm, as follows:

$$Y(t+1) = \zeta \Delta^a Y(t) + \eta \Delta^b Y(t) + (1 - \zeta - \eta) Y^0 \tag{4.2}$$

where ζ and η are the trade-off parameters between Δ^a and Δ^b, where $0 < \zeta, \eta < 1$ and $\zeta + \eta = 1$. The closed-form solution that minimizes the quadratic function of the multi-graph learning algorithm is computed as proposed in [269], as follows:

$$\hat{Y} = (1 - \zeta - \eta)(I - \zeta \Delta^a - \eta \Delta^b)^{-1} Y \tag{4.3}$$

4.3.3 Methods for Computing the Affinity Matrix

The single-label propagation algorithm was implemented using four different methods of computing the affinity matrix and word representations. The first was a semantic lexicon-based approach, called semantic lexicon-based label propagation (SLLP), in which the graph is built based on the semantic relationships between words. The second was a distributional-based approach designated count-based label propagation (CLP), in which the vocabulary and weights are derived from co-occurrence statistics in a corpus. The third method was neural word embedding label propagation (NWELP), and the fourth was a combination of semantic and distributed methods, termed semantic neural word embedding label propagation (SNWELP). The multi-graph label propagation, on the other hand, was implemented using semantic and neural word embeddings. The following subsections explain the different methods for constructing and computing the affinity matrices for the work presented in this thesis.

Semantic Lexicon-Based Label Propagation (SLLP)

Following the general principle of the graph-based label propagation approach described in the previous section, in the SLLP method, a semantic graph is constructed, in which the affinity matrix W is computed using features obtained from semantic lexicons. Two semantic lexicons were used in this algorithm: WordNet dictionary (WN) [171] and the paraphrase database (PPDB) [74].

WordNet, the most commonly used lexical resource or ontology in English, comprises words from all categories. WordNet groups words into synsets (synonyms sets), each of which contains a group of words that refer to the same concept. Words are also linked through other lexical relationships such as antonym, which indicates a semantic contract relationship [e.g., "*happy*" is an antonym of "*sad*"]) and hypernymy (which links general synsets to more specific ones like [e.g., "*colour*" is a hypernymy of "*red*"). The current English version of WordNet contains about 117,659 synsets that correspond to the sense of a word, and a total of 155,287 words. The PPDB is an English database that contains more than 220 million paraphrase pairs, of which 8 million are lexical paraphrases. The PPDB was constructed from a large bilingual parallel corpus using the distributional similarity scores of the paraphrases in Google n-grams [30] and the Annotated Gigaword corpus [86].

The SLLP algorithm constructs a vocabulary V from words derived from the dictionaries and computes the weight matrix W using the synonym and hypernymy lexical relationships in WordNet or the lexical paraphraser relationships in the PPDB. The semantic similarity $w_{i,j}$ of any pair of words x_i and x_j in vocabulary V is calculated as follows:

$$w_{i,j} = \begin{cases} 1.0 & \text{if } x_j \text{ is related to } x_i \\ 0.0 & \text{otherwise} \end{cases} \tag{4.4}$$

Count-Based Label Propagation (CLP)

Distributional word representation is the approach most commonly used in a graph-based learning algorithm for sentiment lexicon induction. The distributional representations are obtained in an unsupervised fashion from co-occurrence statistics aggregated from real-world corpora (e.g., news articles and Twitter), which are then utilized for building the similarity graph in a graph-based learning algorithm. In this work, for the implementation of label propagation using distributional word vectors, co-occurrence statistics aggregated from two real-world news corpora (section 4.4.1) were used. The word vectors are constructed by computing the context distribution smoothing (CDS) of the positive pointwise mutual information (PMI) ($PPMI_\alpha$) [144] from the co-occurrence matrix. This smoothing technique has been found to perform better than PMI and other word embeddings for a variety of natural language processing (NLP) tasks because it reduces the PMI bias toward rare words [144]. This CDS is rendered transferable to the PMI through the raising of the frequency of the context word by a smoothing factor α, as follows:

$$PPMI_\alpha(w, c) = max\{log_2 \frac{p(w, c)}{p(w)p_\alpha(c)}, 0.0\} \tag{4.5}$$

where $p(w, c)$ is the empirical co-occurrence probability, $p(w)$ is the marginal probability of a word w, and $p_\alpha(c)$ is the smoothed marginal probability of word context c. In this work, $\alpha = 0.75$ was used since it was found to give better results [144] [170], and the unsmoothed PPMI was also investigated. The $PPMI_\alpha$ is then factorized with truncated Singular Value Decomposition (SVD) [84], as follows:

$$PPMI_\alpha = U * \Sigma * V^T \tag{4.6}$$

The singular value decomposition word vector (W_{SVD}) is then computed by taking the top k rows of U (in the new method, $k = 100$ and $k = 300$ were used) and by adding an eigenvalue weighting score p (values experimented with were $p = 0.5$, $p = 0.0$) to the eigenvalue matrix Σ_d [144]. Details about hyperparameter tuning approaches for count-based word embeddings are provided in section 2.3.1.

$$W_{SVD} = U_k \Sigma_k^p \tag{4.7}$$

The affinity matrix is then calculated using the cosine similarity between the SVD word vectors in the corpus, as follows:

$$w_{ij} = \cos(v_i, v_j) = \frac{v_i v_j}{\|v_i\| \|v_j\|},\ \forall v_i, v_j \in W_{SVD} \tag{4.8}$$

Neural Word Embeddings Label-Propagation (NWELP)

Neural word embeddings that are learned in an unsupervised manner using feed-forward neural network have recently gained considerable attention in the NLP field because they have provided state-of-the-art results with respect to numerous NLP tasks. This continuous and dense word representation has replaced distributional word representation, including latent semantic analysis (LSA) and latent Dirichlet allocation (LDA) in many NLP problems. The work presented here involved the use of neural word embeddings (word vectors) that capture the syntactic and semantic properties of words in order to induce sentiment lexicons in a semi-supervised fashion. Two neural word embeddings were employed: an SGNS model [170] and the Global Vectors for Word Representation (GloVe) [199]. [1] The NWELP algorithm was implemented on two commonly known pre-trained SGNS and GloVe models as well as a SGNS and GloVe that are trained on a co-occurrence statistic obtained from Wikipedia dataset (section 4.4.1). The graph-based label propagation algorithm described above was implemented by constructing a relational graph that encodes the similarity between neural word embeddings. The graph was constructed using a vocabulary V of all the words in the word embeddings set (with non-alpha words and words that contain digits filtered out), and the affinity matrix W was computed using the cosine similarity (Equation 4.8) between word vectors (each $v_i \in V$ is a 300 dimensional vector). Details about the SGNS model and GloVe are provided in section 2.3.1.

Semantic and Neural Word Embeddings Label-Propagation (SNWELP)

To improve the results of label propagation based on a semantic lexicon and neural word embeddings, prior lexical knowledge (e.g., WordNet, PPDB) was linearly combined with neural word embedding in the label propagation algorithm. With the use of neural word embedding features (SGNS or GloVe) and semantic features obtained from a semantic lexicon (WordNet or PPDB), this problem was treated as a multimodal graph learning problem, with the weight of the semantic and the distributed graphs combined into a multigraph. The multigraph or multimodal graph was then converted into a single weighted graph by means of a pre-propagation fusion strategy that linearly combines the features obtained from the semantic and neural word embeddings. Given the weight matrices of two graphs W^a and W^b, which represent the weights of the semantic and the neural word embeddings graphs, the weight matrix of the combination of W^a and W^b is computed as follows:

[1] Several researchers, such as the author of [144], have considered GloVe to be a neural word embedding model while others have regarded it as a factorization of the co-occurrence statistics. Here, GloVe is considered a neural word embedding model.

$$W_{ij} = (\zeta\, W^a_{ij} + \eta\, W^b_{ij}) \tag{4.9}$$

where ζ and η are trade-off parameters in which $0 \leq \zeta, \eta \leq 1$ and $\zeta + \eta = 1$. The graph in this algorithm is constructed from vocabulary V, which is the intersection between the words in the lexicon and the words in the set of filtered embeddings.

4.4 Evaluation

4.4.1 Dataset

Semantic-based label propagation was performed on two semantic lexicons: WordNet and PPDB. Distributed (neural-network-based) and distributional (count-based) label propagation was implemented in two different settings, one of which involved the use of pre-trained word embeddings and the other, the use of word embeddings trained on a single corpus.

Pre-Trained Word Embeddings

Distributed or neural word embedding label propagation was implemented on two pre-trained word embeddings obtained from co-occurrence statistics aggregated from real-world datasets: the SGNS word2vector [170] and GloVe [199]. The SGNS model is trained on co-occurrence statistics aggregated from a Google News dataset and consisting of ∼3M words, each represented by a 300-dimensional vector. GloVe, on the other hand, is trained on co-occurrence statistics collected from Wikipedia and contains a total of ~ 1.9M words that are also represented by a 300-dimensional vector.

The count-based label propagation method was trained on co-occurrence statistics (n-gram features) obtained from two news corpora: the signal media (SM) one-million-news-article dataset, which contains $\sim$265K blog articles and $\sim$734K news articles [47], and the North American News (NAN) text corpus [85], which has $\sim$931K articles from a variety of news sources. The co-occurrence matrix of the count-based approach was computed for a window size of four words after the removal of stop words, words shorter than three letters, proper nouns, non-alpha words, and bigrams that occur no more than ten times. These heuristics reduced the set to $\sim$80K and $\sim$40K, for SM and NAN corpora, respectively.

Trained Word Embeddings

In the pre-trained approach described above, the label propagation algorithms were implemented using different pre-trained word embeddings, with each being trained on a different corpus. SGNS embeddings are trained on the Google news article dataset while GloVe ones are trained on the Wikipedia dataset, and the count-based (SVD and PMI) versions are trained on another dataset of news articles. To make the results generalizable and independent of the corpus used for training, the label propagation algorithm was trained on SGNS, GloVe, and count-based word embeddings, all obtained from a single corpus (the 10M-word English Wikipedia corpus). The word embedding models were trained on a window size of 10 words, with a minimum frequency of five times. The most frequent and infrequent words were also removed from the vocabulary, which was restricted to only 30K words since both the GloVe and the count-based approaches (SVD and PMI) require an amount of memory that is quadratic in terms of the number of words. The SGNS model was trained on negative sampling $k = 10$ and 1 epoch. The GloVe model was trained on the same set of words with a learning rate $= 0.05$ and 10 epochs. Count-based word embedding was also achieved using the same set of words and co-occurrence matrix, employing different values of the CDS parameter α, the number of truncated rows k, and the eigenvalue weighting parameter p of the $PPMI_\alpha$ and SVD.

4.4.2 Method for Sampling Seed Words

The choice of labelled words (also called paradigm or seed words) has a significant impact on the performance of graph-based learning algorithms. Most of the proposed methods for expanding or regenerating sentiment lexicons have used different sets of seed words that are selected manually [90]. In this study, two sets of seed words were used: 1) fixed seed sets (*fixed-paradigms*) adopted from the research by Osgood et al. [191] (Table 4.1), and 2) words that are randomly sampled from the vocabularies V used in the label propagation algorithm (*vocabulary-paradigms*). The *vocabulary-paradigms* set is randomly sampled from the words with the highest

EPA Seed Words
E+={good, nice, excellent, positive, warm, correct, superior}
E-={bad, awful, nasty, negative, cold, wrong, inferior}
P+={powerful, strong, potent, dominant, big, forceful, hard}
P-={powerless, weak, impotent, small, incapable, hopeless, soft}
A+={active, fast, noisy, lively, energetic, dynamic, quick, vital}
A-={quiet, calm, inactive, slow, stagnant, inoperative, passive}

Table 4.1: Osgood's fixed seed words (+ positive words and - negative words)

EPA Seed Words
E+={good, lovely, excellent, fortunate, pleasant, delightful, perfect, loved, love, happy}
E-={bad, horrible, poor, unfortunate, unpleasant, disgusting, evil, hated, hate, unhappy}

Table 4.2: Standard English seed words(+ positive words and - negative words)

and lowest EPA values (e.g., words with E, P or A ≤ -2.5 or ≥ 2.5). To sample an equal number of words from each polarity, sets were first created of all of the words with the highest and lowest EPA scores (E+, E-, P+,P-, A+, and A-) in the vocabulary and then from each set a sample number of words was taken equal to $(s-1)$ where s is the minimum number of words in all the four lists. The objective of employing this sampling procedure is to use words at the extremes of each E, P, and A dimension as paradigm words in order to propagate these highly influential EPA scores throughout the graph. The algorithms were implemented with seed words that represent a contribution of no more than 1% of all of the words in the vocabulary. Experiments were also conducted with the *fixed-paradigm* sets, but the results of the *vocabulary-paradigms* were significantly better. To compare the newly developed label propagation algorithm against state-of-the-art methods, another set of seed words was also used, called the standard English seed set, which is utilized by some researchers [90] in the field for inducing one-dimensional sentiment lexicons from *standard-English-paradigms* (Table 4.2).

4.4.3 Evaluation Metrics

To assess the performance of the label propagation algorithm with respect to generating a multi-dimensional sentiment lexicon, the most recent manually annotated affective dictionary by Warriner et al. [291] was chosen as a baseline. In the lexicon induction procedure, paradigm words were also sampled from the Warriner et al. dictionary, and the generated lexicon was compared against it. The affective dictionary was randomly divided into *EPA-training* (one-third of the set, equal to 5566 words) and *EPA-testing* (two-thirds of the set, equal to $8,349$ words). The seed words for all algorithms are sampled from *EPA-training* set only, and all results are presented on the *EPA-testing* set. The initial Warriner et al. [291] EPA scores were in the range of $\in [1, 9]$, and they were rescaled to $\in [-4.3, +4.3]$ to conform to the EPA scale used in the other lexicons under consideration [100]. The $[-4.3, +4.3]$ scale is the standard scale used by most sociology researchers who study or measure individuals' emotions toward words.

Four evaluation metrics were used for comparing the induced EPA (*EPA-induced*) scores against the manually annotated EPA (*EPA-testing*) ones: mean absolute error (MAE), Kendall τ rank correlation, F1-binary (positive and negative), and F1-ternary (positive, neutral, and negative). The F1-binary is used for evaluating the binary classification performance of the model (positive ≥ 0 and negative < 0). As in most recently proposed studies in the field [90], the F1-ternary is computed in order to measure the accuracy of the ternary classification: positive $\in (1, 4.3]$, neutral $\in [-1, 1]$, and negative $\in [-4.3, -1)$. The calculation of the F1-ternary was achieved using *class-mass normalization* (CMN) methods [315], which rescales the predicted label ($\hat{y}_{i,l}$) for a point x_i by incorporating the prior class, as follows:

$$\underset{l}{\operatorname{argmax}} \quad w_l \; \hat{y}_{i,l}$$

where w_l is the label mass normalization, which is equal to p_l/m_l, where p_l is the prior probability of a label l (computed from the labelled data), and m_l is the estimated weight of label l over the unlabelled sets. This scaling method is known to improve results more effectively than the typical decision function $\operatorname{argmax}_l \quad \hat{y}_{i,l}$.

4.4.4 Baseline and State-of-the-art Comparison

To compare the new approach against state-of-the-art algorithms for inducing valence (evaluation scores), an unsupervised learning algorithm (PMI-IR [275]) was implemented in order to estimate the sentiment orientation (either positive or negative) of a word by using the co-occurrence statistics aggregated from search engine results in order to compute the difference between the strength of the word associations with positive paradigm words and with negative paradigm words. The results were also compared against the results reported using the orthogonal transformation of word vectors developed by Rothe et al. [218], and against those from a label-spreading algorithm trained on a (domain-specific) SVD word vector model [90]. Experiments were also conducted with a retrofitted word vector model that improves neural word embedding vectors using semantic features obtained from lexical resources (WordNet, PPDB) [66]. To make a fair assessment and to compare the results with those reported in [90] and [218], the label propagation algorithm and the retrofitted word vector approach [66] were implemented in order to recreate the General Inquirer lexicon [253] with a valence score $\in R$ from the Warriner et al. lexicon [291] using *standard-English-paradigms* (Table 4.2). The neutral class was ignored, and the same seed set employed by the authors of [90] and other researchers in the field was used.

All of the results were also compared against the EPA scores obtained from a supervised learning algorithm. A support vector regression (SVR) model was trained on a co-occurrence statistics model derived from the SGNS word embedding model [170] and the sentiment lexicon resource [291]. The SVR model was trained using two different sets of labelled words: the sampled seed words (*vocabulary-paradigms*) and the full training set (*EPA-training*). The SVR model uses a radial basis function (RBF) kernel with $C = 10$, and $\gamma = 0.0$ for training and evaluation against the test set *EPA-testing*.

4.5 Results and Discussion

This section presents an evaluation of the label propagation algorithm using different graph structures, word representations, and sets of seed words, which can be divided into four parts: evaluation of the semi-supervised learning approach with respect to extending a multidimensional lexicon (EPA) using the single-graph approach, pre-trained word embeddings, and sampled seed words (*vocabulary-paradigms*) (*Part One*); assessment of the performance of the label propagation algorithm with respect to regenerating a one-dimensional lexicon using single-graph learning methods, pre-trained word representations, and the *Standard-English-paradigms*, including a comparison of the results against the state-of-the-art methods (*Part Two*); evaluation of the performance of the count-based and distributed-based label propagation model using the trained word embeddings (*Part Three*); and evaluation of the performance of the multigraph learning approach implemented using semantic and distributed word representation against the single-graph learning approach (*Part Four*).

Part One: Table 4.3 shows the results of the comparison of the EPA scores induced using the label propagation algorithms against their corresponding values in the Warriner et al. lexicon [291]. As indicated in the table, the count-based word embeddings label propagation (CLP) algorithm generated the lowest ranking correlation τ and the highest error rate (MAE) compared with the other label propagation methods. The results of the comparison of the induced EPA scores against their true values in the testing set show that the MAE ranged between 0.99 and 1.3 and that, based on the cosine similarity and hard clamping ($\alpha = 1.0$) assumption, the ranking correlation τ was less than 0.2. [2] Experiments were also conducted with the unsmoothed PMI and the shifted positive PMI (SPPMI) [143] (with the number of negative samples $n = 10$), but no significant difference between them was evident. Tests also included different dimensions of the SVD word vector $k = 100$ and $k = 300$, but no significant difference was observed between the 100- and 300-dimension SVD word embeddings.

The semantic lexicon label propagation (SLL) approach, which uses the lexical features obtained from either WordNet or PPDB lexicons, generated a total of ~ 50K words, with ~ 4K words in the testing set (*EPA-testing*). When the induced EPA scores are compared with their corresponding values in (*EPA-testing*), the results reveal a reasonable amount of agreement between the EPA scores induced using dictionary-based features and the manually labelled EPA values. They also show that the two semantic lexicons produced relatively similar results.

The τ correlation scores for both of the neural word embedding label propagating approaches (NWELP) were higher than for the semantic lexicon or the count-based approach, as shown in Table 4.3. The MAE and F1-scores for the semantic-based and neural word embedding label

[2] For all reported τ scores, the p-value is less than 0.001

Method	Corpus	τ			F1-binary			F1-ternary			MAE		
		E	P	A	E	P	A	E	P	A	E	P	A
CLP	**SM**	0.219	0.0263	0.162	0.53	0.44	0.56	0.42	0.45	0.44	1.1	1.09	0.85
	NAN	0.122	0.060	0.084	0.51	0.54	0.54	0.50	0.42	0.45	1.3	1.0	0.99
SLLP	**WN**	0.388	0.244	0.329	0.72	**0.83**	0.73	0.65	0.60	0.75	0.91	**0.79**	**0.71**
	PPDB	0.391	0.181	0.309	0.73	0.76	0.71	0.62	0.60	0.65	0.92	0.89	0.79
NWELP	**SGNS**	0.437	0.283	0.350	0.70	0.80	0.67	**0.69**	**0.65**	**0.79***	**0.84**	1.08	0.88
	GloVe	0.430	0.113	0.357	0.73	0.81	0.70	0.68	0.64	0.78	1.09	1.07	0.84
SNWELP	**PPDB+GloVe**	0.434	0.209	0.360	0.74	0.81	0.70	0.68	0.64	0.77	1.09	1.07	0.84
	WN+GloVe	0.445	0.220	0.366	0.75	0.82	0.71	0.68	0.64	0.78	1.07	1.05	0.84
	PPDB+SGNS	**0.510**	0.284	0.459	**0.76**	0.80	**0.75**	0.68	0.64	0.78	1.10	0.97	0.84
	WN+SGNS	**0.510**	**0.291**	**0.461**	**0.76**	0.80	**0.75**	0.68	0.64	0.78	1.10	0.95	0.83
SVR	**SGNS-Seed Words**	0.52	0.34	0.42	0.81	0.82	0.78	0.62	0.64	0.73	0.75	0.67	0.63
	SGNS-Training Set	**0.628***	**0.422***	**0.500***	**0.83***	**0.84***	**0.78***	**0.72***	**0.65***	0.68	**0.60***	**0.60***	**0.56***

Table 4.3: Results of the label propagation algorithms and the supervised learning (SL) method (support vector regression (WESVR)) using the **sampled seed words** as compared with the ground truth EPA values (Method= the algorithm used for lexicon induction, τ = Kendall's τ correlation, F1-binary= F1 measure of the binary classification, F1-ternary= F1 scores of the ternary classification, MAE=mean absolute error). The highest scores of the label propagation algorithms are in **boldface**. The highest scores of all the algorithms are designated by **boldface***.

propagation were close. A comparison of the results of the two pre-trained neural word embedding models against each other indicates that the skip-gram-based (SGNS) model performed better than GloVe for all three dimensions (EPA). Experiments were performed with different thresholds (0.0, 0.3, and 0.5) of the cosine similarities, and the results of using different thresholds varied a great deal with respect to the number of induced words and the level of accuracy. Higher thresholds provided more accurate results with less noise, but with fewer induced words. The results reported in Table 4.3 and Table 4.5 were obtained using a cosine similarity threshold equal to 0.0 since both the SGNS and GloVe adjacency matrices contain negative values.

