Polymerization Data Mining: A Perspective

Yousef Mohammadi[1*] and Alexander Penlidis[2*]

[1]Petrochemical Research and Technology Company (NPC-rt), National Petrochemical Company (NPC), P.O. Box 14358-84711, Tehran, Iran

[2]Department of Chemical Engineering, Institute for Polymer Research (IPR), University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

To whom correspondence should be addressed:
Dr. Yousef Mohammadi: mohammadi@npc-rt.ir
Prof. Alexander Penlidis: penlidis@uwaterloo.ca

State of affairs

Polymers, gigantic natural and synthetic molecules with a vast variety of intricate micro-molecular and architectural characteristics, play key roles in our daily life. This class of materials is not only massively utilized for general purposes, like packaging and consumer products, but also frequently applied to almost all hi-tech engineering applications. Nowadays, elaborate manipulation of microstructural features to produce very specific macromolecules with outstanding final properties is of great importance to address and meet ever-increasing market demands for smart materials and technologies. To achieve this, 'pioneer' companies attempt to move forward from passive to dynamic and even adaptive selection, design and manufacturing of polymeric materials. More interestingly, leading research centers have stepped into new grounds enlivening the 'futuristic' dream of living and thinking synthetic macromolecules [1,2].

Over the past few decades, some cutting-edge techniques, for instance, controlled living radical (co)polymerizations [3,4], click polymerizations [5], and chain shuttling reactions [6] capable of synthesizing sequence-controlled macromolecules have been proposed and successfully put into practice. The newly developed techniques have been widely applied by several leading companies and many internationally recognized research groups to welcome the advent of ‘advanced macromolecules’. However, an important question arises: Is the development of novel sophisticated polymerization mechanisms/systems the only appropriate option on the table to address all rapid growth demands of the market for specifically tailored polymeric materials with desirable properties?

The question can be properly addressed considering the amount of data on macromolecular chemistry generated daily, and subsequently reported and published worldwide by many different companies and research groups in lab-, pilot- and industrial-scales. Is ‘big data’ appropriately processed to enhance our knowledge on various aspects of the complex world of macromolecules? Like in many other disciplines, the answer is the same (and quite clear). The fact is that a huge amount of data on polymerization kinetics, macromolecular reaction engineering, process monitoring, and characterization of polymers is routinely generated and reported, which is not thoroughly and effectively analyzed applying advanced data processing/refinement tools and techniques. Some claim that about 60-75% of the data points/information collected is ignored (not analyzed or taken into account). Hence, managing the explosive growth of data not only in polymerization systems but also in all other aspects of human endeavors is a perplexing challenge. In fact, we are drowning in data, but we are still starved for knowledge that would help our diagnostic skills.

Data mining

Nowadays, ‘Data mining’ is widely proposed by data scientists as the most accepted and powerful approach to properly handle the information explosion. Data mining is defined as the extraction of interesting patterns and knowledge from huge amounts of data. It should be noted that the word ‘interesting’ refers to ‘non-trivial’, ‘implicit’, ‘previously unknown’, and ‘potentially useful’. Generally, data mining projects are composed of three essential steps including data pre-processing, processing, and post-processing. The first step, i.e. data pre-processing, is mostly applied for data cleaning, data integration, data transformation, and also dimensionality reduction. Data processing, the heart of all data mining projects, results in knowledge discovery as the main outcome of data mining, applying powerful modeling and optimization techniques.

Post processing, the last step of data mining, is mostly employed to appropriately interpret, visualize, and present the processed outputs.

The main functions of data mining are generalization, pattern discovery, classification, clustering, outlier analysis, time and ordering (sequential pattern, trend, and evolution analysis), and structure/network analysis. Data mining is the confluence of multiple disciplines including Statistics, visualization technology, high-performance computing, database technology, algorithm design, machine learning, and pattern recognition, with a wide variety of applications. It is mostly due to (1) a tremendous amount of data being generated (i.e. 'big data'), (2) the high-dimensionality of data, (3) the high-complexity of data, and (4) the emergence of new novel and sophisticated applications. Today, data mining has been implemented and applied over a vast range of applications, like web page analysis, market basket analysis, fraud and intrusion detection, banking, telecommunication, customer relationship management, bioinformatics, educational technology, software engineering, criminal investigation, medical and health systems, text analysis, voice recognition, social and information networks, and the analysis of large amounts of unstructured information in the oil and gas industry.

Polymerization data mining, like in other disciplines, can be considered as the measurement, collection, analysis, and reporting of data about polymerization systems for purposes of understanding, controlling, and optimizing macromolecular reactions and the environments in which they occur. In fact, polymerization data mining is an effective and intelligent processing/analysis of massive datasets frequently generated in polymerization systems.

In general, for all macromolecular reaction engineering projects, several polymerization recipes are predefined applying experimental design techniques first. Then, the polymerization processes are separately performed for each recipe. Afterwards, the produced macromolecules are precisely analyzed applying available experimental techniques to determine their micromolecular characteristics and also final properties. The microstructure and architecture of the synthesized chains is precisely quantified by well-defined micromolecular indices either as average or distributional properties. Also, the final properties including chemical, physical, thermal, mechanical, optical, and/or biological properties determine the appropriateness of the produced macromolecules in different applications. Undoubtedly, understanding the intricate interrelationships between polymerization recipe, microstructure, and ultimately the polymer properties is the key to tailor-make complex macromolecules. Hence, the ultimate goal of polymerization data mining is to 'crack' the complexity of recipe-architecture-property interrelationships via masterful processing of the collected data.