Linearly combining the semantic and neural word embedding features improved the results with τ ranging between 0.43 and 0.51 and the MAE ≤ 1.1 for the evaluation scores (E). Exper-

Word	Method	Induced EPA	True EPA
injustice	WN	[-1.9, 0.3, -1.7]	[-2.7, 1.6, -1.86]
injustice	GloVe	[-1.3, 1.4 , -1.8]	[-2.7, 1.6, -1.86]
injustice	GloVe+WN	[-1.4, 0.2, -1.3]	[-2.7, 1.6, -1.86]
injustice	SG+WN	[-1.9, 0.3, -1.7]*	[-2.7, 1.6, -1.86]
evil	PPDB	[-1.3 , 0.05, -1.1]	[-2.9, 0.7, -1.5]
evil	GloVe	[-2.1, 2.5, -3.1]	[-2.9, 0.7, -1.5]
evil	GloVe+PPDB	[-1.7 , 0.08, -1.2]	[-2.9, 0.7, -1.5]
evil	SG+PPDB	[-2.1, 0.1, -1.5]	[-2.9, 0.7, -1.5]
successful	SG	[2.15, 0.04, 1.6]	[2.97, 0.09, 2.9]
successful	SG+PPDB	[2.5, -0.6, 2.0]	[2.97, 0.09, 2.9]

Table 4.4: Examples of the induced EPA scores and their EPA ratings from the Warinner et al. lexicon, and the induced EPA values using label propagation and different word representations. WN = WordNet, PPDB = paraphrase database, SGNS = skip-gram negative-sampling word vector, and GLoVe = the Global Vector for Word Representation. The starred example * shows no change after the addition of neural word vector features.

iments were conducted with different values of the trade-off parameters (e.g., $\zeta = 0.8$, $\eta = 0.2$ and $\zeta = 0.5$, $\eta = 0.5$), but no significant differences between them were evident. The reported results were obtained with obtained with $\zeta = 0.5$, $\eta = 0.5$. The findings also show that the results produced with the semi-supervised methods were in line with those from the supervised SVR model that is trained on the seed words (*vocabulary-paradigms*). However, the SVR model trained on the whole training set significantly outperformed the semi-supervised methods.

A comparison of the results across the different affective dimensions (E, P, and A) reveals that, in both the semi-supervised and the supervised learning algorithms, the rank correlation score τ for comparing the potency (P) values to their counterpart scores in the testing set was low relative to the scores for evaluation (E) and activity (A). In contrast, in all the algorithms, the rank correlations τ were the highest for the evaluation (E) scores, followed by the rank correlation τ values for activity (A), which indicates that the semantic and distributed similarities between words are strong indicators of how close the words are in the evaluation and activity space. Table 4.4 shows some of the induced EPA scores and their corresponding values in the Warriner

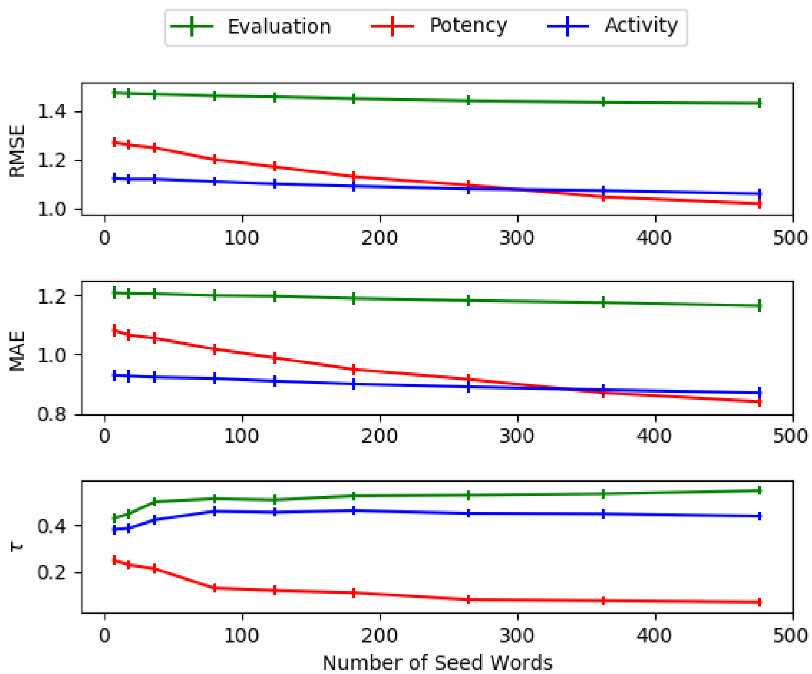


Figure 4.1: Root mean square error (RMSE), mean absolute error (MAE), and τ rank correlation scores using different numbers of seed words

et al. [291] dataset. The table also provides some examples of the same words and their induced EPA scores when different word representations are used.

Figure 4.1 shows the performance of the label propagation algorithm with the use of different numbers of seed words. As indicated in the figure, the addition of more seed words improves the error rate and τ correlation scores for evaluation and activity, but it lowers the τ correlation scores for the potency. The figure shows that, for evaluation and activity, the error rate decreases and the τ correlation score increases as the number of seed words rises, while for potency, the error rate and τ correlation score both increase as the number of seed words decreases. The reported error rates and correlation scores were obtained using the semantic and neural word embedding label propagation (SNWELP) model, and an equal number of samples are taken from each dimension.

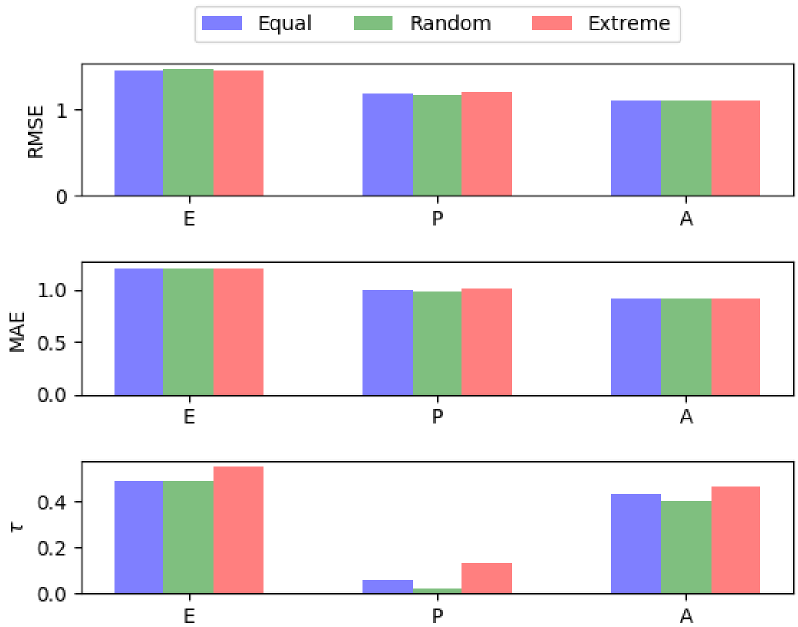


Figure 4.2: Root mean square error (RMSE), mean absolute error (MAE), and τ rank correlation by sampling method

The performance of the label propagation methods was also compared using three different sampling strategies: sampling from the set of words at the extremes of each dimension (Extreme); equal sampling from each category in the EPA space (neutral $\in [-1, 1]$, negative $\in [-4.3, -1)$, and positive$\in (1, 4.3]$) (Equal); and random sampling from the entire EPA space (Random). The results of a comparison of the error rates associated with the three sampling strategies reveal no significant difference among them, but for all of the EPA dimensions, the τ correlation scores for the approach that samples from the words with the highest and lowest EPA scores were higher than those for the other two approaches (Figure 4.2). The number of seed words was identical for each of the sampling strategies.

Part Two: Table 4.5 shows the results when the new approach for inducing evaluation scores (E) from the General Inquirer lexicon was compared with state-of-the-art methods. The results indicate that the label propagation algorithms performed significantly better than the unsupervised learning method and that the semantic neural word embedding (SNWELP) model outper-

Method	τ	F1-ternary	ACC
SNWELP (SGNS+WN)	0.51	0.68	0.94
Hamilton et al. [90]	0.50	0.62	0.93
NWELP (SGNS)	0.48	0.67	0.94
Rothe et al. [218]	0.44	0.59	0.91
Faruqui et al. [66]	0.40	0.62	0.84
SLLP (WN)	0.16	0.65	0.64
Turney et al. [274]	0.14	0.47	0.55

Table 4.5: Results of comparing the evaluation (E) scores for the lexicon induced from the **General Inquirer** induced lexicon using **pre-trained** neural word embedding label propagation (NWELP) and the semantic neural word embedding label propagation (SNWELP) against the results reported for state-of-the-art methods for lexicon induction (τ = Kendall's τ correlation, ACC= the binary accuracy, F1= the ternary F-measure)

formed both the Rothe et al. [218] and the Hamilton et al. [90] approaches. The neural word embedding and semantic neural word embedding algorithms also perform better than label propagation based on the retrofitted word vector approach. The results reported were obtained using the improved skip-gram (SG) model and the semantic features obtained from WordNet [66]).

Part Three: To evaluate the word representations or embeddings independently of the corpora, the distributed and the count-based word embedding models were trained on the same corpus and co-occurrence statistics. As shown in Table 4.6, the results of the comparison of the induced evaluation scores (E) in the General Inquirer lexicons against their corresponding values in the hand-coded lexicon were consistent with the results obtained using the pre-trained models, since the SGNS model provided the best performance. However, the results with the count-based approaches that use CDS ($PPMI_\alpha + SVD$, and $PPMI_\alpha$) were better than those using GloVe. The findings also reveal that the use of the truncated SVD improved the correlation score of the $PPMI_\alpha$. As indicated in the table, the results from all of the trained word embedding approaches are inferior to those from their corresponding pre-trained models because the trained word embeddings were obtained using a relatively smaller vocabulary set (|V|= $30K$) than that used with the pre-trained model.

Part Four: Multigraph label propagation produced outcomes that were almost identical to

Method	τ	F1-ternary	ACC	MAE
SGNS	**0.43**	0.65	**0.70**	**1.19**
$PPMI_\alpha + SVD$	0.35	0.65	0.66	1.21
$PPMI_\alpha$	0.33	0.67	0.68	1.20
GloVe	0.27	0.65	0.66	1.21

Table 4.6: Results of a comparison of the evaluation (E) scores from the lexicon induced from the **General Inquirer** using the **trained** neural word embedding label propagation (NWELP) against count-based word embeddings (τ = Kendall's τ correlation, ACC= the binary accuracy, F1-ternary= the ternary F-measure, MAS=mean absolute error)

Trade-off parameters	τ	F1-ternary	ACC	MAE
$\zeta = 0.8\ \eta = 0.2$	**0.47**	0.67	0.70	1.2
$\zeta = 0.5\ \eta = 0.5$	0.46	0.67	0.71	1.2
$\zeta = 0.2\ \eta = 0.8$	0.39	0.67	0.71	1.2

Table 4.7: Results of a comparison of the evaluation (E) scores from the lexicon induced from the **General Inquirer** using **multi-graph label propagation** based on different values of the trade-off parameters (τ = Kendall's τ correlation, ACC= the binary accuracy, F1-ternary= the ternary F-measure, MAE=mean absolute error)

those that resulted from the single-graph learning approach. The results reported in Table 4.7 were obtained using semantic and neural word embeddings with different values of the trade-off η and ζ. The findings show no significant differences among the error rates and accuracy scores obtained using different values of trade-off scores, but the greater weight associated with semantic similarity (0.8) than with distributed similarity (0.2) significantly improved the rank correlation scores.

4.6 Conclusion

One of the methods most commonly used for addressing SA problems is the use of sentiment lexicons in addition to distributional and semantic features. Sentiment dictionaries are typically composed of a relatively small set of words that are manually annotated by human judges or automatically generated using distributional word representations (count-based statistics) and some seed words. These lexicons often fail to include new words that emerge on the Internet or through social media (e.g., selfie, sexting, photobomb), fail to measure current attitudes toward these words and terms (e.g., homosexuality, abortion), and represent human emotions in only one-dimensional space.

With the goals of overcoming these limitations, reducing the costs associated with manual annotation, and increasing the number of annotated terms, this thesis presents a novel approach that implements a graph-based learning method using a linear combination of semantic and distributed word representations to regenerate sentiment lexicons as well as to provide an intensive evaluation of corpus- and thesaurus-based graph-based semi-supervised algorithms. The proposed approach was evaluated with respect to the regeneration of a three-dimensional and a one-dimensional sentiment lexicon, and the results were compared against manually annotated scores and the performance of some of the state-of-the-art sentiment induction methods.

The results of the evaluation show that incorporating distributed features into label propagation based on a semantic lexicon significantly improves performance. When different word embedding approaches (SGNS, GloVe, and count-based word embeddings) were evaluated, the results were generally consistent with those from other studies whose goal was to assess the performance of word vector models in regard to word ontology and other NLP tasks, since the word2vec (SGNS) model outperformed the count-based and GloVe algorithms with respect to pre-trained and trained word embeddings. When the results produced by the semantic neural label propagation algorithm are viewed against those generated using a supervised learning algorithm (SVR) trained on word embeddings and a sentiment lexicon, the results for the semi-supervised and unsupervised learning methods are comparable. The semi-supervised algorithm, however, does not require a large training dataset and permits the re-estimation of the labels of previously-labelled words.

Chapter 5

Entity-Based Reader Affect Recognitions with Single-Event Sentences

5.1 Introduction

Sentiment analysis (SA) studies and applications have been a recent target of increased attention in the scientific community, especially in the computer science and natural language processing (NLP) fields. The mainstream research in SA focuses on analysing subjective text, such as review documents, that describes an explicit opinion or belief about an object or entity (e.g., *"Seventh Son is a terrible movie"*). They also utilizes count-based or distributional word features to classify textual content to a set of pre-defined labels (e.g., happy, sad, and neutral). Using the distributional or count-based word features revealed many limitations. The distributional features might impair the sentiment analysis capability, since they are built upon corpus frequency that unlikely to be sentiment-relevant (e.g., "good" and "bad" may have similar distributional word representations if they appeared in a similar context). They also make the sentiment analysis models context-sensitive since they are obtained from corpus-based feature (e.g., movie review, product review), and thereby not generalizable to other type of corpora. Most of the studies in the sentiment analysis field have also performed coarse-grained analyses in which a piece of text (such as a sentence or an article) is mapped to a label or a point between polarities. Only a handful of studies have performed entity-based sentiment analyses where sentiment labels or scores are attached to an aspect of an object, such as "the lens of the camera was stuck many times". There have been an insufficient number of studies that have incorporated linguistic structure, or psychological or linguistics theory or knowledge into statistical language models.

The focus of this chapter is the analysis of fact-based or objective sentences that describe events. Despite limited research related to fact-based SA, many factual statements carry or trigger sentiments, including such sources as news articles and blog comments. Subjective documents contain highly opinionated words or phrases about a situation or aspect that is classifiable by means of shallow parsing or a basic word look-up strategy. In contrast, performing SA on fact-based textual information requires an understanding of the meaning and syntactic structure of the sentence. A need thus exists for a new approach that analyzes the sentiment associated with objective, or factual, text as well as its components. For example, a sentence such as *"x kills y"* will clearly evoke a negative sentiment for the reader, and highly negative (yet different) sentiments toward each of *x* and *y* (e.g., angry at *x* and sorry for *y*). Further, the initial sentiment felt for each entity participating in an event might affect the reader's judgement regarding the situation. For example, if it is known that *x* is a victim and *y* is a criminal, then the sentiment triggered is likely to be positive in general and positive toward *x*, yet remain negative toward *y*.

This chapter introduces a fine-grained SA approach that utilizes socio-mathematical theory, or affect control theory (ACT), to model reader emotions arising from text about real-world events and their associated entities (i.e., actor and object) in a three-dimensional space. ACT is based on the premise that social situations involve a range of culturally shared sentiments toward the entities (identities and behaviours) in a social situation, and that these sentiments develop or generate a situated sentiment (transient impression) whereby individuals try to remain close to their fundamental sentiment. An additional ACT principle posits that human social and emotional responses are controlled by differences between culturally shared and transient sentiments.

The approach presented in this chapter is based on ACT impression and emotion formation equations for predicting human sentiments about events by first extracting the sentence components (subject, verb, and object) using the syntactic event extraction method described in (Chapter 3). The next step was to establish the affective meaning of each element from an augmented affective meaning lexicon. The affective meaning dictionary was expanded using a semi-supervised graph-based learning algorithm built from lexical resources (WordNet) (Chapter 4). ACT equations were then used for computing the predicted emotions toward the actor, behaviour, and object. The estimated emotions are then compared against manually annotated news headline dataset. This approach yielded accuracy levels of 72% and 85% with respect to predicting reader emotions regarding the subject and object in the news headlines. These results are significantly better than those obtained from a model trained using bag-of-words and the SentiWordNet lexicon.

5.2 Background

The hypothesis underlying ACT is that individuals perceive identities, behaviour, and objects, including words, in terms of affective meanings represented by a score along a three-dimensional space (evaluation, potency, and activity (EPA)). These affective meanings are associated with interpersonal social identity and behaviour control perceptions and interactions. In any given situation, these initial affective meanings or *fundamental sentiments* develop *transient impressions* that individuals use as a means of remaining close to the *fundamental sentiments* in the EPA space. Concretely, ACT models an emotion as a function of the difference between the *transient impressions* and the *fundamental sentiments*. Thus, when a situation causes a discrepancy between fundamental and transient impressions, the emotions are likely to be more prominent. For example, in a situation such as "*An employee yelled at an employer*", the employer, who initially felt powerful and has been made to feel powerless (situated feeling), is likely to feel "angry".

ACT provides a mathematical formulation for modelling the situated impression, as well as the emotions arising from fundamental sentiment interactions (i.e., their effective meanings) regarding self, others, the settings in which they are embedded, and the behaviours occurring throughout. These mathematical formulations are obtained through empirical studies and regression analysis. This chapter provides a brief review of fundamental and transient impression calculation and then focuses on emotional dynamics and deflection equations. Additional detail regarding ACT is included in section 2.2.

5.2.1 ACT Mathematical Model

In ACT, events are represented in terms of three or four main elements: subject (S) or actor, verb (V) or behaviour, object (O), and sometimes the setting or the location (L) where the event is embedded. Three values (EPA) represent the "*fundamental sentiments*" of these elements, capturing the culturally shared sentiments they evoke in terms of evaluation, potency, and activity. A nine-dimensional vector (or 12-dimensional vector when the locations or settings are considered) represents the *fundamental sentiment* of an event:

$$F = \{f_{s_e}, f_{s_p}, f_{s_a}, f_{v_e}, f_{v_p}, f_{v_a}, f_{o_e}, f_{o_p}, f_{o_a}\}$$

where, e.g., f_{s_e} is the fundamental sentiment of the subject (S) on the evaluative dimension. The situated or *transient impression* arising from an event is also represented in a nine-dimensional vector:

$$T = \{t_{s_e}, t_{s_p}, t_{s_a}, t_{v_e}, t_{v_p}, t_{v_a}, t_{o_e}, t_{o_p}, t_{o_a}\}$$

The transient impression T is computed by multiplying a vector of the combinations of terms from the *fundamental sentiment* f, by a matrix M of prediction coefficients estimated based on impression-formation research. Sections 2.2 and 6.2 provide additional details.

5.2.2 Events Likelihood and The Deflection

The difference between a fundamental sentiment and a situated impression that an event generates is an important measure of the likelihood or normality of an event. A large deflection ($d > 4$) indicates an inconsistency between the behaviour and the identities involved in an event, such as one that describes " *a student yells at a teacher*", for which the level of deflection is equal to 7.7. On the other hand, a small deflection indicates consistency between the behaviour performed and the role identities of the subject and object. For example, in an event such as "*a teacher teaches a student*", the deflection level is equal to 1.1. In ACT, deflection is computed either as the squared Euclidean distance between the *fundamental sentiments* (F) and the *transient impressions* (T) (Equation 2.5), or as a probability distribution as proposed by Hoey et al. [107, 106, 108].

$$d = e^{-\sum \sigma (F-T)^2} \tag{5.1}$$

where σ is a weight vector. In this study, the deterministic Euclidean distance approach was used for computing deflection.

5.2.3 Emotion Dynamics

In ACT, like identities and behaviours, emotions are grounded in three affective dimensions (EPA). For example, feeling "thrilled" has an EPA profile of $\{2.08, 1.95, 2.37\}$, meaning that such an emotion feels quite good, slightly potent, and quite active. ACT parameters specify that individuals experience emotional responses as result of the difference between the fundamental sentiments and the transient impressions that arise from an event. ACT models the emotions that a person feels when interacting in a social event as a function of the fundamental identity and transient identity of the actor $F_r \equiv \{f_{s_e}, f_{s_p}, f_{s_a}\}$ and $T_r \equiv \{t_{s_e}, t_{s_p}, t_{s_a}\}$ or of the fundamental identity and transient identity of the object $F_r \equiv \{f_{o_e}, f_{o_p}, f_{o_a}\}$ and $T_r \equiv \{t_{o_e}, t_{o_p}, t_{o_a}\}$. The emotion ε that each participant in the event is experiencing is defined in terms of an empirically measured equation that is approximately equal to the difference between the fundamental F_r and the transient identity T_r and is represented in three-dimensional space $\{\varepsilon_e, \varepsilon_p, \varepsilon_a\}$. The emotional response that individuals in a social event experience is computed as follows:

$$\varepsilon \propto E\,(T_r - I\,F_r - \delta) \tag{5.2}$$

where E is a 3×3 matrix coefficient of the emotion profile, I is a 3×3 matrix coefficient for identity, and δ is a vector of equation constants. Chapter 6 includes a detailed discussion of this function.

Amalgamation of the identities and the modifiers, such as the salient characteristics or the mood, plays a significant role in forming the emotional response of the participants of an event. For example, a "physician" is associated with an identity considered to be $\{2.01, 1.67, -0.10\}$, while the EPA profile of " female physician" is equal to $\{1.58, 0.22, -0.23\}$. This latter value is less positive, less powerful, and less active than "physician". In ACT, this amalgamation of the modifiers and the identities is achieved through the linear combination of the EPA profile of the identity and the EPA profile of the modifier, on one hand, with the multiplication of empirically driven coefficients, on the other (Equation 2.4).

Researchers have analyzed the empirically driven emotion equation (Equation 5.2) in order to determine the main factors involved in shaping the emotional experience that accompanies social events [156]. They found that fundamental identity F_r plays an important part in shaping individuals' emotions by setting expectations regarding the role that each participant plays in social events. The positivity of individuals' emotions ε_e in particular, has been found to be affected primarily by the evaluation of transient identities. For example, if an actor performs an act that makes her/him look good, he/she will feel good. The positivity of emotions is also affected by the transient activity (i.e., t_{o_a} or t_{s_a}) and the fundamental activity (i.e., f_{o_a} or f_{s_a}). For example, if the transient impression of one's activity is greater than one's fundamental activity, one will feel more positive emotions, and vice versa.

The potency factor associated with an individual's emotions ε_p, which corresponds to the dominance or vulnerability aspects of their emotion (e.g., anger or fear), also varies with the evaluation and activity, and evaluation and potency processes. For example, if the evaluation (i.e., t_{o_e} or t_{s_e}) is negative and the potency (i.e., t_{o_p} or t_{s_p}) is positive, individuals will feel negative but potent feelings (e.g., anger). If the transient has a higher evaluation (i.e., t_{o_e} or t_{s_e}) level and activity (i.e., t_{o_a} or t_{s_a}) level than those of the fundamental, a higher-potency emotion will be experienced. The dynamics of the emotion on the activity dimension ε_a also vary with the transient activity and evaluation (i.e., t_{o_a} or t_{s_a}, and t_{o_e} or t_{s_e}). For example, if the transient activity is higher than the fundamental activity, a high-arousal emotion will result.

5.3 Method

To compute reader emotions triggered by event-based sentences, the event entities (i.e., subject, verb, object, setting, and modifiers) are first extracted, and the EPA scores of each entity are then obtained from the extended EPA lexicon (section 5.3.2). Next, the amalgamation of the modifiers and identities is computed using Equation 2.4, and the fundamental sentiments (F) are constructed based on the EPAs of the extracted and modified entities. Given the fundamental sentiment (F), the transient impression (T) is calculated using Equations 2.1 and 2.2. The deflection d is then determined using Equation 2.5, the emotions about the subject and object are calculated using Equation 5.2, and the EPAs of the estimated emotions are mapped to the nearest emotion label in ACT by means of a Euclidean distance measure. The ACT dataset [161] has 135 emotion labels, each with an EPA score (e.g., delighted=$\{2.04, 0.96, 1.48\}$). The final step is to compare the estimated emotions ε about the subject and object of a given event to the corresponding ground truth in the news headline dataset (section 5.4).

5.3.1 Extracting Sentence Quintets from Event-Based Text

The extraction of sentence entities or quintets involved the development of a syntactic-based algorithm inspired by the one created by Rusu et al. [225], which takes a syntactic parse tree of a given sentence and returns the subject, verb, and object. The event extraction algorithm constructed for this study includes consideration of the syntactic structures of news headlines. News headlines do not follow typical English syntactic structures, and may have several noun phrases without articles or may be missing the verb "to be". They even use the present tense to describe current or past events, such as " *Terror strikes police base*". Past-tense verbs are rarely used in news headlines, and passive voice sentences are common. Passive voice sentences are often written without an auxiliary verb, making it more difficult for standard parsers to accurately distinguish the subject, verb, and object (e.g., "*Six killed in accident*"). Details about syntactic-based event extraction methods are provided in Chapter 3. Models implemented for this work also include an ACT-based SA model that employs a semantic and syntactic-based event extraction method that provides the highest level of accuracy of the available approaches (Chapter 3).

Consideration also included whether a verb type is transitive, which is a direct indication of positive or negative sentiment about something (e.g., x killed y), or intransitive, which transfers sentiments into nouns (e.g., x provides help to y). For intransitive verbs, the second verb is chosen as the behaviour (verb) in the sentence (e.g., "x provides help to y" will be "x helps y"). The process also involved the use of part-of-speech tagging as a means of determining the elements of an event and identifying places and names. The gender of the names is taken into account through the training of a naïve Bayes classifier on names-gender corpus of $5,001$ female and $2,943$ male names[1], yielding a gender classification accuracy level of 86%.

[1] www.nltk.corpus

5.3.2 Lexicon Induction Expansion Using WordNet

To augment the affective meaning dictionary, a graph-based semi-supervised label-propagation algorithm was implemented in order to build a graph that encodes similarity between nodes and propagates the labels from the labelled nodes to the rest of the graph. Further details about the algorithm are provided in Chapter 4. This chapter describes the implementation of the graph-based label algorithm according to the semantic similarity obtained from the WordNet dictionary and from randomly sampled seed words derived from an affective meaning dictionary.

To obtain the seed words, or the labelled nods, for the label propagation algorithm, a relatively large affective meaning dictionary was created by combining three manually annotated English lexicons: an ACT lexicon that contains $2,293$ words (ACT-lexicon) [161]; the Affective Norm for English Words (ANEW) [29] lexicon, which has $2,476$ English words; and the Warriner et al. data-set [291] that contains $13,915$ words. Both ANEW and Warriner et al. datasets were scaled from a range of [1,9] to a range of [-4.3,4.3] [92]. These two affective meaning lexicons were selected to be added because they are the most recent affective meaning dictionaries that contain larger sets of words. Combining the three dictionaries generated a lexicon containing $17,347$ words, which was randomly divided into a training-EPA-lexicon and a testing-EPA-lexicon with $5,782$ and $11,565$ words, respectively.

The vocabulary sets in the label propagation algorithm were formed through the addition of all the synonyms and lemmas of a specific part of speech (verb, noun, adjective, or adverbs) in the WordNet dictionary, and the labels of the seed words were obtained from the training-EPA-lexicon. The weight matrix or adjacency matrix was computed using the Wu and Palmer (*wup*) similarity measure [298], based on the idea that concepts with greater depth in each taxonomy are more similar. The Wu and Palmer similarity measure is equal to the depth of the least common subsumer (LCS), also called the least common ancestor, divided by the summation of the depth of the two words in the WordNet taxonomy:

$$wup(w_1, w_2) = \frac{2 * depth(LCS(w_1, w2))}{depth(w_1) + depth(w_2)}$$

where w_1 and w_2 are two words in the dictionary, $depth(LCS(w_1, w2))$ is the depth of the least common ancestor of w_1 and w_2 in a taxonomy, and $depth(w)$ is the depth of a given word.

POS	W	RMSE			MSA		
		E	P	A	E	P	A
Adjectives	378	1.2	1.2	0.9	0.9	1.0	0.8
Adverbs	5	1.0	1.1	1.3	0.7	0.8	0.9
Verbs	2,787	1.2	1.1	0.8	1.0	0.9	0.6
Nouns	7,079	1.3	1.1	0.9	1.0	0.9	0.7
Total	10,249	1.3	1.1	0.9	1.0	0.9	0.7

Table 5.1: Results of a comparison of the lexicon induced using label propagation and the ground truth EPA values (POS = part-of-speech, W = the number of induced words, MAS = mean absolute error, and RMSE = root mean square error)

The *wup* similarity measure will take a value $\in (0, 1]$, in which 1 indicates that the two concepts are the same. This formulation also indicates that any value that is close to zero means that the two concepts are entirely different (*wup* similarity cannot be equal to zero because the depth of the root of a taxonomy is equal to one). For example, the *wup* similarity measure of "*car*" and "*bike*" is equal to 0.82, where *depth(car)= 9*, *depth (bike)=8*, and *depth(LCS(car,bike))= depth(wheeled vehicle)=7*.

The label propagation algorithm generated 167 adverbs; $3,809$ adjectives; $11,531$ verbs; and $81,347$ nouns. Of these, $10,249$ were in the testing-EPA-lexicon. A comparison of the EPA ratings generated for these 10,249 words against those from the testing-EPA-lexicon reveals a reasonable amount of agreement between the induced and manually labelled EPA values, as indicated in Tables 5.1 and 5.2. The results also show that adjectives and adverbs were associated with the highest level of agreement. It is worth mentioning that, due to the limited numbers of adverbs in the ACT lexicon and the fact that the WordNet similarity measure compares only words of the same part of speech (POS), only a limited number of adverbs were generated using the label propagation algorithm.

Words	Testing-EPA-lexicon	LP-lexicon
Incapable (adj.)	[-1.83, -1.40, -0.54]	[-1.56, -1.18, -2.59]
Wrongly (adv.)	[-1.96, -0.22, 0.17]	[-2.02, -0.23, 0.18]
Gauge (v.)	[0.12, -0.55, 0.13]	[0.18, -1.61, 0.25]
Loser (n.)	[-1.30, -1.75, 0.30]	[-1.14, -1.52, 0.28]

Table 5.2: Words and their EPA ratings from the testing-EPA-lexicon and the LP-lexicon (LP = label propagation)

5.4 Datasets

The approach proposed in this study was evaluated using a newly collected news headline dataset. A total of $2,080$ news headlines were aggregated from a group of online newswires and news archives (BBC, CNN, Reuters, The Telegraph, The New York Times, and several other newswire sites.) from the period 1999 to 2014. Participants were recruited to use Mechanical Turk to annotate the headlines according to several criteria: being in North America, having more than 500 accepted hits, and having an acceptance rate of 90% or more. Participants were given a description of the annotation scheme (details in Appendix A) as well as 26 randomly selected headlines. They were also asked to identify the subject (actor), behaviour (verb), and object of each headline as well as their sentiment towards them in the EPA format $\in [-4.3, +4.3]$ (where -4.3 indicates a strongly negative EPA value and +4.3 indicates a strongly positive EPA value) (Appendix A).

At least three participants annotated each headline in the dataset, but any annotations with blanks, zeros, or similar values in all fields were excluded, as were answers that identified an inappropriate subject, verb, or object form (e.g., behaviour = Obama, subject = as). Also eliminated were all answers rated by fewer than three participants. This screening resulted in $1,658$ headlines, all of which had mean EPA ratings of 0.80, 1.04, and 1.02. Of these, 995 headlines had a positive evaluation score and 663 had a negative one.

The inter-rater agreement (Kappa Cohen inter-rater agreement) among the three participants ranged from 0.6 to 0.62, 0.55 to 0.60, and 0.36 to 0.39 with respect to annotating the subject, behaviour, and object, respectively (Table 3.2). The inter-rater agreement among participants in regard to rating their emotions about the subject, behaviour, and object ranged from 0.18 to 0.26 and 0.11 to 0.15 for annotating the evaluation of the subject and the object, respectively, which is considered unsatisfactory agreement. The levels of inter-rater agreement with respect to the potency and activity dimensions were even lower. These scores were obtained from a comparison of the discretized emotions (i.e., negative ≤ 0 and positive > 0). All of the entries provided by the participants for annotating the subject, verb, and object were also considered, and the average of the EPA ratings by the three annotators for each emotion rating was computed. Table 5.5 exhibits some examples from this dataset.

5.5 Evaluation

This study included an evaluation of the proposed model with respect to predicting the sentiment evoked toward the headlines as a whole based on a comparison of the discretized averaged evaluation scores of the emotions toward the subject (ε_{s_e}) and the object (ε_{o_e}) against the ground truth (examples can be found in Table 5.5). The proposed approach was assessed for a number of configurations (Table 5.3): 1) users' manually annotated triplets (ACT-UA) (i.e., subject, verb, and object); 2) parse tree triplet methods (only the subject, verb, and object) (ACT-PTT) (details in section 5.3.1); 3) the parse tree quintet approach (i.e., subject, verb, object, modifiers, and settings) (ACT-PTQ); and 4) the semantic-syntactic parser (ACT-SSQP) (Chapter 3) plus the extended affective meaning dictionary augmented by the support vector regression model (SVR) and the ACT lexicon (Chapter 4). The results were also compared against those obtained using a standard sentiment classifier (STD-classifier) that applies a voting strategy based on the frequencies of positive vs. negative words according to the SentiWordNet dictionary [53].

To evaluate the effectiveness of the ACT-based approach to fine-grained sentiment analysis of the subject and object, the estimated emotions of the subject and object were compared to the corresponding ground truth in the news headline dataset (section 5.4). For this phase of the study, root mean square error (RMSE) and mean absolute error (MAE) metrics were employed for assessing the real EPA scores $\in [-4.3, +4.3]$ as a means of establishing the ground truth. To compare the discretized EPA scores for the estimated emotions ε (i.e., positive > 0 versus negative ≤ 0) with the ground truth, an F1-measure was also computed, and the level of accuracy

Classifier	**Precision**	**Recall**	**F1-score**
ACT-SSQP	**.90**	**.82**	**.85**
ACT-PTQ	.82	.67	.73
ACT-UA	.75	.62	.67
ACT-PTT	.71	.63	.66
STD-classifier	.57	.51	.53

Table 5.3: Results of a comparison of reader sentiment (discretized evaluation score (E)) with regard to news headlines against the ground truth using ACT with different parsing techniques and a standard sentiment classier (ACT-SSQP = the semantic and syntactic quintet approach, ACT-PTQ = syntactic parse tree quintet, ACT-UA = users' annotations, ACT-PTT = parse tree triplet, STD-classifier = standard classifier built from SentiWordNet)

Emotions	**ACC**			**F1-score**			**RMSE**			**MAE**		
	E	P	A	E	P	A	E	P	A	E	P	A
ε_s	.72	.83	.89	.69	.79	.85	1.27	1.0	.95	0.99	0.78	0.75
ε_o	.85	.26	.83	.80	.14	.75	0.74	1.5	0.80	0.95	1.72	0.99

Table 5.4: Results of a comparison of the predicted emotions toward the subject ε_s and object ε_o against the ground truth (MAS = mean absolute error, RMSE = root mean square error, F1-score = F1-measure of the discretized EPA value, ACC = the accuracy of the discretized EPA value)

was examined. For the semantic-syntactic parsing approach (ACT-SSQP), the event entities of this task were extracted because, relative to the other parsing methods, this technique yielded the best event extraction performance (Table 5.4).

5.6 Results

Table 5.3 lists the results obtained from the manually annotated triplets and indicates how these outcomes yielded an F1-score of 67% compared to the ground truth. When the parse tree triplet

(ACT-PTT) method was used, the F1-score dropped to 66%. Adding adjectives (modifiers) and settings to the subject, verb, and object (ACT-PTQ) elevated the performance, resulting in a 73% F1-score compared to the ground truth. The highest F1-score of 85% was obtained using the semantic and syntactic event extraction algorithm along with the newly generated affective meaning dictionary that contains $\sim 3M$ words (Chapter 4). According to the results, the proposed approach outperformed the classifier built from SentiWordNet, which yielded an F1-score of 53% compared to the ground truth (Table 5.3).

The evaluation of the ACT-based approach with respect to modelling the reader affect toward the subject and object, as shown in Table 5.4 yielded an RMSE and MAS that were less than 1.3, and an F1-score that varied from 14% to 85% across the discretized EPA values for subject and object. The results also demonstrate that the accuracy of the scores for estimating the emotions (EPA) toward the subject were more accurate than toward the object. Overall, these findings are consistent with the results of the fine-grained event extraction methods, in which the extraction of the subject proved easier than the extraction of the object (Chapter 3). Contextually, a difference of 1.4 in the EPA space would equate to the difference between "accusing" ($\{-1.03, 0.26, 0.29\}$) and "punishing" someone ($\{0.19, 0.79, 0.76\}$), or between the identity of "mother" ($\{2.48, 1.96, 1.15\}$) and "girl"($\{1.96, 0.67, 0.99\}$), or between the emotion of "joyful" ($\{2.43, 1.97, 1.33\}$) and "euphoric" ($\{1.42, 1.09, 0.99\}$). These pairs of words seem, in an affective sense, quite close, indicating that the proposed SA method can uncover sentiments at a level that is reasonable on an intuitive level. These considerations are also evidence of the power of the new method with respect to uncovering sentiments about specific elements of sentences.

5.7 Discussion

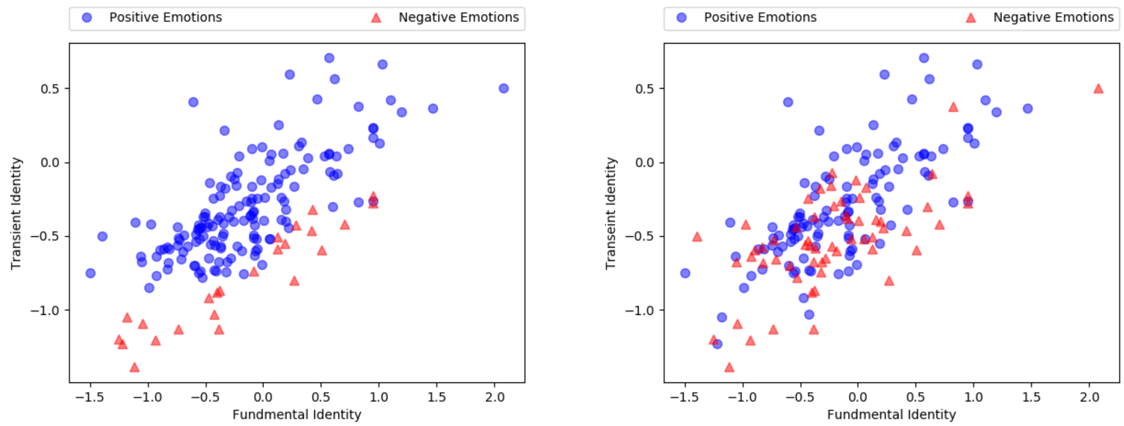
The approach proposed in this chapter entails fine-grained sentiment analysis of reader emotions triggered by event-based textual information. Using ACT, the emotions of the readers were modelled in a three-dimensional space that had been empirically and theoretically verified by several researchers [70, 15, 221, 191] to provide a comprehensive and universal representation of human emotions. ACT models emotions as a function of the difference between the pre-event (fundamental) sentiment and the post-event (transient) sentiment. Empirical survey data

Headline	d	ε_s	ε_o	ϵ_e	ε_e	ε_{s_e}	ε_{o_e}
Press sees hope in Mecca talks	2.57	happy	reverent	1.33	2.50	1.53	1.59
Brazil deploys troops to secure borders for World Cup	2.32	proud	apathetic	1.70	1.61	0.66	1.22
Gunfire injures three Napoli fans	6.80	furious	melancholy	-1.13	-0.86	-1.25	0.54
Three political *candidates* slain before Iraqi vote	11.24	furious	sorry	-1.33	-1.46	-1.80	0.05
Lily *Allen* wins web music award	2.74	proud	reverent	1.67	2.45	1.46	1.36
Finland Air *crash* kills skydivers	12.54	furious	cheerless	-1.33	-3.20	-3.0	-0.10
Bomb kills 18 on military bus in Iran	3.40	impatient	overwhelmed	-1.60	-1.23	-2.30	-0.11
Russia says 4 militants killed in Dagestan siege	11.37	furious	sorry	-0.20	-1.46	-2.34	-0.58
Baby dies after being left in car for over 8 hours	17.42	furious	overwhelmed	-1.67	-2.10	-1.57	0.06
Female *astronaut* sets record	0.79	contented	reverent	3.50	0.91	1.37	0.46

Table 5.5: ACT model results for news headlines (d=deflection, ε_s,ε_o= emotion toward subject and object, ϵ_e = evaluation of emotion toward the event (ground truth), ε_e=estimated evaluation of emotion toward an event (average of ε_{s_e}, ε_{o_e}), and ε_{s_e}, ε_{o_e} = evaluation of emotion toward the subject and object. Parse elements are coded as: *subject*, verb, object, and setting.

drove the emotional dynamics that guided the modelling of the emotions of participants in social events. The dynamics and impression-formation equations were employed for modelling reader emotions regarding event-based textual information.

As with the emotional dynamics of the interactants and their relationship with transient and fundamental sentiments (section 5.2.3), it was found that the positivity of the reader impressions toward the participants in an event (subject or object) varied greatly and were dependent on the positivity of the transient identity, with a Pearson correlation of ($r = 0.30$, $p < 0.01$). For example, if the subject held a positive identity and he/she performed an act that made him/her look negative, the reader valence toward the subject will also be negative. The intensity of the readers' emotional pleasure or displeasure toward the actor or the object was also affected by the extremity of the positive or negative situated impression of the actor or the object. It was also discovered that the dynamics of the readers' emotional potency correlated with the transient evaluation (T_e) and the transient potency (T_p), with Pearson correlations of ($r = 0.35$, $p < 0.01$ and $r = 0.16$, $p < 0.1$), respectively. The dynamics of the activity also varied with the transient evaluation and the transient potency (T_p), with Pearson correlations of ($r = 0.34$, $p < 0.01$ and $r = 0.20$, $p < 0.01$), respectively.



(a) Relationship among F_e, T_e, and ε_e

(b) Relationship among F_e, T_e, and ϵ_e

Figure 5.1: For emotions toward the subject, relationships among the fundamental evaluation, the transient evaluation, and (a) the estimated emotion evaluation (ε_e) and (b) the readers' true emotion evaluation (ϵ_e) (Here we show the emotions toward the subject).

Unlike interactant emotions, the reader valence in the developed dataset showed a negative weak correlation with the transient activity, as evidenced by a Pearson correlation of ($r = -0.14$, $p < 0.01$). There was also no significant correlation between the readers' true emotions and their fundamental identities. Despite the small number of participants in the experiment and the low level of inter-rater agreement among the annotators, many similarities were evident between the readers' true emotions and the interactants' emotions. Using the emotion equation, the readers' triggered emotion was estimated with a high degree of accuracy, as shown in Table 5.4. Figure 5.1a illustrates the relationships among the fundamental evaluation (F_e), the impression evaluation (T_e), and the emotion evaluation (ε_e) (i.e., positive > 0 and negative ≤ 0), as estimated using the emotion equations. Figure 5.1b, on the other hand, depicts the relationships among the fundamental evaluation (F_e), the impression evaluation (T_e), and the evaluation of the readers' true emotion (ε_e). The figures show how the fundamental and impression evaluations were combined to provide an estimation of the emotional response of individuals to social events, and how these estimated emotions were similar to readers' true emotions toward participants.

To evaluate the proposed approach, the sentiments arising from news headlines were chosen for analysis because they represent real-world events that describe particular single events. Analyzing sentiment in news headlines is challenging for several reasons: 1) news headlines are non-grammatically structured, and standard syntactic parsers fail to extract the events and their associated entities correctly; (2) they are written to be short and precise, providing little information to enable the effective functioning of typical bag-of-words classifiers; and (3) they are objective, containing words that might not exist in the commonly used sentiment lexicon. To overcome such challenges and facilitate fine-grained SA using ACT, the events and their associated entities were extracted based on consideration of the grammatical and syntactic structure of the headlines, the ACT lexicon was augmented using label propagation built on semantic similarity, and interactions between words in a sentence were taken into account through the computation of transient and fundamental sentiments.

The label propagation algorithm generated $96,853$ words, of which $10,249$ are in the testing dataset. In the affective space, the EPA values for these words were found to be quite close to those for their corresponding ground truths (Table 5.1). Table 5.2 also shows several examples of generated EPAs and their corresponding ground truths. The label-propagation results were further improved by an assessment of similarity between word embeddings and the employment of another similarity measure (Chapter 4). As indicated in Tables 5.4, 5.3, and 5.5 the prediction results with respect to sentiments toward events and their entities were computed using only ACT augmented lexicons, impression formation, and emotion dynamics equations. Such a simple parsimonious and theoretically well-grounded approach enables the performance of fine-grained event-based sentiment analysis that models emotions toward a subject and object in a multidimensional space and that provides a relatively high degree of accuracy.

Mapping these three-dimensional EPA scores to a specific emotion label offers detailed information about the emotions that individuals feel toward the objects and subjects within an event, a process demonstrated in Table 5.5. For example, in a sentence such as *" Russia says 4 militants killed in Dagestan siege"*, the reader will experience negative feelings and at the same time entertain different emotions toward the subject *"furious"* and the object *"sorry"*. Table 5.5 (obtained using ACT-PTQ) lists the deflection (d), the emotions toward the event (ε), and those toward the subject and object (e_a) and (e_o) for some of the examples in the dataset. Table 5.5 shows that the deflection (d) is very high when an unexpected event occurs (e.g.,*"Baby dies after being left in car for over 8 hours"*). The deflection in this sentence is high (17.42) because the EPA value of the object *"baby"* is equal to $[2.40, -2.28, 2.58]$, which is considered a quite good, quite weak, and quite active identity, and *"car"* is considered a quite positive place, with an EPA value equal to $[1.62, 1.65, 2.01]$. However, if an event took place in a war zone or if the subject has a negative evaluation, the deflection will not be very high (e.g., *"Bomb kills 18 on military bus in Iran"*), for which the deflection is equal to 3.40 because *"bomb"* has a negative evaluation. Table 5.5 demonstrates that the estimated emotions evaluation (ε_e) and the ground truth evaluation toward events (ϵ_e) are often quite close.

The above results were obtained based on the extraction of a single subject, verb, object, modifier, and setting, but they could be further improved if adverbs (e.g., *"lived happily"*), phrasal verbs (e.g.,*"get along" and "get back"*), numbers (e.g.,*"45 killed"*), and negations (e.g.,*"no more funding"*) were taken into account as well. An additional consideration is that accumulating the emotions related to multiple consequent behaviours could also be very useful. For example, in the sentence *"Man arrested after beating cops in a restaurant"* the previous behaviour *"beating"* is not taken into account and only *"arrested"* is considered at this stage. This deficiency could be addressed if complex parse trees were used and if the emotions of multiple behaviours were accumulated through consideration of the previously generated sentiment as the fundamental sentiment. A final possibility is that the use of ACT predictions to bootstrap supervised learning could also lead to enhanced performance.

5.8 Conclusion

A new direction in sentiment analysis has been proposed, whereby ACT is employed for assigning different emotions toward event-based/objective textual information and its associated entities (subject, object). Unlike the majority of SA models, which researchers have trained on highly subjective words in order to obtain descriptive labels, the new model is unique in that it incorporates ACT. ACT models emotions as points in three-dimensional space and analyzes how objective texts trigger different emotions toward event components. Using an extended affective meaning lexicon and a syntactic-semantic event extraction method, an integration of ACT and sentiment analysis was performed and assessed on real-world dataset. The proposed model was evaluated on a news headline dataset providing a higher degree of accuracy than a bag-of-word sentiment classifier. The sentiment evaluation using ACT (actors/subjects and the objects) as it pertained to news headlines was also analyzed, resulting in 72% and 85% accuracy, for the estimation of emotional evaluations of subjects and objects, respectively. These results were obtained without the performance of any supervised learning and without consideration of any consequent behaviours, phrasal verbs, or sentence negations. The results demonstrate the potential of ACT for SA use. Since ACT can also handle consequent behaviours and modifiers, future plans include the augmentation of the developed method with the addition of more complex levels of detail, the collection of more extensive datasets, and the evaluation of ACT with respect to more precise and detailed sentiments.

Chapter 6

Modelling Interactants' Emotions and Optimal Behaviour in Event-Based Articles

6.1 Introduction

Our emotional perception of the social events we observe or participate in every day has a significant impact on our judgements and reactions toward those events and any associated entities (identities or objects). The pre-event and post-event feelings of the entities involved in an event are the main drivers behind changes in individuals' reactions to and feelings about the event. The fundamental (pre-event) sentiments that people have about any entities may change or develop over the course of the event. Textual information, whether opinionated or fact-based, is also written with the intent to trigger or change individuals' opinions or emotions toward the identities/objects described. News articles, for example, represent a set of real-world or political events that impact or modify individuals' initial sentiments toward event participants. For example, individuals who feel outraged by a political event may protest or go on strike owing to the difference between their pre-event and post-event feelings. Fairy tales are another example of textual information written to influence and change individuals' emotions over the course of the story. This point can be illustrated by this manually annotated example from a fairy tale dataset [4], “*Then the door opened, and the King walked in, and there stood a maiden . . . .*

The maiden was frightened when she saw . . . .

The King looked kindly at her, stretched out his hand, and said, “Will you go with me to my palace and be my dear wife?” . . . ”

This example shows the development of the feelings of the main characters throughout the story which goes from “surprise” in the first sentence, to “*afraid*” in the second, and “*happy*” in the third.

This thesis presents a proposed approach for the performance of fine-grained emotion/sentiment analysis of event-based/fact-based documents from the perspectives of the interactants. Unlike most previously proposed sentiment analysis (SA) methods that train a supervised machine-learning (ML) algorithm on a “bag-of-words” distributional representation in order to classify textual information into predefined classes, the approach proposed here integrates socio-mathematical theory (affect control theory (ACT)), linguistic-based sentence structure analysis, and distributional word representations. Together, these constituents model human emotions associated with event-based articles in a three-dimensional space. The proposed method models the dynamics of these emotions over a series of interactions, but it also predicts the most affectively aligned behaviour each participant will take at the next event. The new methodology involves the recursive computation of the dynamics of emotions and behaviour, so that previously produced emotions or transient impressions are considered while the identities are being re-identified in the current event. For example, if an actor such as “*a teacher*”, who holds an identity that considered “good, potent, and active” performs a very negative behaviour, such as “*slap*”, toward “a very positive, weak, and active” object, in this case “*child*”, he or she will be perceived as “bad, more potent, and less active”. This outcome is the result of the fundamental identity of “*a teacher*” being modified by the negative action, and this modified identity is thenceforth used for estimating the sentiments and behaviour associated with subsequent events.

The proposed method models the entities in each event as a distribution of all of their related words, allowing consideration of the other aspects/roles of the identities and behaviours with respect to estimating pre-event and post-event feelings. For example, the evaluation, potency, and activity (EPA) of an actor described as “*a father, professional kick-boxer, and shop owner*” will be drawn from a distribution of *“father”*, “*professional kick-boxer*” and “*shop owner*”. The sampling approach was also compared to a method that uses a recursive linear combination of the words that describe the entities (i.e., identities and behaviour). The new model was evaluated using two text corpora (i.e., fairy tales and news articles), and the results generated were then compared against the ground truth [4, 254]. The results were also applied for a toy event-based story that was created in order to assess the efficacy of the fine-grained emotional modelling for simple structure sentences.

A premise of ACT [99], which is a socio-mathematical theory that models individuals' emotions and actions in social interactions, is that people's social perceptions, actions, and emotional experiences are governed by a psychological need to minimize the difference between culturally shared *fundamental sentiments* about social situations and *transient impressions* resulting from the dynamic behaviours of the interactants in those situations. The generated transient impressions progress over the course of the interaction and impact the interactants' behavioural decision-making processes according to empirically measured temporal dynamics. If the transient impression of an actor who initially holds a positive fundamental identity (e.g.,"*good, potent, and active*") is very negative as a result of performing a very negative behaviour, the new estimated identity will influence the transient impression estimation in subsequent events, that is, unless the actor performed a very positive behaviour. The interactants' emotions are also represented in a three-dimensional EPA vector (e.g., "delighted" $\equiv [2.45, 1.8, 1.62]$) [161] and modelled as a function of the *fundamental sentiments* and *transient impressions*. The optimal behaviour that each interactant will perform after any event is computed as the behaviour that minimizes the difference between the *fundamental sentiment* and the *transient impression*. In this study, the focus is on the emotional dynamics and optimal behaviour estimations within the theory. Section 2.2 provides detailed background information about ACT.

To compute the interactants' emotions and optimal behaviour, each sentence in the articles was parsed into events, and event entities (i.e., subject, verb, object, location, and modifiers) were extracted using the semantic and syntactic parsing and extraction method (Chapter 3). The *fundamental sentiment* was then estimated based on sampling from the distributions of the EPA values for each entity in the event, and the *transient impression* and the interactants' emotions were computed in a deterministic manner through existing impression-formation and emotion equations. The work involved the use of an extended three-dimensional lexicon that was generated using a pre-trained word2vec model [168] and a support vector regression (SVR) model

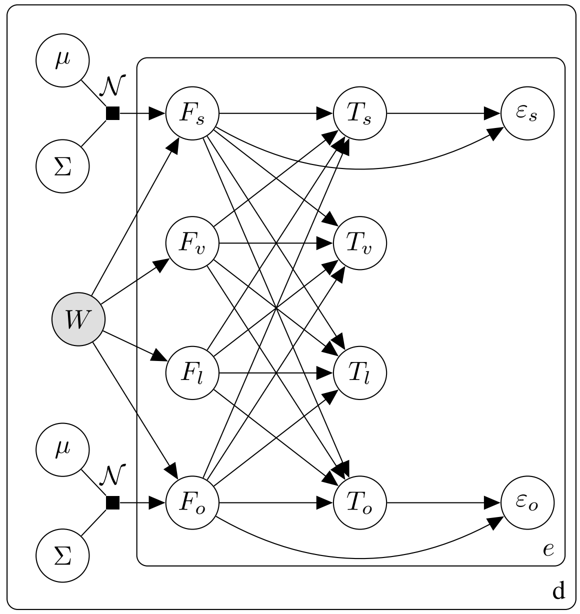


Figure 6.1: Sentiment analysis model based on affect control theory

(Chapter 4). The new ACT-based approach has shown promising results with respect to the handcrafted story, yielding an accuracy level ranging from 67% to 93%; however, the evaluation of the proposed approach for the real-world event-based sentences uncovered a number of challenges, as these sentences might describe non-behavioural events or multiple events that occur at different times.

The results of the ACT-based approach also demonstrate that estimating the evaluative factor of the emotions and the optimal behaviour is more accurate than estimating other dimensions of the affective space. It was also found that estimates of object emotions and optimal behaviours were, in general, more accurate than estimations of subject emotions and behaviours. An additional finding was that the sampling approach generated results that were very similar to those obtained using the recursive linear fusion method.

6.2 Methods

This research has resulted in the proposal of a model that computes the interactants' emotions and optimal behaviours/reactions that are produced by a single behavioural event or a sequence of such events (e.g., news articles or fairy tales). Assuming a corpus of documents $C = \{d_1, d_2, ...d_m\}$ with each document $d \in C$ containing a sequence of events $E_d = \{e_1, e_2, ...e_n\}$, each event is itself a phrase or a sentence that can be parsed using the event extraction method described in Chapter 3. Each event has at least three entities: actor (subject, s), behaviour (verb, v), and object (o). The i^{th} event of the document, for example, will be represented as $e_i = \{s_i, v_i, o_i\}$ (Figure 6.1). Events might also include the location l of the event.

Each element (term) is associated with *fundamental sentiment* that represents the commonly shared EPA toward this term (identity and behaviour) within specific cultures $F = \{f_s, f_v, f_o\}$, where for example, f_s is the fundamental sentiment of the subject(s), and each event generates a transient feeling $T = \{t_s, t_v, t_o\}$. Each element of the fundamental and transient is represented in a three-dimensional EPA vector. A nine-dimensional or 12-dimensional vector represents the fundamental and transient sentiments with or without the l, respectively.

$$F = \{f_{s_e}, f_{s_p}, f_{s_a}, f_{v_e}, f_{v_p}, f_{v_a}, f_{o_e}, f_{o_p}, f_{o_a}, f_{l_e}, f_{l_p}, f_{l_a}\}$$

where, for example, f_{s_e} represents the fundamental sentiment of the subject (s) with respect to the evaluation (e) (i.e., how nice or bad the subject is).

$$T = \{t_{s_e}, t_{s_p}, t_{s_a}, t_{v_e}, t_{v_p}, t_{v_a}, t_{o_e}, t_{o_p}, t_{o_a}, t_{l_e}, t_{l_p}, t_{l_a}\}$$

The transient impression T is computed deterministically by multiplying $\mathscr{G}(F)$, a 29-dimensional or a 34-dimensional vector, with or without the location of the event of the interactions between the terms from the *fundamental sentiment* F, by a matrix M of prediction coefficients estimated based on impression-formation research and regression analysis [100]:

$$\begin{aligned} \mathscr{G}(F) = (1 \quad & f_{s_e} \quad f_{s_p} \quad f_{s_a} \quad f_{v_e} \quad f_{v_p} \quad f_{v_a} \quad f_{o_e} \quad f_{o_p} \quad f_{o_a} \quad f_{l_e} \quad f_{l_p} \quad f_{l_a} \\ & f_{s_e}f_{v_e} \quad f_{s_e}f_{v_p} \quad f_{s_e}f_{o_e} \quad f_{s_p}f_{v_e} \quad f_{s_p}f_{v_p} \quad f_{s_p}f_{v_a} \quad f_{s_p}f_{o_e} \quad f_{s_p}f_{o_p} \\ & f_{s_p}f_{o_a} \quad f_{s_a}f_{v_p} \quad f_{s_a}f_{v_a} \quad f_{v_e}f_{o_e} \quad f_{v_e}f_{o_p} \quad f_{v_p}f_{o_e} \quad f_{v_p}f_{o_p} \quad f_{v_p}f_{o_a} \\ & f_{v_a}f_{o_p} \quad f_{s_e}f_{v_e}f_{o_e} \quad f_{s_e}f_{v_p}f_{o_p} \quad f_{s_p}f_{v_p}f_{o_p} \quad f_{s_p}f_{v_p}f_{o_a}) \end{aligned} \tag{6.1}$$

$$T = M\mathscr{G}(F) \tag{6.2}$$

Modelling a Sequence of Emotions

Fundamental and transient feelings can be used for computing the interactants' emotions ε in each time step i as well as the ultimate emotions ϵ generated from a sequence of events. The emotions at each time step ε_i are computed as proposed by [100] deterministically as a function of the fundamental identity of the subject (actor) or object, $F_r \equiv \{f_{s_e}, f_{s_p}, f_{s_a}\}$ or $\equiv \{f_{o_e}, f_{o_p}, f_{o_a}\}$, and the transient identity of the actor or object, $T_r \equiv \{f_{s_e}, f_{s_p}, f_{s_a}\}$ or $\equiv \{f_{o_e}, f_{o_p}, f_{o_a}\}$. To compute the emotions that a subject (actor) or object ε_r is experiencing after a single interaction at time step i, the difference between the fundamental identity (F_{r_i}) and the transient identity (T_{r_i}) is computed as follows:

$$\varepsilon_{r_i}(F_{r_i}, T_{r_i}) \propto E\,(T_{r_i} - I\,F_{r_i} - \delta) \tag{6.3}$$

where E is a 3×3 matrix coefficient of the emotion profile, I is a 3×3 matrix coefficient for identity, and δ is a vector of equation constants. The interactants' emotions at any time step i are computed using Equation 6.2 and 6.3 if $i = 0$ or if the identity has not been observed before. However, the transient impression of the previous event $T_{r_{i-1}}$ is used to compute the transient identity of the current event T_{r_i} instead of the fundamental sentiment if the two events e_{i-1} and e_i are relevant/related (i.e., the role identity of the subject or object has been observed before).

The ultimate sentiments/emotions of the interactants described in a sequence of events can be computed using Equations 6.2 and 6.3 with consideration to previous transient impressions. Event-based articles describe either related events or unrelated events in which the subject and/or the object of an event e_i are the same as in a previous event e_{i-1}, or the subject and the object of an event e_i are not the same as in any previous events e_{i-1}. If the object or subject of any given event e_{i+1} happens to be mentioned in a previous event, the fundamental feeling of the event e_{i+1} will be the transient feeling of e_i, but if they are different, the emotions are computed using Equation 6.3.

Optimal Behaviour

The optimal behaviour $\mathscr{V}_{i+1}$ that the subject (s) or object (o) would perform during the next time step $i+1$ can be computed in a deterministic way as proposed in [100]. In ACT, the optimal behaviour is typically regarded as the one that minimizes the unlikeliness υ of an event. The deterministic unlikeness of an event υ is the difference between the fundamental and transient:

$$\upsilon = k + \sum_{j=f_{s_a}}^{f_{o_e}} W_j (F_j - T_j)^2 \tag{6.4}$$

where k is an empirically estimated arbitrary constant, and j is the index of the elements in F and T. The behaviour can be computed analytically by calculating the partial derivative of the unlikeness υ and solving for the behaviour, as proposed in [100].

To compute the optimal behaviour that minimizes unlikeliness, as proposed by Hesie [100], the first step is to construct a vector $\mathscr{X}(F\ ,\mathscr{G}(T))$ where F is the *fundamental sentiment* and $\mathscr{G}(T)$ is a vector of the features of the terms from the *transient impression* T (Equation 6.1), but the behaviour will be recalled from memory and remains unchanged over any set of interactions.

$$\begin{aligned}\mathscr{X}(F\ ,\mathscr{G}(T)) = (&f_{s_e}\ \ f_{s_p}\ \ f_{s_a}\ \ f_{v_e}\ \ f_{v_p}\ \ f_{v_a}\ \ f_{o_e}\ \ f_{o_p}\ \ f_{o_a}\ \ f_{l_e}\ \ f_{l_p}\ \ f_{l_a}\\ &1\ \ t_{s_e}\ \ t_{s_p}\ \ t_{s_a}\ \ f_{v_e}\ \ f_{v_p}\ \ f_{v_a}\ \ t_{o_e}\ \ t_{o_p}\ \ t_{o_a}\ \ t_{l_e}\ \ t_{l_p}\ \ t_{l_a}\\ &t_{s_e}f_{v_e}\ \ t_{s_e}f_{v_p}\ \ t_{s_e}t_{o_e}\ \ t_{s_p}f_{v_e}\ \ t_{s_p}f_{v_p}\ \ t_{s_p}f_{v_a}\ \ t_{s_p}t_{o_e}\ \ t_{s_p}t_{o_p}\\ &t_{s_p}t_{o_a}\ \ t_{s_a}f_{v_p}\ \ t_{s_a}f_{v_a}\ \ f_{v_e}t_{o_e}\ \ f_{v_e}t_{o_p}\ \ f_{v_p}t_{o_e}\ \ f_{v_p}t_{o_p}\ \ f_{v_p}t_{o_a}\\ &f_{v_a}t_{o_p}\ \ t_{s_e}f_{v_e}t_{o_e}\ \ t_{s_e}f_{v_p}t_{o_p}\ \ t_{s_p}f_{v_p}t_{o_p}\ \ t_{s_p}f_{v_p}t_{o_a})\end{aligned} \tag{6.5}$$

From $\mathscr{X}$, $\mathscr{X}_v$ can be constructed to hold only the value of the behaviour and I_v, which is a vector that contains the terms in $\mathscr{X}$ that indicate identity (subject S , object O, and location L), but not the behaviour:

$$
\begin{aligned}
\mathscr{Z}_v = (&1 \;\; 1 \;\; 1 \;\; f_{v_e} \;\; f_{v_p} \;\; f_{v_a} \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \\
&1 \;\; 1 \;\; 1 \;\; 1 \;\; f_{v_e} \;\; f_{v_p} \;\; f_{v_a} \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \\
&f_{v_e} \;\; f_{v_p} \;\; 1 \;\; f_{v_e} \;\; f_{v_p} \;\; f_{v_a} \;\; 1 \;\; 1 \\
&1 \;\; f_{v_p} \;\; f_{v_a} \;\; f_{v_e} \;\; f_{v_e} \;\; f_{v_p} \;\; f_{v_p} \;\; f_{v_p} \\
&f_{v_a} \;\; f_{v_e} \;\; f_{v_p} \;\; f_{v_p} \;\; f_{v_p})
\end{aligned}
$$

$$
\begin{aligned}
I_v = (&f_{s_e} \;\; f_{s_p} \;\; f_{s_a} \;\; 1 \;\; 1 \;\; 1 \;\; f_{o_e} \;\; f_{o_p} \;\; f_{o_a} \;\; f_{l_e} \;\; f_{l_p} \;\; f_{l_a} \\
&1 \;\; t_{s_e} \;\; t_{s_p} \;\; t_{s_a} \;\; 1 \;\; 1 \;\; 1 \;\; t_{o_e} \;\; t_{o_p} \;\; t_{o_a} \;\; t_{l_e} \;\; t_{l_p} \;\; t_{l_a} \\
&t_{s_e} \;\; t_{s_e} \;\; t_{s_e}t_{o_e} \;\; t_{s_p} \;\; t_{s_p} \;\; t_{s_p} \;\; t_{s_p}t_{o_e} \;\; t_{s_p}t_{o_p} \\
&t_{s_p}t_{o_a} \;\; t_{s_a} \;\; t_{s_a} \;\; t_{o_e} \;\; t_{o_p} \;\; t_{o_e} \;\; t_{o_p} \;\; t_{o_a} \\
&t_{o_p} \;\; t_{s_e}t_{o_e} \;\; t_{s_e}t_{o_p} \;\; t_{s_p}t_{o_p} \;\; t_{s_p}t_{o_a})
\end{aligned}
$$

$\mathscr{Z}_v$ is formed by means of selection matrices S_v and G_v, where $\mathscr{Z}_v = S_v\, b + G_v$. S_v is a matrix of 34 by 3, which has ones where f_{v_e}, f_{v_p}, and f_{v_a} are located, and zeros elsewhere. G_v is a 34-vector of ones where there is no behaviour and one elsewhere. H here is a 34x34 matrix of constant parameters of W (a diagonal matrix of the weight) and M. The unlikelihood of an event will therefore be equal to:

$$
\upsilon = k + (S_v\, b + G_v) I_v\; H\; I_v (S_v\, b + G_v) \tag{6.6}
$$

Following the computation of the partial derivative and the solving for the behaviour, the optimal behaviour at time $i + 1$ that minimizes the unlikelihood of the event is computed using the following equation:

$$
\begin{gathered}
\mathscr{V}_{i+1}(F_i, T_i) = (S'_v\; I_v\; H\; S_v\; I_v)^{-1} \\
(S'_v\; I_v\; H\; G_v)
\end{gathered} \tag{6.7}
$$

The inclusion of the previously produced sentiment ε_i, or mood, in calculating the optimal behaviour can be achieved through the addition to Equation 6.7 of a vector h of the fundamental emotions ε_f and transient emotions ε_t:

$$\begin{aligned}\mathscr{V}_{i+1}(F_i, T_i, \varepsilon_i) = (S_v' \; I_v \; H \; S_v \; I_v)^{-1} \\ (S_v' \; I_v \; H \; G_v) + \omega(S_v' I_v) h\end{aligned} \tag{6.8}$$

where $h' = [\varepsilon_f, M\varepsilon_t]$ is a vector of the pre-event and post-event sentiments, M is a coefficient matrix, and ω is a constant $= 0.5$.

These equations can help predict any missing values for behaviour or identities or predict the behaviour during the next time step. In this study, optimal behaviour equations have been used for predicting the optimal behaviour that each identity (subject or object) could take during the next time step, more details can be found in [99].

Optimal Role Identity

To estimate the optimal role identity $\mathscr{R}_i$ for missing or implicit identities (subject (s) or object(o)) or to re-estimate the identity of the subject or object according to current events (e.g., a father who is physically abusive to their children should have the identity of predator), the optimal identity is computed similar to the optimal behaviour by choosing the identity that minimizes the unlikeliness of the described event. Using the analytical approach proposed by Heise [99], the optimal identity (solving for the subject S) is computed as follows:

$$\begin{aligned}\mathscr{R}_i(F_{i-1}, T_{i-1}) = (S_s' \; I_s \; H \; S_s \; I_s)^{-1} \\ (S_s' \; I_s \; H \; G_s)\end{aligned} \tag{6.9}$$

Similar to Equation 6.8 taking the emotions from the previous time step ε_{i-1} into consideration is calculated as follows:

$$\begin{aligned}\mathscr{R}_i(F_{i-1}, T_{i-1}, \varepsilon_{i-1}) = (S_s' \; I_s \; H \; S_s \; I_s)^{-1} \\ (S_s' \; I_s \; H \; G_s) + \omega(S_s' I_s) h\end{aligned} \tag{6.10}$$

6.2.1 Sampling

Instead of modelling each entity in an event as a point in the affective space (EPA), and to consider the other roles of the subject and object (e.g, "father and widower") and other aspects of a given behaviour (e.g., "killed accidentally"), each entity in the event (i.e, subject, behaviour, and object) was represented as a distribution of the EPAs of all the relevant words. Let $\{w_1, \ldots, w_d\}$ be a set of words that refers to an entity in an event (e.g., the subject), with each word represented in terms of a set of EPA values $\{x_1, \ldots, x_n\} \in \mathbb{R}^d$ independently drawn from a multivariate distribution $X \sim \mathcal{N}_k(\vec{\mu}, \Sigma)$, where $\vec{\mu}$ is a mean vector of $|k|$, k is the number of random variables (in this case the EPA), and Σ is a $k \times k$ covariance matrix with $\sigma = \{\sigma_1, \ldots, \sigma_k\}$. To obtain the EPA values for each entities in a given event, a multivariate distributions d is first generated where d is a set of words describing a single entity (e.g., *John, young, man*). Then, n points were sampled from each distribution, and the mean of all of the samples was computed to obtain the *fundamental sentiment* (F), followed by a deterministic computation of the *transient impression* (τ). To compute the emotions and the optimal behaviour, an n samples were taken from F and T to estimate ε and $\mathscr{V}$.

For example, for a set of words (e.g., *John, young, man*) that describes the subject (S), the EPA values are first obtained for each word from the extended lexicon, in this case, the English US Indiana lexicon [71], which was extended using SVR and word2vec [170] (Chapter 4). A multivariate distribution is then generated based on the EPA values of the words $\vec{\mu}$ and Σ, and n samples are drawn from each distribution. In this example, if $n = 100$, there will be a total of 300 samples to represent the EPA values of the subjects. The mean of the weighted samples is then computed in order to obtain the fundamental sentiment of the subject F_s. Values for F_b, F_o, and F_l are similarly obtained as a means of composing the *fundamental sentiment* (F).

6.3 Datasets

The proposed ACT-based model was evaluated using three datasets that describe various kinds of behavioural events: a toy handcrafted story, a fairy tale dataset, and a news article dataset. For effective testing of the model, a short story was used because it would permit an assessment of how the model would respond to simply structured sentences that describe a single event and would provide an opportunity to test it on a dataset annotated with three-dimensional EPA scores. The story was inspired by real-world events and has a total of 15 sentences. Two external judges (an undergraduate computer engineering student and a graduate computer science student) annotated the story. They were asked to annotate each event in the story by identifying the subject, verb, object, location, and modifiers, as well as the subjects', objects', and readers' emotions.

The fairytale [4] corpus comprises three sets of children's stories by different writers: Beatrix Potter, H. C. Anderson, and the Brothers Grimm. The corpus contains 176 fairy tales, each having an average of 68 sentences, and each annotated with the feelers' (main characters') primary emotions and moods. The judges annotated the primary emotions and moods according to Ekman's basic emotions [60], namely angry (A), disgusted (D), fearful (F), happy (H), neutral (N), sad (S), positively surprised (Su+), and negatively surprised (Su-). For example, the annotation for a sentence such as "*Once upon a time there was a small town*" is N:N N:N, where the first part (on the left) is the label representing the primary emotion identified by the first and the second annotators, and the second part is the label for the mood indicated by the first and second annotators, respectively. The results of our proposed method were compared against the original annotations for the primary emotions.

The goodFor/badFor dataset is part of the Multi-Perspective Question Answering (MPQA) sentiment analysis datasets and contains 134 documents from political articles that have been manually annotated to indicate how the events described positively or negatively affect (good for/bad for) the objects and what the writer's attitudes are toward the agents and the objects in the events. For example, the annotation for a sentence such as "*John helped Sally*" is the subject="*John*", the behaviour="*helped*," the object ="*Sally*", and the polarity ="*good for*". This corpus contains $1,764$ annotated good for/bad for triplets, of which 691 are good for and $1,073$ are bad for. Additional information about the dataset and the annotation scheme can be found in [56, 254].

6.4 Evaluation

To evaluate the performance of the model for the annotated datasets, the emotions and optimal behaviours of the main characters were computed after each interaction using Equations 6.2, 6.3, 6.7, and 6.8. The estimated emotions were compared to the ground truth, and the optimal behaviour, to the behaviours exhibited in the subsequent event. To compare the resultant three-dimensional emotions with the hand-coded annotations, the emotion labels in the fairy tale and news article datasets were mapped to their EPA values from the original EPA lexicon (e.g., sad $\equiv$ [-1.88,-1.46,-1.25]). The labels missing from the EPA lexicon, such as neutral, positively surprised (Su+), negatively surprised (Su-), good for, and bad for were mapped to the EPA values of "no emotions", "awe-struck", "shocked", "good", and "bad", respectively. To evaluate the optimal behaviour estimations, the subject's and object's behaviours were extracted from subsequent events and mapped to their EPA values. For example, in these two consecutive events, e_1 = " *John said hi to the elderly woman*" and e_2 = "*The elderly woman welcomed John*", the optimal behaviour of the object in e_1 is "*welcome*".

To compute the interactants' emotions and optimal behaviours, each sentence is parsed and divided into a set of events $E = \{e_1, e_2, e_3, e_4 \ldots\}$, and the subject-verb-object-setting and their modifiers are then extracted from each event e_i (Chapter 3). The fundamental F_i and transient T_i sentiments are then computed based on the extracted entities, for which two approaches were employed for computing the fundamental and transient sentiments: a sampling approach (section 6.2.1) and a recursive linear fusion (RLF) approach.

In the recursive approach, if event e_i has multiple identities and modifiers describing the subject or object, they are recursively combined using Equation 2.4 according to the word order from right to left. For example, if the subject of an event is "*Super Bowl-winning quarterback Russell Wilson divorces wife*", the model will first compute the linear combination of "*quarterback*" and " *Bowl-winning*" and then the linear combination between "*Super*" and "*Bowl-winning quarterback*". F_i and T_i are then constructed using the recursively computed identities and behaviour, and finally, the emotion ε_i is computed using Equation 6.3 and the optimal behaviour $\mathscr{V}_i$ is calculated using Equations 6.7, and 6.8.

Through the sampling approach, a distribution of the identities (subjects and objects) and the behaviours is created from the extracted entities, and samples from these distributions are used for constructing the fundamental F and transient T sentiments. For example, if an event has multiple identities and modifiers, all describing the subject (e.g., "*a poor farmer who has three children and who recently lost his wife*"), the identity of the subject will be a weighted averaged of samples drawn from a distribution of "farmer", "poor", "father", and "widower". The emotions ε_i and the optimal behaviour $\mathscr{V}_i$ are then computed by sampling from F_i, and T_i with ε_i being computed using Equations 6.3 and $\mathscr{V}_i$ using Equations 6.7, and 6.8.

The EPA scores of the extracted words were obtained from the augmented lexicon using an SVR model trained on word features that were acquired from the pre-trained word2vec model [170] and an EPA lexicon [71] (original-EPA lexicon) (Chapter 4). When a word is not contained in the augmented lexicon, the default value of the identity or behaviour is obtained, which is equal to the average of all the identities or the average of all the behaviours in the original-EPA-lexicon, respectively. For example, the default identity in original-EPA lexicon $\equiv [0.95, 1.15, 0.55]$, which is close to the EPA value of "*consultant*".

To evaluate the proposed model, the resultant emotions and optimal behaviour were compared with the ground truth based on four metrics: the τ rank correlation and mean absolute error (MAE) for comparing the real value EPA scores, and an F-1 measure and the accuracy level (ACC) for comparing the binary EPA scores (i.e., positive ≥ 0 and negative < 0).

6.5 Results and Discussion

This section presents the results of the evaluation of the proposed model with respect to three different datasets: a toy handcrafted story (*Part One*), a fairy tale dataset (*Part Two*), and a news article dataset (*Part Three*).

ID	Subject	SM	Verb	VM	Object	OM	Location
0	John	young man	visit		town	small	
1	John	walk					the street
2	John		say hi		woman	elderly	
3	woman	elderly	welcomed		John		
4	woman	elderly	introduce		John		nice places
5	woman	elderly	say goodbye		John		
6	biker	driving very fast	hit	accidentally	woman	elderly	
7	woman	elderly	fall down and injured				the street
8	John		help		woman	elderly	
9	paramedics		help -take		woman	elderly	hospital
10	policeman		arrest		biker		
11	biker		found guilty				
12	biker		apologize		woman	elderly	
13	woman	elderly	thank		John		
14	woman	elderly	invited		John		her house

Table 6.1: Manual annotations of the handcrafted story (ID≡event ID, SM ≡ subject modifiers, VM≡verb modifiers, and OM ≡object modifiers)

Part One involved the computation of the interactants' emotions and optimal behaviours exhibited in the handcrafted story, using the manually annotated subjects, verbs, objects, settings (locations), and modifiers shown in Table 6.1. Both the sampling and recursive linear fusion approaches were employed for calculating fundamental identities and behaviours for each event. Both methods provided highly correlated interactant emotions and optimal behaviours, but it was the sampling approach that yielded results that were slightly closer to the ground truth. Figure 6.2 and Table 6.2 elaborate on this.

Fusion Methods	Task	τ			F1			ACC			MAE		
		E	P	A	E	P	A	E	P	A	E	P	A
RLF	**SE**	0.29	0.19	0.31	0.75	0.53	0.80	0.80	0.67	0.87	1.58	1.71	1.25
	OE	0.43	0.22	-0.18	0.90	0.65	0.53	0.93	0.67	0.67	1.37	1.3	1.31
Sampling	**SE**	0.40	0.27	0.29	0.75	0.53	0.80	0.80	0.67	0.87	1.51	1.69	1.22
	OE	0.38	0.20	-0.24	0.90	0.59	0.53	0.93	0.60	0.67	1.34	1.31	1.30

Table 6.2: Estimated subject emotions (SE) and object emotions (OE) using the sampling and recursive linear fusion (RLF) methods

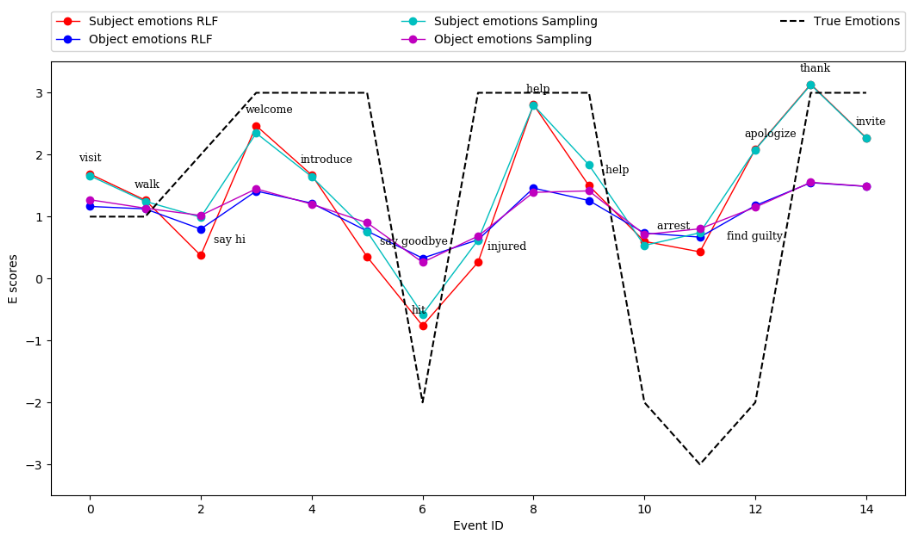


Figure 6.2: Emotional development (E) of the interactants in the handcrafted story using the sampling method and the recursive linear fusion (RLF) method

Figure 6.2 provides linear representations of the emotional development of the subject and object plotted alongside the subjects' behaviours. The figure clearly shows the high amount of agreement between the estimated emotions and the ground truth, as well as the fact that when an actor performs a negative behaviour (e.g., "*hit*"), he/she is perceived negatively, but if the actor performs a positive action after a negative act (e.g., "*apologize*") people view him/her more positively. The figure also demonstrates that the performance of a negative action affects both the subject and object as individuals, and the observers, according to ACT, tend to reconceptualize the other identities involved in the events in order to reduce the difference between the fundamental and transient sentiments. Figure 6.3 shows how the transient impression (EPA) evolves for the three main characters in the story. The figure indicates slight changes in the EPA scores of "*John*" and "*the woman*", especially following "*say goodbye*" and "*help*". The variations are more pronounced in the potency dimension: for example, after "*the woman says goodbye to John*", she is viewed as being stronger, while he is perceived as being more powerless. Moderate variations were also evident in the EPA scores for "*the biker*" after he/she performed the negative behaviour "*hit*" and the positive behaviour "*apologize*".

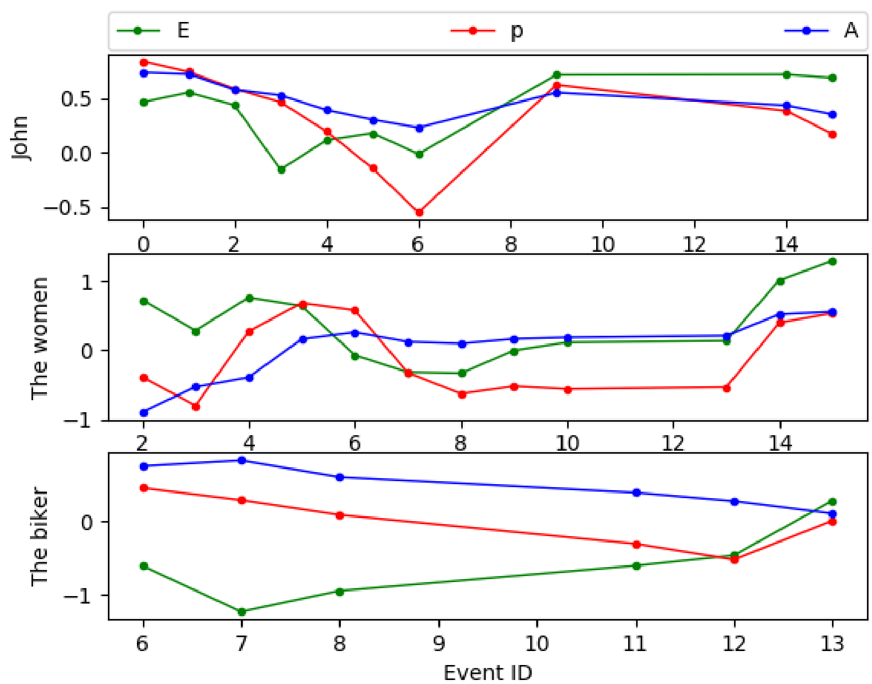


Figure 6.3: Development of the transient impression (EPA) for the three main identities in the story: John, the woman, and the biker

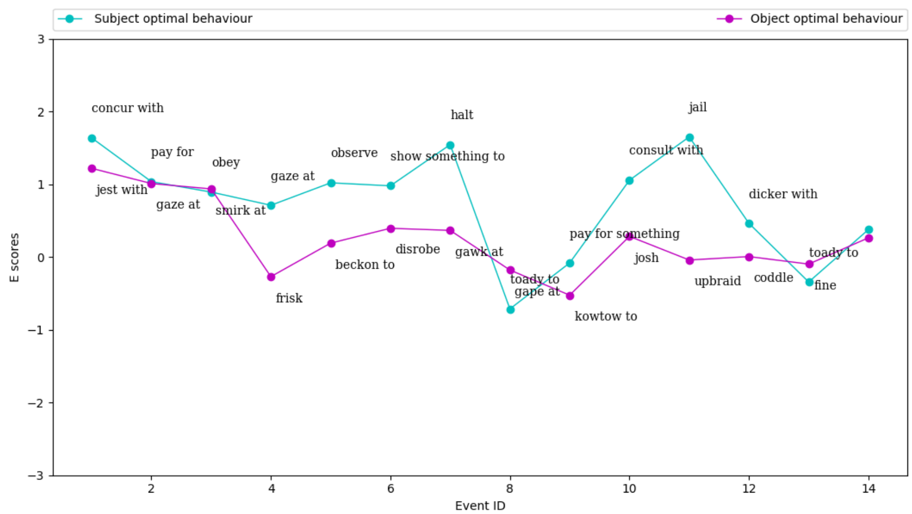


Figure 6.4: Optimal behaviour (E) of the subject and object, along with the nearest label in ACT dictionary(y-axis = the evaluation scores, x-axis = event ID in the handcrafted story

Figure 6.4 displays the results of the subjects' and objects' optimal behaviours along with the nearest label in the original-EPA-lexicon. The outcomes of the RLF and sampling procedures were relatively similar with respect to estimating the optimal behaviours, but the results shown in this figure were obtained using the sampling approach. Comparing the predicted optimal behaviour of the interactants to their behaviour in the subsequent event yielded a relatively low degree of accuracy (< 0.5), F1-score (< 0.5), and $\tau(< 0.1)$ scores (Table 6.2). These levels might be due to the choice of true labels because not all event-based articles contain very descriptive accounts of the interactions between the subject and the object. Despite the low evaluation scores, many true estimations of the interactants' reactions were observed and are worth further analysis; for example, the optimal behaviour of the subject after performing the act "*hit*" is "*halt*" while the optimal behaviour of the object is "*gawk at*", and the optimal behaviour for the policeman after arresting the biker is "*jail*".

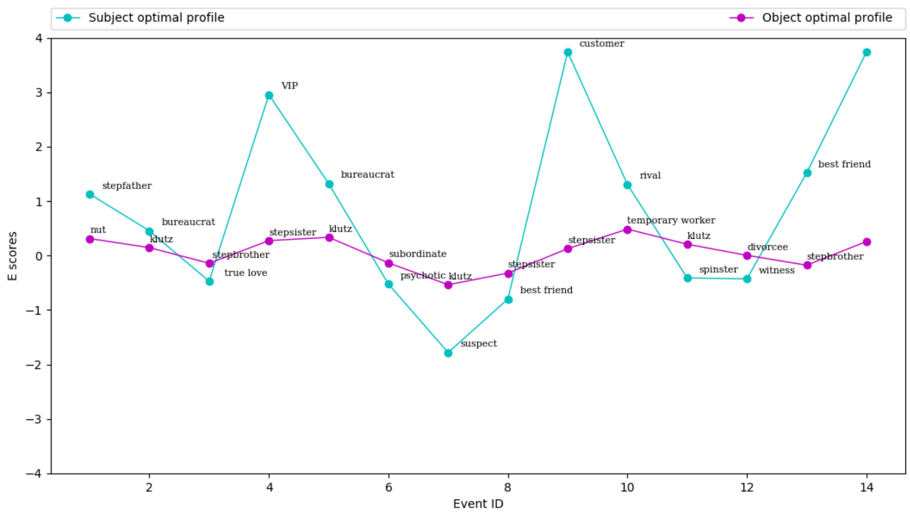


Figure 6.5: Optimal profile (E) of the subject and object along with the nearest label (y-axis = the evaluation scores, x-axis = event ID in the handcrafted story

Figure 6.5 shows the re-estimated subject and object identities based on the events described along with the labels of the nearest identity. The results of the re-estimated identity merit further analysis; however, the figure shows interesting results, such as the fact that the optimal identity of the subject "*the biker*" after performing the negative action "*hit*" is "*suspect*", while the object's re-estimated identity is "*klutz*". The optimal behaviour and optimal profile (optimal identity) were also computed based on the inclusion of the previously triggered mood or emotions in Equation 6.8, and 6.10. The results were inconsistent with the ground truth and with the optimal behaviour and optimal profile obtained without taking into account the previously triggered emotions. After further analysis, these results will be the topic of future studies.

In *Part Two*, the estimated performance of the interactants' emotions and optimal behaviours were evaluated with respect to a real-world fairy tale dataset. After the sentences in the fairy tales were parsed using the semantic syntactic parser and the event entities were extracted, the interactants' emotions and optimal behaviours were then computed. Table 6.3 reports the results, which were compared across all the sentences in the dataset against the first annotator annotations. No significant difference was evident between the performances identified by the first and

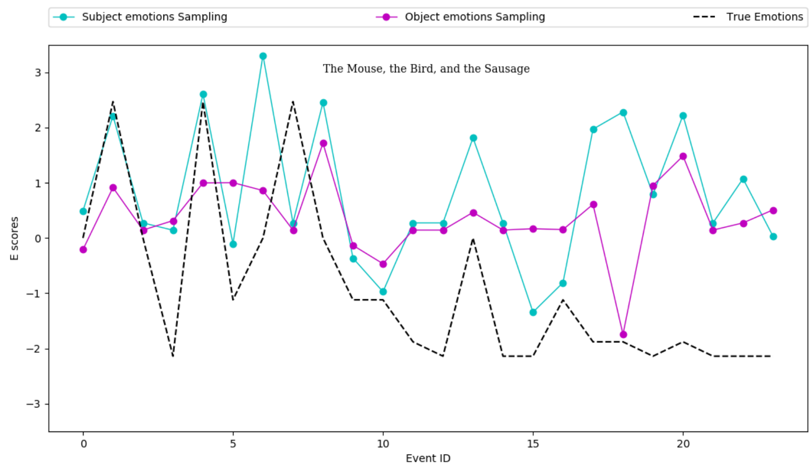


Figure 6.6: Plots of variations in the emotions predicted for the subject and object compared with the ground truth emotions (the annotation of the first annotator). The evaluation scores of the labels are compared to the estimated evaluation scores. (y-axis = the evaluation scores, x-axis = event ID in the fairy tale)

second annotators. The results of the evaluation of the performance of the proposed approach with respect to the fairy tale dataset show there was a moderate agreement between estimated emotions and their corresponding actual values with a correlation scores ($\tau < 0.3$) and accuracy scores ranging from 0.46 to 0.76 for all dimensions. The agreement scores were different, varying greatly from story to story. The highest levels of agreement ($\tau > 0.2$ and accuracy > 0.70) were obtained with stories that used a simple sentence structure to describe anthropomorphic interactions between two or three characters, such as "*The mouse, the bird, and the sausage*","*The tailor in heaven*", and "*The straw, the coal, and the bean*", while the least amount of agreement was obtained with stories that described interactions among many characters, such as "*The salad story*", "*The seven ravens*", and "*Sweetheart Roland*".

Figure 6.6 shows the estimated interactant emotions in one of the fairy tales and the ground truth scores. The figure also demonstrates the agreement between the estimated emotions and the ground truth in this story, as well as indicating that the stories generally induce "neutral" or "no emotion" feelings at the beginning, while other emotions, such as surprise and anger, might be evoked in the middle and toward the end. The results also led to the conclusion that the

Corpus	Task	F1			ACC			MAE		
		E	P	A	E	P	A	E	P	A
GFBF	SE	0.67	0.66	0.65	0.56	0.52	0.48	1.4	1.1	1.1
	SO	0.78	0.12	0.78	0.64	0.40	0.63	1.3	1.4	0.51
	SOB	0.78	0.85	0.55	0.67	0.78	0.44	0.93	0.66	0.49
	OOB	0.72	0.93	0.61	0.60	0.86	0.50	0.80	0.74	0.47
FT	SE	0.76	0.59	0.60	0.64	0.46	0.60	1.47	0.99	0.80
	OE	0.78	0.78	0.76	0.66	0.67	0.63	1.40	0.97	0.68
	SOB	0.59	0.63	0.93	0.44	0.47	0.88	1.53	1.38	0.51
	OOB	0.68	0.89	0.75	0.54	0.88	0.61	1.32	0.66	0.85

Table 6.3: Results of the comparison of interactant optimal behaviour and emotions with the ground truth in the good for/bad for (GFBF) and the fairy tale (FT) datasets (SE = subject's emotions, OE = object's emotions, SOB = subject's optimal behaviour, OOB = object's optimal behaviour, F-1 = F-1 binary score, ACC = binary accuracy, and MAE = mean absolute error)

estimates of the objects' emotions and optimal behaviours were closer to the annotations of the main characters' feelings and subsequent behaviours (Table 6.3).

In *Part Three*, for which the testing was conducted using news articles, the goodFor/badFor dataset was annotated with the subjects, behaviours, and objects, with each of these elements being either a word or a phrase. To evaluate the proposed model using this dataset, each phrase is parsed, the EPA scores of each word are extracted, and the fundamental sentiment is computed by employing a linear fusion approach using Equation 2.4 (section 6.4). The fundamental F, the transient T, the optimal behaviour $\mathscr{V}$, and the subject/object emotions ε are then estimated using Equations 6.3 and 6.7. Each sentence in the dataset is annotated with a "good for/bad for" value $\in$ {goodfor, badfor}, indicating that the event described is positively or negatively affecting the object as well as the writers' sentiment toward the subject (SE) and toward the object (OE) $\in$ {positive, negative}. The binary scores of the SE and OE were compared against their true value and the EPA score of the true behaviour was compared with the estimated optimal behaviour by employing ACT. The results show a high level of agreement between the estimated optimal behaviour and the true behaviour, with an MAE < 0.93 and an F1-score > 0.61. A comparison of the estimated sentiment toward the subject and the object with the ground truth revealed agreement between them, with evaluation scores of ACC ≥ 0.56 and F1-score ≥ 0.67 for the evaluation.

6.6 Conclusion

Most sentiment analysis studies are devoted to opinionated documents/articles, with researchers devoting less attention to event-based documents. This study has proposed a socio-mathematical-based approach that performs fine-grained three-dimensional modelling of the interactants' emotions and their affectively aligned behaviours in event-based texts. The documents are first parsed into sentences and events, followed by the computation of the interactants' emotions and optimal behaviour after each event, with the previously estimated impression also being taken into account. The estimated interactants' emotions and behaviours were evaluated using two real-world datasets and a handcrafted story composed of simple event-based sentences. The results of the performance evaluation using the handcrafted dataset yielded correlation scores of ($\tau > 0.30$) and an accuracy level of (< 80) for evaluation (E). Extraction of the event entities from the real-world datasets, firy tales and news articles, proved challenging, since they involve multiple events and conversational behaviours (e.g.,"*One small business owner says the new law's tax credits to help small business owners provide health insurance for their employees is working*"). Despite problems with the structure of the sentences and the accuracy of the annotation, the results observed were promising and exhibited a high level of agreement between the estimated emotions and behaviours and their corresponding ground truths.

Chapter 7

Summary and Future Research

Sentiment analysis (SA) has received greater attention in the last few years, as it plays a significant role in finding solutions for numerous real-world problems in a variety of contexts. In commercial settings, for example, effective SA applications can assist with the building of brand awareness for a given product on social media and with the management of its online reputation. For brands and companies that are launching new products, SA will help them obtain nearly real-time feedback about how consumers perceive the new products. SA applications can also contribute to the enhancement of consumer experience, the identification and prediction of market trends, and the creation of consumer-tailored advertising aligned with individual interests. In social science settings, SA tools can measure historical changes and shifts in individuals' opinions or sentiments, which will help with the tracking of those changes and the determination of the most influential factors that have contributed to the shifts. Similarly, in political settings, SA techniques can be used for following public attitudes over time and establishing the primary factors that impact people's opinions, thereby steering a political campaign toward success.

SA researchers have made great strides with respect to the analysis of user sentiments regarding textual information, focusing in general on opinionated and review documents, along with the use of machine-learning (ML) techniques and corpus-based frequency features. Most of the proposed ML approaches in the SA area, however, have produced black-box models that are hard to fine-tune or adapt to other types of textual information because they are trained on a specific type of annotated data. In addition to the scalability issue, most SA research has been related to unsupervised vector representations of words (count-based or distributional-based features), which might be able to model rich lexical meanings but will most likely fail to capture the sentiment information central to many SA and natural language processing (NLP) tasks. The majority of the proposed SA methods have also involved coarse-grained sentiment analyses that map a piece of text to a categorical label or to a point between two polarities. Only a handful of attempts have been made to incorporate linguistic structure and psychological theory or knowledge into statistical-based SA models, and most of these attempts have met with only modest success.

This thesis has proposed a contextual model that performs fine-grained SA of non-review and objective text and that models the emotional responses arising from this textual information in a three-dimensional space through the incorporation of features associated with affect control theory (ACT). Analysis of the sentiments emerging from non-review or event-based documents requires a deep understanding of the syntactic structure of the sentences and the event described, since they do not express an explicit opinion about a product or services and also do not contain highly opinionated words and adjectives. ACT and, more specifically, impression-formation equations represent a fruitful resource for measuring and predicting individuals' emotions and reactions toward social events, because they are theoretically grounded and have been empirically validated through extensive impression-formation studies. The socio-mathematical ACT equations are built on affective meaning, a three-dimensional representation of sentiments that is thought to provide a universal and comprehensive representation of people's sentiments. Like the compositional semantic model, which states that the meaning of a sentence is a function of its words, the empirically driven ACT equations also include consideration of the interactions between the terms that describe an event, and these factors are incorporated into the calculation of the impressions and emotions that emerge from the event. For many NLP tasks, including sentiment analysis, structured and contextual-based models that take into account the interaction between words often perform better than flat bag-of-words-based methods.

In ACT, each event is modelled as a vector of the actor, behaviour, object, and setting, and each of these elements is grounded in a three-dimensional affective meaning. To model individuals' sentiments that arise from interactions during social events or that are triggered by event-based text, an implementation of a fine-grained event extraction method that uses semantic and syntactic knowledge was performed to extract the events' entities. The developed algorithm first conducts a semantic search for the predicate of the sentence, and then, based on the extracted predicate, performs a breadth and depth search in the syntactic parse tree in order to identify the subject and object. With respect to the extraction of event components from grammatically or ungrammatically structured sentences, the semantic-syntactic event extraction method has shown results that are competitive with those produced using syntactic-based, dependency-based, and ML-based approaches.

ACT comprises a set of lexicons that measure the emotional or affective meanings associated with concepts (identities, behaviour, settings, and modifiers) in various cultures. To expand affective meaning dictionaries that contain a relatively small number of terms, this research included the development of a semi-supervised graph-based label propagation algorithm based on a similarity graph that encodes the semantic- and distributional-based similarity between words in order to propagate the affective meaning from the labelled nodes to the unlabelled nodes. The performance of the semantic- and neural-embedding-based approach was compared against several semi-supervised learning methods built on semantic, distributed, and distributional word similarities; a supervised learning approach trained on neural word embedding features; and specific state-of-the-art methods for valence-based lexicon expansion. The results showed that the performance of the semantic neural word embedding label propagation approach was better than that of the other similarity measures and state-of-the-art methods and that it approached that of the semi-supervised learning algorithm.

In ACT an emotional dynamic is computed as the difference between a culturally shared pre-event sentiment and the post-event (transient) sentiment emerging from an event. The event extraction approach, the augmented affective meaning lexicons, and the empirically driven impression-formation and emotion dynamic equations were applied for the creation of a proposed fine-grained SA approach that models readers' triggered emotions toward the actor and object in event-based sentences (e.g., a news headline). This simple and theoretically grounded approach yielded performance superior to that obtained with a bag-of-words-based sentiment classifier. Also modelled were the emotional dynamics and optimal behaviour of the interactants in event-based stories or articles (e.g., a fairy tale or news article). The emotional responses of the interactants were computed as a function of the pre-event and post-event sentiment, and an optimal behaviour was calculated as the behaviour that minimizes the difference between those sentiments. This approach included consideration of the dynamics of emotions that emerge from a sequence of consecutive events, and the identities and behaviours were modelled as distributions of the words associated with them. The proposed model was evaluated using a simple hand-crafted story, a fairy tale dataset, and a news article dataset, and despite the challenges associated with working on unstructured real-world text, the proposed approach achieved reasonable agreement with the user annotations. Interesting patterns that are worth further analysis were also uncovered with respect to emotional response development and affectively aligned behaviour.

7.1 Future Research

The approach presented in this thesis represents the first attempt to tackle the problem of modelling sentiments toward different entities (i.e., actors and objects) in an event-based text, the first to incorporate socio-mathematical ACT into an SA model, the first to include multimodal features (semantic and distributed) into a lexicon induction and label propagation algorithm, and the first to extend a multidimensional sentiment lexicon in a semi-supervised or supervised fashion. As a result, in each of the four proposed pioneering methods, a substantial number of improvements are possible, and numerous directions are available for future research.

With respect to the event extraction approach, a potential enhancement to the semantic and syntactic event extraction method would be to identify "performed" events and "non-performed" events. In the current version, the assumption is that all events are completed, when in fact, many unperformed events might exist in real-world articles (e.g.,"*He is thinking about moving to the state*", or"*He might join the army*"). The performance of the event extraction method could be further elevated by the inclusion of other lexical resources (Verbnet [239], and FrameNet [9]) for extracting the predicate of the sentence and by the utilization of semantic parsing for locating the subject and object of the sentence. The results could also be improved if multi-word verbs (e.g.," *rolls out* "), negation (e.g.,"*did not visit*"), numbers (e.g.,"*3 killed in an accident*"), and other types of phrases (e.g.,"*Our neighbour who works as an officer had an accident yesterday*") were taken into account. An interesting future direction would be the evaluation of the extraction algorithms using social media data (e.g., Twitter, Facebook, or GitHub), which can contain spelling mistakes, acronyms, abbreviations, emoticons, hyperlinks, or expressive lengthening (e.g., "*Noooooo*","*omgggg*").

With regard to the semi-supervised lexicon expansion method, several approaches can be applied for enhancing the performance of the label propagation algorithm, such as improving the distributional similarities between words by adding some supervised lexical constraints to the unsupervised word embedding learning algorithm. Another possible avenue for obtaining affective meaning ratings would be to perform a directed principle component analysis (directed-PCA), which reduces the dimensionality of the neural word embeddings based on knowledge obtained from affective meaning lexicons. Another area for future investigation could be the computation of the affective meaning (EPA) of phrases and multi-word terms (e.g.,"Chinese restaurant") using the label propagation algorithm and the similarity between neural word embeddings.

The ACT-based approach can be enhanced through the re-estimation of the coefficient matrices using real-world events aggregated from social media. It is also important to estimate the emotional dynamics of the writers, readers, and interactants for a single dataset and then to compare them in order to identify the similarities and differences associated with each of the dynamics. Many potential applications could be explored in relation to the proposed event-based SA method, which include but are not limited to the tracking of public attitudes toward an event in social media, the detection of prejudice and discrimination in news media, the estimation of event likelihood using deflection (i.e., the difference between pre-event and post-event sentiments), the predicting of trial outcomes through the analysis of case text using ACT, and the uncovering of cyberbullying in social media. The proposed model can also be used for improving the text-to-speech quality of auto-readers by incorporating the affective modelling of human emotions into the text-to-speech model. The ACT-based SA model can also be used for analyzing the emotional development of readers and interactants as a means of predicting the popularity of a story or of event-based content [311]. This model could be also employed for auto-generating children's fairy tales by, for example, giving the model two identities and creating the entire story based on the estimated interactant emotions and optimal behaviours.

Collectively, the approaches proposed in this thesis provide the basis for fine-grained and event-based sentiment analysis and contribute toward an enhanced understanding of the ways event-based text and the associated entities influence individuals' emotions and their situational impressions. Numerous directions for future investigation can be explored with respect to this still insufficiently investigated research problem, and many real-world applications could be modelled using the proposed framework.

References

[1] David Ahn. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8. Association for Computational Linguistics, 2006.

[2] Areej Alhothali and Jesse Hoey. Good news or bad news: Using affect control theory to analyze readers reaction towards news articles. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*, Denver, CO, 2015.

[3] Areej Alhothali and Jesse Hoey. Semi-supervised affective meaning lexicon expansion using semantic and distributed word representations. *arXiv preprint arXiv:1703.09825*, 2017.

[4] Cecilia Ovesdotter Alm and Richard Sproat. Emotional sequencing and development in fairy tales. In *Affective Computing and Intelligent Interaction*, pages 668–674. Springer, 2005.

[5] Alina Andreevskaia and Sabine Bergler. Semantic tag extraction from wordnet glosses. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC'06)*. Citeseer, 2006.

[6] Chinatsu Aone and Mila Ramos-Santacruz. Rees: a large-scale relation and event extraction system. In *Proceedings of the sixth conference on Applied natural language processing*, pages 76–83. Association for Computational Linguistics, 2000.

[7] Ramon F Astudillo, Silvio Amir, Wang Ling, Bruno Martins, Mário Silva, Isabel Trancoso, and Rua Alves Redol. Inesc-id: A regression model for large scale twitter sentiment lexicon induction. *SemEval-2015*, page 613, 2015.

[8] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[9] Mark Baker. Incorporation: A theory of grammatical function changing, univer-sity of chicago press, chicago. *BakerIncorporation: A Theory of Grammatical Function Changing1988*, 1988.

[10] David Bakhurst. Wittgenstein and social being. *The social self*, pages 30–46, 1995.

[11] Alexandra Balahur and Ralf Steinberger. Rethinking sentiment analysis in the news: from theory to practice and back. *Proceeding of WOMSA*, 2009.

[12] Alexandra Balahur, Ralf Steinberger, Mijail Alexandrov Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. In *LREC*, 2010.

[13] Shumeet Baluja, Rohan Seth, D Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In *Proceedings of the 17th international conference on World Wide Web*, pages 895–904. ACM, 2008.

[14] Marco Baroni, Georgiana Dinu, and Germán Kruszewski. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *ACL (1)*, pages 238–247, 2014.

[15] Lisa Feldman Barrett. Are emotions natural kinds? *Perspectives on psychological science*, 1(1):28–58, 2006.

[16] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, 7(Nov):2399–2434, 2006.

[17] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends in Machine Learning*, 2(1):1–127, 2009.

[18] Yoshua Bengio, Olivier Delalleau, and Nicolas Le Roux. Label propagation and quadratic criterion. *Semi-supervised learning*, pages 193–216, 2006.

[19] Yoshua Bengio, Holger Schwenk, Jean-Sébastien Senécal, Fréderic Morin, and Jean-Luc Gauvain. Neural probabilistic language models. In *Innovations in Machine Learning*, pages 137–186. Springer, 2006.

[20] Joseph Berger and Morris Zelditch. *New Directions in Contemporary Sociological Theories*. Rowman & Littlefield, 2002.

[21] Steven Bethard and James H Martin. Identification of event mentions and their semantic class. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 146–154. Association for Computational Linguistics, 2006.

[22] Plaban Kumar Bhowmick. Reader perspective emotion analysis in text through ensemble based multi-label classification framework. *Computer and Information Science*, 2(4):P64, 2009.

[23] Jiang Bian, Bin Gao, and Tie-Yan Liu. Knowledge-powered deep learning for word embedding. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 132–148. Springer, 2014.

[24] Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A Reis, and Jeff Reynar. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, volume 14, 2008.

[25] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100. ACM, 1998.

[26] Erik Boiy, Pieter Hens, Koen Deschacht, and Marie-Francine Moens. Automatic sentiment analysis in on-line text. In *ELPUB*, pages 349–360, 2007.

[27] Danushka Bollegala, Alsuhaibani Mohammed, Takanori Maehara, and Ken-ichi Kawarabayashi. Joint word representation learning using a corpus and a semantic lexicon. *arXiv preprint arXiv:1511.06438*, 2015.

[28] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[29] MM Bradley and PJ Lang. Affective norms for english words (anew): Affective ratings of words and instruction manual (technical report c-2), 2010.

[30] Thorsten Brants and Alex Franz. Web 1t 5-gram version 1. 2006.

[31] Scott Brave and Clifford Nass. Emotion in human-computer interaction. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, pages 81–96, 2002.

[32] Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479, 1992.

[33] Rafael A. Calvo and Sunghwan Mac Kim. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 2012.

[34] Walter B Cannon. The james-lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, 39(1/4):106–124, 1927.

[35] Liangliang Cao, Jiebo Luo, and Thomas S Huang. Annotating photo collections by label propagation according to multiple similarity cues. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 121–130. ACM, 2008.

[36] Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2), 2008.

[37] Sabrina Cerini, Valentina Compagnoni, Alice Demontis, Maicol Formentelli, and G Gandini. Language resources and linguistic theory: Typology, second language acquisition,

english linguistics, chapter micro-wnop: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Franco Angeli Editore, Milano, IT*, 2007.

[38] François-Régis Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 422–425. Association for Computational Linguistics, 2007.

[39] Danqi Chen and Christopher D Manning. A fast and accurate dependency parser using neural networks. In *EMNLP*, pages 740–750, 2014.

[40] Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 129–136. Association for Computational Linguistics, 2006.

[41] Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In *EMNLP*, volume 4, pages 33–40, 2004.

[42] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

[43] Stephen Clark. Type-driven syntax and semantics for composing meaning vectors. *Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse*, pages 359–377, 2013.

[44] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*, 2010.

[45] Allan M Collins and M Ross Quillian. Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2):240–247, 1969.

[46] Charles Horton Cooley. *Human nature and the social order*. Transaction Books, 1992.

[47] David Corney, Dyaa Albakour, Miguel Martinez, and Samir Moussa. What do a million news articles look like? In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016.*, pages 42–47, 2016.

[48] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[49] Koby Crammer and Yoram Singer. Ultraconservative online algorithms for multiclass problems. *Journal of Machine Learning Research*, 3(Jan):951–991, 2003.

[50] Lorand Dali and Blaz Fortuna. Triplet extraction from sentences using svm. *Proceedings of SiKDD*, 2008, 2008.

[51] Lorand Dali, Delia Rusu, Blaz Fortuna, Dunja Mladenic, and Marko Grobelnik. Question answering based on semantic graphs. In *Proceedings of the workshop on semantic search (sem-search 2009)*, 2009.

[52] Yan Dang, Yulei Zhang, and Hsinchun Chen. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, 25(4):46–53, 2010.

[53] Amitava Das and Sivaji Bandyopadhyay. Sentiword-net for bangla. *Knowledge Sharing Event-4: Task*, 2, 2010.

[54] Dipanjan Das and Noah A Smith. Graph-based lexicon expansion with sparsity-inducing penalties. In *Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: human language technologies*, pages 677–687. Association for Computational Linguistics, 2012.

[55] Joel R Davitz. *The language of emotion*. Academic Press, 2016.

[56] Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. Benefactive/malefactive event and writer attitude annotation. In *Annual Meeting of the Association for Computational Linguistics (ACL-2013, short pap*, 2013.

[57] Paramveer S Dhillon, Partha Pratim Talukdar, and Koby Crammer. Inference-driven metric learning for graph construction. In *4th North East Student Colloquium on Artificial Intelligence*, 2010.

[58] Xiaowen Ding, Bing Liu, and Philip S Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining*, pages 231–240. ACM, 2008.

[59] Stéphane Dupont and Juergen Luettin. Audio-visual speech modeling for continuous speech recognition. *IEEE transactions on multimedia*, 2(3):141–151, 2000.

[60] Paul Ekman. Are there basic emotions? 1992.

[61] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[62] Andrea Esuli and Fabrizio Sebastiani. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 617–624. ACM, 2005.

[63] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

[64] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A high-coverage lexical resource for opinion mining. *Evaluation*, pages 1–26, 2007.

[65] Nigel Fabb. *Sentence structure*. Routledge, 2004.

[66] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014.

[67] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4647–4657. ACM, 2016.

[68] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[69] John Rupert Firth. *A synopsis of linguistic theory, 1930-1955*. 1957.

[70] Johnny RJ Fontaine, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth. The world of emotions is not two-dimensional. *Psychological science*, 18(12):1050–1057, 2007.

[71] Clare Francis and David R Heise. Mean affective ratings of 1,500 concepts by indiana university undergraduates in 2002-3 [computer file]. distributed at affect control theory website, program interact [distributor]. 2006.

[72] Gottlob Frege. On sense and reference. *Ludlow (1997)*, pages 563–584, 1892.

[73] Nico H Frijda. Varieties of affect: Emotions and episodes, moods, and sentiments. *The nature of emotions: Fundamental questions*, pages 197–202, 1994.

[74] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. PPDB: The paraphrase database. In *Proceedings of NAACL-HLT*, pages 758–764, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

[75] Ashutosh Garg, Gerasimos Potamianos, Chalapathy Neti, and Thomas S Huang. Frame-dependent multi-stream reliability indicators for audio-visual speech recognition. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–I. IEEE, 2003.

[76] Nikhil Garg and James Henderson. Temporal restricted boltzmann machines for dependency parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 11–17. Association for Computational Linguistics, 2011.

[77] Manoochehr Ghiassi, James Skinner, and David Zimbra. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16):6266–6282, 2013.

[78] Daniel Gildea and Daniel Jurafsky. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288, 2002.

[79] Ana-Maria Giuglea and Alessandro Moschitti. Semantic role labeling via framenet, verbnet and propbank. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 929–936. Association for Computational Linguistics, 2006.

[80] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. Large-scale sentiment analysis for news and blogs. *ICWSM*, 7, 2007.

[81] Andrew B Goldberg and Xiaojin Zhu. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pages 45–52. Association for Computational Linguistics, 2006.

[82] Harry F Gollob. Some tests of a social inference model. *Journal of Personality and Social Psychology*, 29(2):157, 1974.

[83] Harry F Gollob. The subject-verb-object approach to social cognition. *Psychological Review*, 81(4):286, 1974.

[84] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. *Numerische mathematik*, 14(5):403–420, 1970.

[85] David Graff. North american news text corpus, 1995.

[86] David Graff and C Cieri. English gigaword corpus. *Linguistic Data Consortium*, 2003.

[87] Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011.

[88] Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. Concrete sentence spaces for compositional distributional models of meaning. *arXiv preprint arXiv:1101.0309*, 2010.

[89] Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. Emotion detection in email customer care. *Computational Intelligence*, 2012.

[90] William L Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. Inducing domain-specific sentiment lexicons from unlabeled corpora. *arXiv preprint arXiv:1606.02820*, 2016.

[91] John M Hammersley and Peter Clifford. Markov fields on finite graphs and lattices. 1971.

[92] Jiawei Han. *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, Amsterdam Boston, 2012.

[93] Zellig S Harris. *Distributional structure*. Springer, 1981.

[94] Vasileios Hatzivassiloglou and Kathleen R McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 174–181. Association for Computational Linguistics, 1997.

[95] Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

[96] David R Heise. Affect control theory: Concepts and model. *Journal of Mathematical Sociology*, 13(1-2):1–33, 1987.

[97] David R Heise. Understanding social interaction with affect control theory. *New directions in contemporary sociological theory*, pages 17–40, 2002.

[98] David R Heise. Sentiment formation in social interaction. *Purpose, meaning, and action: Control systems theories in sociology*, pages 189–211, 2006.

[99] David R Heise. *Expressive order: Confirming sentiments in social actions*. Springer, 2007.

[100] David R. Heise. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Wiley, 2010.

[101] David R Heise. Cultural variations in sentiments. *SpringerPlus*, 3(1):1, 2014.

[102] David R Heise and Neil J MacKinnon. Affective bases of likelihood judgments. *Journal of Mathematical Sociology*, 13(1-2):133–151, 1987.

[103] Felix Hill, Roi Reichart, and Anna Korhonen. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 2016.

[104] Geoffrey E Hinton. Distributed representations. 1984.

[105] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[106] Jesse Hoey, Tobias Schroder, and Areej Alhothali. Bayesian affect control theory. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 166–172. IEEE, 2013.

[107] Jesse Hoey, Tobias Schroeder, and Areej Alhothali. Affect control processes: Probabilistic and decision theoretic affective control in human-computer interaction. In *University of Waterloo School of Computer Science Technical Report*, number CS-2013-03, 2013.

[108] Jesse Hoey, Tobias Schroeder, and Areej Alhothali. Affect control processes: Intelligent affective interaction using a partially observable markov decision process. *Artificial Intelligence*, 230, 2016.

[109] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[110] Sheng Huang, Zhendong Niu, and Chongyang Shi. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems*, 56:191–200, 2014.

[111] Sean P Igo and Ellen Riloff. Corpus-based semantic lexicon induction with web-based corroboration. In *Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*, pages 18–26. Association for Computational Linguistics, 2009.

[112] Daisuke Ikeda, Hiroya Takamura, Lev-Arie Ratinov, and Manabu Okumura. Learning to shift the polarity of words for sentiment classification. In *IJCNLP*, pages 296–303, 2008.

[113] Julie A Jacko. *Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. CRC press, 2012.

[114] Anuja Jaiswal and Vinai George. A modified approach for extraction and association of triplets. In *Computing, Communication & Automation (ICCCA), 2015 International Conference on*, pages 36–40. IEEE, 2015.

[115] William James. What is an emotion? *Mind*, (34):188–205, 1884.

[116] Mario Jarmasz and Stan Szpakowicz. Roget's thesaurus and semantic similarity1. *Recent Advances in Natural Language Processing III: Selected Papers from RANLP*, 2003:111, 2004.

[117] Valentin Jijkoun, Maarten de Rijke, and Wouter Weerkamp. Generating focused topic-specific sentiment lexicons. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 585–594. Association for Computational Linguistics, 2010.

[118] Ms Anjali Ganesh Jivani, Ms Amisha Hetal Shingala, and Paresh V Virparia. The multi-liaison algorithm. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 2(5), 2011.

[119] Kenneth Joseph and Kathleen M Carley. Relating semantic similarity and semantic association to how humans label other people. *NLP+ CSS 2016*, page 1, 2016.

[120] Kenneth Joseph, Wei Wei, Matthew Benigni, and Kathleen M Carley. A social-event based approach to sentiment analysis of identities and behaviors in text. *The Journal of Mathematical Sociology*, 40(3):137–166, 2016.

[121] Kenneth Joseph, Wei Wei, and Kathleen M Carley. Girls rule, boys drool: Extracting semantic and affective stereotypes from twitter. 2017.

[122] Mahesh Joshi and Carolyn Penstein-Rosé. Generalizing dependency features for opinion mining. In *Proceedings of the ACL-IJCNLP 2009 conference short papers*, pages 313–316. Association for Computational Linguistics, 2009.

[123] Jugal Kalita. Detecting and extracting events from text documents. *arXiv preprint arXiv:1601.04012*, 2016.

[124] Jaap KAMPS. Using wordnet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1115–1118, 2004.

[125] Jaap Kamps and MJ Marx. Words with attitude. 2001.

[126] Kamps, Jaap and Marx, MJ and Mokken, Robert J and De Rijke, Maarten. Using wordnet to measure semantic orientations of adjectives. *European Language Resources Association (ELRA)*, 2004.

[127] Phil Katz, Matthew Singleton, and Richard Wicentowski. Swat-mp: the semeval-2007 systems for task 5 and task 14. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 308–313. Association for Computational Linguistics, 2007.

[128] Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125, 2006.

[129] Dan Klein and Christopher D Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 423–430. Association for Computational Linguistics, 2003.

[130] Dan Klein, Christopher D Manning, et al. Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10, 2003.

[131] Efstratios Kontopoulos, Christos Berberidis, Theologos Dergiades, and Nick Bassiliades. Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, 40(10):4065–4074, 2013.

[132] Terry Koo, Xavier Carreras, and Michael Collins. Simple semi-supervised dependency parsing. 2008.

[133] Zornitsa Kozareva, Borja Navarro, Sonia Vázquez, and Andrés Montoyo. Ua-zbsa: a headline emotion classification through web information. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 334–337. Association for Computational Linguistics, 2007.

[134] Zornitsa Kozareva, Ellen Riloff, and Eduard H Hovy. Semantic class learning from the web with hyponym pattern linkage graphs. In *ACL*, volume 8, pages 1048–1056, 2008.

[135] Lars Kuchinke, Arthur M Jacobs, Claudia Grubich, Melissa L-H Võ, Markus Conrad, and Manfred Herrmann. Incidental effects of emotional valence in single word processing: an fmri study. *Neuroimage*, 28(4):1022–1032, 2005.

[136] Marco La Cascia, Saratendu Sethi, and Stan Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. In *Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on*, pages 24–28. IEEE, 1998.

[137] Igor Labutov and Hod Lipson. Re-embedding words. In *ACL (2)*, pages 489–493, 2013.

[138] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[139] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[140] Richard S Lazarus. Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, 46(8):819, 1991.

[141] Claudia Leacock and Martin Chodorow. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283, 1998.

[142] Yann LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–50. Springer, 1998.

[143] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems*, pages 2177–2185, 2014.

[144] Omer Levy, Yoav Goldberg, and Ido Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225, 2015.

[145] Michael D Lewis, Jeannette M Haviland-Jones, and Lisa Feldman Barrett. *Handbook of emotions*. Guilford Press, 2010.

[146] Shoushan Li, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang. Sentiment classification with polarity shifting detection. In *Asian Language Processing (IALP), 2013 International Conference on*, pages 129–132. IEEE, 2013.

[147] Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375–384. ACM, 2009.

[148] Dekang Lin and Xiaoyun Wu. Phrase clustering for discriminative learning. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1030–1038. Association for Computational Linguistics, 2009.

[149] KH-Y Lin, Changhua Yang, and Hsin-Hsi Chen. Emotion classification of online news articles from the reader's perspective. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*, volume 1, pages 220–226. IEEE, 2008.

[150] Bing Liu. Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3):76–80, 2010.

[151] Bing Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:568, 2010.

[152] Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[153] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.

[154] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[155] Quan Liu, Hui Jiang, Si Wei, Zhen-Hua Ling, and Yu Hu. Learning semantic word embeddings based on ordinal knowledge constraints. In *ACL (1)*, pages 1501–1511, 2015.

[156] Kathryn J Lively and David R Heise. Emotions in affect control theory. In *Handbook of the Sociology of Emotions: Volume II*, pages 51–75. Springer, 2014.

[157] Hector Llorens, Estela Saquete, and Borja Navarro. Tipsem (english and spanish): Evaluating crfs and semantic roles in tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 284–291. Association for Computational Linguistics, 2010.

[158] Yao Lu, Xiangfei Kong, Xiaojun Quan, Wenyin Liu, and Yinlong Xu. Exploring the sentiment strength of user reviews. In *International Conference on Web-Age Information Management*, pages 471–482. Springer, 2010.

[159] Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 142–150. Association for Computational Linguistics, 2011.

[160] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[161] Neil J. MacKinnon. Mean affective ratings of 2, 294 concepts by guelph university undergraduates, ontario, canada. In *2001-3 [Computer file]*. 2006.

[162] Neil J MacKinnon and Alison Luke. Changes in identity attitudes as reflections of social and cultural change. *Canadian Journal of Sociology*, pages 299–338, 2002.

[163] Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. Structured models for fine-to-coarse sentiment analysis. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 432. Citeseer, 2007.

[164] Ryan T McDonald and Fernando CN Pereira. Online learning of approximate dependency parsing algorithms. In *EACL*, pages 81–88, 2006.

[165] George Herbert Mead. *The philosophy of the act*, volume 3. Univ of Chicago Pr, 1938.

[166] Albert Mehrabian. *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain Cambridge, MA, 1980.

[167] Yelena Mejova and Padmini Srinivasan. Exploring feature definition and selection for sentiment classifiers. In *ICWSM*, 2011.

[168] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *INTERSPEECH*, pages 1045–1048, 2010.

[169] Tomas Mikolov, Stefan Kombrink, Lukas Burget, JH Cernocky, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5528–5531. IEEE, 2011.

[170] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[171] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[172] Scott Miller, Jethran Guinness, and Alex Zamanian. Name tagging with word clusters and discriminative training. In *HLT-NAACL*, volume 4, pages 337–342, 2004.

[173] Jeff Mitchell and Mirella Lapata. Vector-based models of semantic composition. In *ACL*, pages 236–244, 2008.

[174] Andriy Mnih and Geoffrey E Hinton. A scalable hierarchical distributed language model. In *Advances in neural information processing systems*, pages 1081–1088, 2008.

[175] Saif M Mohammad. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741, 2012.

[176] Saif M Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion measurement*, pages 201–238, 2015.

[177] Saif M Mohammad and Peter D Turney. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34. Association for Computational Linguistics, 2010.

[178] Saif M Mohammad and Tony Wenda Yang. Tracking sentiment in mail: how genders differ on emotional axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011*, pages 70–79, 2011.

[179] Rodrigo Moraes, JoãO Francisco Valiati, and Wilson P GaviãO Neto. Document-level sentiment classification: An empirical comparison between svm and ann. *Expert Systems with Applications*, 40(2):621–633, 2013.

[180] Frederic Morin and Yoshua Bengio. Hierarchical probabilistic neural network language model. In *Proceedings of the international workshop on artificial intelligence and statistics*, pages 246–252, 2005.

[181] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In *EMNLP*, volume 4, pages 412–418, 2004.

[182] Tony Mullen and Robert Malouf. A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 159–162, 2006.

[183] Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794. Association for Computational Linguistics, 2010.

[184] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77. ACM, 2003.

[185] Martina Naughton, Nicola Stokes, and Joe Carthy. Investigating statistical techniques for sentence-level event classification. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 617–624. Association for Computational Linguistics, 2008.

[186] Kamal Nigam and Rayid Ghani. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the ninth international conference on Information and knowledge management*, pages 86–93. ACM, 2000.

[187] Zheng-Yu Niu, Dong-Hong Ji, and Chew Lim Tan. Word sense disambiguation using label propagation based semi-supervised learning. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 395–402. Association for Computational Linguistics, 2005.

[188] Joakim Nivre, Johan Hall, and Jens Nilsson. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of LREC*, volume 6, pages 2216–2219, 2006.

[189] Yoshiki Niwa and Yoshihiko Nitta. Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In *Proceedings of the 15th conference on Computational linguistics-Volume 1*, pages 304–309. Association for Computational Linguistics, 1994.

[190] Charles E Osgood. Dimensionality of the semantic space for communication via facial expressions. *Scandinavian journal of Psychology*, 7(1):1–30, 1966.

[191] Charles Egerton Osgood. *The measurement of meaning*, volume 47. University of Illinois Press, 1957.

[192] Charles Egerton Osgood, William H May, and Murray S Miron. *Cross-cultural universals of affective meaning*. University of Illinois Press, 1975.

[193] Martha Palmer, Daniel Gildea, and Nianwen Xue. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103, 2010.

[194] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.

[195] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[196] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[197] W Parrott. *Emotions in social psychology: Essential readings.* Psychology Press, 2001.

[198] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004*, pages 38–41. Association for Computational Linguistics, 2004.

[199] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[200] William Phillips and Ellen Riloff. Exploiting strong syntactic heuristics and co-training to learn semantic lexicons. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 125–132. Association for Computational Linguistics, 2002.

[201] Rosalind W. Picard. *Affective computing*. MIT press, 2000.

[202] Robert Plutchik. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, 1(3):3–33, 1980.

[203] Christopher Potts. On the negativity of negation. In *Proceedings of SALT*, volume 20, pages 636–659, 2011.

[204] Sameer S Pradhan, Wayne Ward, and James H Martin. Towards robust semantic role labeling. *Computational Linguistics*, 34(2):289–310, 2008.

[205] Vasin Punyakanok, Dan Roth, Wen-tau Yih, and Dav Zimak. Semantic role labeling via integer linear programming inference. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1346. Association for Computational Linguistics, 2004.

[206] James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. The timebank corpus. In *Corpus linguistics*, volume 2003, page 40, 2003.

[207] James Pustejovsky, Marc Verhagen, Xue Nianwen, Robert Gaizauskas, Mark Hepple, Frank Schilder, Graham Katz, Roser Saurí, Estela Saquete, Tommaso Caselli, et al. Tempeval2: Evaluating events, time expressions and temporal relations. *SemEval Task Proposal*, 2009.

[208] Delip Rao and Deepak Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics, 2009.

[209] Reinhard Rapp. Word sense discovery based on sense descriptor dissimilarity. In *Proceedings of the Ninth Machine Translation Summit*, pages 315–322, 2003.

[210] Kumar Ravi and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46, 2015.

[211] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.

[212] Ellen Riloff and Jessica Shepherd. A corpus-based approach for building semantic lexicons. *arXiv preprint cmp-lg/9706013*, 1997.

[213] Ellen Riloff, Janyce Wiebe, and William Phillips. Exploiting subjectivity classification to improve information extraction. In *Proceedings of the national conference on artificial intelligence*, volume 20, page 1106. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[214] Julie M Robillard, Areej Alhothali, Sunjay Varma, and Jesse Hoey. Intelligent and affectively aligned evaluation of online health information for older adults. In *AAAI Workshop on Health Intelligence*, San Francisco, CA, 2017. AAAI.

[215] Dawn T Robinson, Lynn Smith-Lovin, and Allison K Wisecup. Affect control theory. In *Handbook of the sociology of emotions*, pages 179–202. Springer, 2006.

[216] Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. Inducing a semantically annotated lexicon via em-based clustering. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 104–111. Association for Computational Linguistics, 1999.

[217] Ronald Rosenfeld. Two decades of statistical language modeling: Where do we go from here? *Proceedings of the IEEE*, 88(8):1270–1278, 2000.

[218] Sascha Rothe, Sebastian Ebert, and Hinrich Schütze. Ultradense word embeddings by orthogonal transformation. *arXiv preprint arXiv:1602.07572*, 2016.

[219] Herbert Rubenstein and John B Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633, 1965.

[220] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[221] James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.

[222] James A Russell. Emotion, core affect, and psychological construction. *Cognition and Emotion*, 23(7):1259–1283, 2009.

[223] James A Russell, Maria Lewicka, and Toomas Niit. A cross-cultural study of a circumplex model of affect. *Journal of personality and social psychology*, 57(5):848, 1989.

[224] James A. Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.

[225] Delia Rusu, Lorand Dali, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. Triplet extraction from sentences. In *Proceedings of the 10th International Multiconference" Information Society-IS*, pages 8–12, 2007.

[226] Delia Rusu, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. Semantic graphs derived from triplets with application in document summarization. *Informatica (Slovenia)*, 33(3):357–362, 2009.

[227] M Rushdi Saleh, Maria Teresa Martín-Valdivia, Arturo Montejo-Ráez, and LA Ureña-López. Experiments with svm to classify opinions in different domains. *Expert Systems with Applications*, 38(12):14799–14804, 2011.

[228] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[229] Inaki San Vicente, Rodrigo Agerri, German Rigau, and Donostia-San Sebastián. Simple, robust and (almost) unsupervised generation of polarity lexicons for multiple languages. In *EACL*, pages 88–97, 2014.

[230] Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. Evita: a robust event recognizer for qa systems. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 700–707. Association for Computational Linguistics, 2005.

[231] Roser Saurí, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. Timeml annotation guidelines. *Version*, 1(1):31, 2006.

[232] Stanley Schachter and Jerome Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379, 1962.

[233] Earl S Schaefer. A circumplex model for maternal behavior. *The Journal of Abnormal and Social Psychology*, 59(2):226, 1959.

[234] Klaus R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.

[235] Klaus R. Scherer, Tanja Banziger, and Etienne Roesch. *A Blueprint for Affective Computing: A sourcebook and manual*. Oxford University Press, 2010.

[236] Harold Schlosberg. The description of facial expressions in terms of two dimensions. *Journal of experimental psychology*, 44(4):229–237, 1952.

[237] Harold Schlosberg. Three dimensions of emotion. *Psychological review*, 61(2):81, 1954.

[238] T SchroĻder. Mean affective ratings of 1,100 concepts by a diverse sample of germans in 2007 [computer file]. 2008.

[239] Karin Kipper Schuler. Verbnet: A broad-coverage, comprehensive verb lexicon. 2005.

[240] Fei Sha and Fernando Pereira. Shallow parsing with conditional random fields. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 134–141. Association for Computational Linguistics, 2003.

[241] James G Shanahan, Yan Qu, and Janyce Wiebe. *Computing attitude and affect in text: theory and applications*, volume 20. Springer, 2006.

[242] Marcin Skowron. *Affect listeners: Acquisition of affective states by means of conversational systems*, pages 169–181. Development of Multimodal Interfaces: Active Listening and Synchrony. Springer, 2010.

[243] HW Smith and Y Cai. Mean affective ratings of 1,146 concepts by shanghai undergraduates, 1999 [computer file]. retrieved from the affect control theory website, program interact. 2006.

[244] HW Smith, T Matsuno, S Ike, and M Umino. Mean affective ratings of 1,894 concepts by japanese undergraduates, 1989–2002 [computer file]. retrieved from the affect control theory website, program interact. 2006.

[245] Richard Socher, John Bauer, Christopher D Manning, and Andrew Y Ng. Parsing with compositional vector grammars. In *In Proceedings of the ACL conference*. Citeseer, 2013.

[246] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, pages 926–934, 2013.

[247] Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 151–161. Association for Computational Linguistics, 2011.

[248] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer, 2013.

[249] Evgeni N Sokolov and Wolfram Boucsein. A psychophysiological model of emotion space. *Integrative Physiological and Behavioral Science*, 35(2):81–119, 2000.

[250] Pontus Stenetorp. Transition-based dependency parsing using recursive neural networks. In *NIPS Workshop on Deep Learning*, 2013.

[251] Jan Stets and Jonathan H Turner. *Handbook of the Sociology of Emotions*. Springer, 2007.

[252] Maja Stikic, Diane Larlus, and Bernt Schiele. Multi-graph based semi-supervised learning for activity recognition. In *Wearable Computers, 2009. ISWC'09. International Symposium on*, pages 85–92. IEEE, 2009.

[253] Philip Stone, Dexter C Dunphy, Marshall S Smith, and DM Ogilvie. The general inquirer: A computer approach to content analysis. *Journal of Regional Science*, 8(1):113–116, 1968.

[254] Veselin Stoyanov, Claire Cardie, and Janyce Wiebe. Multi-perspective question answering using the opqa corpus. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 923–930. Association for Computational Linguistics, 2005.

[255] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics, 2007.

[256] Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086, 2004.

[257] Amarnag Subramanya and Jeff Bilmes. Semi-supervised learning with measure propagation. *Journal of Machine Learning Research*, 12(Nov):3311–3370, 2011.

[258] Yen-Jen Tai and Hung-Yu Kao. Automatic domain-specific sentiment lexicon generation with label propagation. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, page 53. ACM, 2013.

[259] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 133–140. Association for Computational Linguistics, 2005.

[260] Partha Talukdar and Koby Crammer. New regularized algorithms for transductive learning. *Machine Learning and Knowledge Discovery in Databases*, pages 442–457, 2009.

[261] Partha Pratim Talukdar, Joseph Reisinger, Marius Paşca, Deepak Ravichandran, Rahul Bhagat, and Fernando Pereira. Weakly-supervised acquisition of labeled class instances using graph random walks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 582–590. Association for Computational Linguistics, 2008.

[262] Hristo Tanev and Bernardo Magnini. Weakly supervised approaches for ontology population. In *Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 129–143. Citeseer, 2008.

[263] Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *EMNLP*, pages 1422–1432, 2015.

[264] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*, pages 1555–1565, 2014.

[265] Egidio Terra and Charles LA Clarke. Frequency estimates for statistical word similarity measures. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 165–172. Association for Computational Linguistics, 2003.

[266] Michael Thelen and Ellen Riloff. A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 214–221. Association for Computational Linguistics, 2002.

[267] Ani Thomas, MK Kowar, Sanjay Sharma, and HR Sharma. Extracting noun phrases in subject and object roles for exploring text semantics. *International Journal on Computer Science and Engineering (IJCSE) vol-3*, 2011.

[268] Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics, 2006.

[269] Hanghang Tong, Jingrui He, Mingjing Li, Changshui Zhang, and Wei-Ying Ma. Graph based multi-modality learning. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 862–871. ACM, 2005.

[270] Kristina Toutanova, Aria Haghighi, and Christopher D Manning. Joint learning improves semantic role labeling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 589–596. Association for Computational Linguistics, 2005.

[271] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180. Association for Computational Linguistics, 2003.

[272] Joseph Turian, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394. Association for Computational Linguistics, 2010.

[273] Peter Turney. Measuring semantic similarity by latent relational analysis. 2005.

[274] Peter Turney and Michael L Littman. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. 2002.

[275] Peter Turney, Michael L Littman, Jeffrey Bigham, and Victor Shnayder. Combining independent modules to solve multiple-choice synonym and analogy problems. 2003.

[276] Peter D Turney. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proceedings of the 12th European Conference on Machine Learning*, pages 491–502. Springer-Verlag, 2001.

[277] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[278] Peter D Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37:141–188, 2010.

[279] Wiltrud KESSLER Hinrich SCH UTZE. Classification of inconsistent sentiment words using syntactic constructions. In *24th International Conference on Computational Linguistics*, page 569, 2012.

[280] Naushad UzZaman and James F Allen. Trips and trios system for tempeval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 276–283. Association for Computational Linguistics, 2010.

[281] Olga Vechtomova. Disambiguating context-dependent polarity of words: An information retrieval approach. *Information Processing & Management*, 53(5):1062–1079, 2017.

[282] Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. The viability of web-derived polarity lexicons. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 777–785. Association for Computational Linguistics, 2010.

[283] Andrew Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2):260–269, 1967.

[284] Ekaterina P Volkova, Betty J Mohler, Detmar Meurers, Dale Gerdemann, and Heinrich H Bülthoff. Emotional perception of fairy tales: Achieving agreement in emotion annotation of text. In *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, page 98, 2010.

[285] Bin Wang, Bruce Spencer, Charles X Ling, and Harry Zhang. Semi-supervised self-training for sentence subjectivity classification. In *Conference of the Canadian Society for Computational Studies of Intelligence*, pages 344–355. Springer, 2008.

[286] Fei Wang and Changshui Zhang. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, 20(1):55–67, 2008.

[287] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, 57:77–93, 2014.

[288] Meng Wang, Xian-Sheng Hua, Xun Yuan, Yan Song, and Li-Rong Dai. Optimizing multi-graph learning: towards a unified video annotation scheme. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 862–871. ACM, 2007.

[289] Sida Wang and Christopher D Manning. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 90–94. Association for Computational Linguistics, 2012.

[290] Grant Wardlaw. *Political terrorism: Theory, tactics and counter-measures*. Cambridge University Press, 1989.

[291] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

[292] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[293] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 625–631. ACM, 2005.

[294] Dominic Widdows and Beate Dorow. A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics, 2002.

[295] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.

[296] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354. Association for Computational Linguistics, 2005.

[297] Xiao-Ming Wu, Zhenguo Li, Anthony M So, John Wright, and Shih-Fu Chang. Learning with partially absorbing random walks. In *Advances in Neural Information Processing Systems*, pages 3077–3085, 2012.

[298] Zhibiao Wu and Martha Palmer. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. Association for Computational Linguistics, 1994.

[299] Wilhelm Max Wundt. *Grundriss der psychologie*. W. Engelmann, 1896.

[300] Rui Xia, Feng Xu, Jianfei Yu, Yong Qi, and Erik Cambria. Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. *Information Processing & Management*, 52(1):36–45, 2016.

[301] Chang Xu, Yalong Bai, Jiang Bian, Bin Gao, Gang Wang, Xiaoguang Liu, and Tie-Yan Liu. Rc-net: A general framework for incorporating knowledge into word representations. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1219–1228. ACM, 2014.

[302] Huong Nguyen Thi Xuan, Anh Cuong Le, and Le Minh Nguyen. Linguistic features for subjectivity classification. In *Asian Language Processing (IALP), 2012 International Conference on*, pages 17–20. IEEE, 2012.

[303] Nianwen Xue and Martha Palmer. Calibrating features for semantic role labeling. In *EMNLP*, pages 88–94, 2004.

[304] Yuto Yamaguchi, Christos Faloutsos, and Hiroyuki Kitagawa. Omni-prop: Seamless node classification on arbitrary label correlation. In *AAAI*, pages 3122–3128, 2015.

[305] Yuto Yamaguchi, Christos Faloutsos, and Hiroyuki Kitagawa. Camlp: Confidence-aware modulated label propagation. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 513–521. SIAM, 2016.

[306] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Writer meets reader: Emotion analysis of social media from both the writer's and reader's perspectives. In *Proceed-*

ings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, pages 287–290. IEEE Computer Society, 2009.

[307] Ainur Yessenalina and Claire Cardie. Compositional matrix-space models for sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 172–182. Association for Computational Linguistics, 2011.

[308] Mo Yu and Mark Dredze. Improving lexical embeddings with semantic knowledge. In *ACL (2)*, pages 545–550, 2014.

[309] Annie Zaenen and Livia Polanyi. Contextual valence shifters. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pages 106–111, 2004.

[310] Ziqiong Zhang, Qiang Ye, Zili Zhang, and Yijun Li. Sentiment classification of internet restaurant reviews written in cantonese. *Expert Systems with Applications*, 38(6):7674–7682, 2011.

[311] Aojia Zhao. Predicting popularity of fanfiction stories based on title and summary.

[312] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. *Advances in neural information processing systems*, 16(16):321–328, 2004.

[313] Xiaojin Zhu. Semi-supervised learning literature survey. 2005.

[314] Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.

[315] Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, volume 3, pages 912–919, 2003.

APPENDICES

Appendix A

News headlines Annotation Task

Part One: Survey Instructions :

This survey consists of three parts. The first part will provide you with instructions and examples to fill out the survey.
In part two, you will be asked to fill out some demographic questions that are going to be used for remuneration purposes and will remain confidential.
In part three, you will be asked to rate 26 news headlines that are collected from different news resources in time between 2000-2014.

- Each of the headlines describes at least one event (or a fact) and each event has actor (subject), behaviour (verb), and object.
- Rate your emotions toward the event in general as well as toward the actor, behaviour, and object with respect to the context in which they appear.
- Please do the rating in three stages:
  1. First, rate your emotions as a reader toward the event in general.
     - If the sentence contains sequence of events, rate your emotion toward the event in a cumulative manner.
       For example, in rating a sentence like 'A woman set her boyfriend's car on fire after he cheated on her', you should take in your consideration the previous event(cheated on her).
  2. Second, find the (actor AC., behaviour B, and object O) triple by selecting the word from the drop-down list.
     - If the sentence contains multiple behaviours, actors, and/or objects choose the main verb, actor and/or object in the sentence that describes the significant event.
       For example, in the above example the main event is 'setting the car on fire.'
     - Some statements might not have or only include implicit actor or/and object. In this case, choose either 'blank' or 'implicit' from the list.
       For example, the subject is implicit in 'the court appeal was denied' while in 'no one helped John' there is no subject.
       Note: you don't have to enter an emotion rating for the 'blank' actor or object (you can choose not applicable n/a in this case), but you still need to provide an emotion rating for the implicit actor or object.
  3. Then, determine your emotions as a reader toward each of the event's components.
     - Emotions are defined as sentiments towards an object in terms of Evaluation (E) measures good vs. bad, Potency (P) measures strong vs. weak, and activity (A) measures active vs. inactive on a scale from -4 to +4.
       For example, a policeman would normally be considered quite strong. Whereas a child would be considered quite weak.
       An example for a negative evaluation would be criminals as they are considered to be quite bad. Whereas a mother would be considered to be quite good.
       An example for active vs. inactive would be a child and a senior in which a child would be considered quite active while a senior would be considered quite inactive.

Evaluation:	Bad/Awful										Good/Nice
Potency:	Powerless/Weak	____|____	____|____	____|____	____|____	____|____	____|____	____|____	____|____	____|____	Powerful/Strong
Activity:	Inactive/Quiet										Active/Lively
		-4	-3	-2	-1	0	+1	+2	+3	+4	
		Infinitely	Extremely	Quite	Slightly	Neutral	Slightly	Quite	Extremely	Infinitely	

- Here's an example of the rating :

Sentence	Sentiment toward event	Actor	Sentiment toward Actor	Behaviour	Sentiment toward behaviour	Object	Sentiment toward Object
A criminal killed two policemen	E=-3.0 P=1.8 A=1.4	Criminal	E=-2.10 P=0.06 A=0.54	Killed	E=-3.33 P=1.32 A=0.65	Policemen	E=1.23 P=1.87 A=0.78

- Note: in this example the rating for subject and object are not what you are generally feeling towards criminals or policemen but towards these particular policemen and this particular criminal.

Part Tow: Demographic Information:

* required field.

Please fill the information below and click Agree to proceed to the next page
Name: *
Email: *
Gender: Male: ○ Female: ○ *
Level of Education: Select.. *
☐ I have read and agree to the above instrctions *

Submit

Part three: Text and Sentiment Annotations:

Show Instructions

Note: Please rate the following statements appropriately as they will be read and judged before approval/rejection.

#	News Headline	Emotions toward event	Actor	Emotions toward actor	Behaviour	Emotions toward behaviour	Object	Emotions toward object
1	Scotland votes to stay in UK	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
2	Guam Schools See Dramatic Decline in Bullying	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
3	Teenage girls abducted in Nigerian school raid by 'Boko Haram' gunmen	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
4	276 kidnapped Nigerian schoolgirls' plight gripping country	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
5	U.S. FCC chief details pledge to police Internet 'fast lanes'	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
6	The sordid legal battles of LA Clippers owner Donald Sterling laid bare	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
7	Islamist president says Egypt won't reverse course	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
8	Naked man doing push-ups in middle of busy road struck and killed by car	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
9	Israel says hopes to keep Lebanon border calm	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
10	New York's Giuliani Makes Crime His Issue	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
11	US STOCKS-Futures flat a day after Dow hits record close	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
12	Bombs wound three policemen in Cairo	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
13	Japan warns China over 'airspace violations'	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
14	North Dakota abortion ban overturned	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
15	Republicans name seven to U.S. House panel on Benghazi attacks	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
16	Five Killed In University House Party Stabbing	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
17	One dead, 100 Hurt in Anti-China Riot in Vietnam	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
18	One dead and several missing after Congo boat capsizes	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
19	Germany, Japan say G7 won't waiver on further Russian sanctions	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
20	Car blast kills two in Bahrain	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
21	Social media campaign for kidnapped girls reaches 1m	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
22	Two meals a day 'can treat diabetes'	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
23	Former NSA contractor Snowden expects to remain in Russia	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
24	US senators remove requirement for disclosure over drone strike victims	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
25	Kashmiris wary as Modi challenges for power in India	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:
26	Argentina's options shrink after U.S. debt ruling	E: P: A:	Ac.: Select	E: P: A:	B: Select	E: P: A:	O: Select	E: P: A:

Submit Clear