A bit of history

Over the past decades, classical computational techniques as the most available processing tools have frequently been applied to polymerization systems. Considering the stochastic nature of macromolecular systems/processes, classical stochastic mathematical tools including the method of Moments, method of Markovian chains, and Monte Carlo methods have mostly been employed to handle modeling, simulation, and optimization in macromolecular processes. In fact, Statistics and Probability have been considered as the main computational framework to monitor, predict, and fine-tune the quality of the produced polymers. Whatever occurs in macromolecular systems, however, is so perplexing to be thoroughly grasped merely by classical techniques. In other words, classical computational tools are not powerful enough to effectively challenge extremely complex macromolecular systems and appropriately handle the generated information and collected data in such perplexing systems. The huge numbers of macromolecular species involved and the corresponding distributional properties and architectures make these days the use of classical techniques rather limited.

Basically, there exist two main data generation sources in polymerization systems resulting in experimental and theoretical datasets. In experimental datasets the information is either collected during the course of polymerization (mostly on polymerization kinetics) or reflects the outcomes of all experimental characterization analyses/measurements on synthesized macromolecules and/or the corresponding final products. For instance, the reaction time, the feeding policies and reactants' consumption rates, the conversion, and the rate of incorporation/propagation reaction channels are some variables monitored and reported as useful information during the course of polymerization. On the other hand, molecular weight distribution, sequence length distribution, branching density/frequencies (and distribution), melt flow properties, infrared spectra, and molecular weight averages/polydispersity index are some typically measured variables applying well-developed techniques like GPC, NMR, and FTIR to uncover micro-molecular (micro-structural) characteristics, while some others like crystallization, glass transition temperature, melting point, solubility, filmability, phase separation patterns, creep and stress relaxation plots, roughness, and rheological behavior are distinguished instances of the experimental datasets measured/collected to quantify the final properties of the produced macromolecules.

In contrast, theoretical datasets are mostly gathered applying either available algebraic/differential equations or in-house modelers/simulators along with commercial software packages. Nowadays, there exist many mathematical equations to study the polymerization kinetics and the micro-molecular and final properties of produced macromolecules. For instance, in macromolecular reaction engineering, the most widely accepted approach for kinetic studies and determination of average micro-molecular characteristics is the method of Moments. Also, many in-house modelers and simulators have been developed mostly based on classical stochastic mathematical models, e.g. molecular simulation techniques like Molecular Dynamics and Kinetic Monte Carlo approaches, to study polymerization systems and polymer processing properties in detail. Furthermore, several software packages have been successfully commercialized to model/simulate the polymerization kinetics/processes (e.g. PREDICI and Aspen Polymers) or predict the processability and final properties of produced macromolecules (e.g. ABAQUS, ADF Modeling Suite, LAMMPS, BIOVIA Materials Studio, COMSOL Multiphysics, ANSYS, and CheFEM).

Both experimental and theoretical datasets can be of different types, including numerical, text, graph (e.g. stress-strain curves, FTIR spectra), image (e.g. SEM images), video (e.g. the nano-/micro-structure evolution during phase separation), etc. Among all, numerical data are the simplest ones to handle. Other data types should be decoded/translated into standard formats first in order to be understood by the processing units. This can be managed by developing/applying appropriate image/video processing techniques and/or text/pattern recognizing algorithms. Both powerful image/video processors and text/pattern recognizers can be designed/established applying well-developed approaches based on Computational Intelligence techniques like Deep Learning algorithms. After being well-trained, they will be capable of effectively decoding complex images, patterns, spectra, plots, etc. as the main outputs of polymerization processes and/or polymer characterization/processing into comprehendible datasets. The tabulated numerical datasets are the most acceptable types which can be of different discrete or continuous attributes, including binary, nominal, ordinal, and numeric/quantitative interval-scaled and ratio-scaled.

What we can do currently

After translating the raw data into standard formats and extracting the most important information, the obtained datasets are pre-processed by a Data Refinery unit to enhance the quality of the experimental/theoretical datasets before being challenged by the processing unit. Pre-processing consists of different statistical techniques/modules, as follows [7-9]:

1. 'Data cleaning': Data cleaning handles missing data, smoothes noisy data, identifies or removes outliers, and resolves inconsistencies. Real-world data points are 'dirty' as lots of potentially incorrect data are generated daily due to faulty measuring processes/instruments, personal biases, human/computer errors, and transmission errors. Missing data may be observed due to equipment malfunction, deletion because of being inconsistent with other recorded data, not considered at the time of entry, not entered because of misunderstanding, and not registered/updated history or changes of data. Noisy data as random errors or variances in a measured variable, however, may occur due to faulty data collection instruments, data entry problems, data transmission problems, technology limitations, duplicate records, incomplete data, and inconsistent data. Binning, regression, clustering, and semi-supervised (combined computer and human inspection) methods are the most important techniques to handle noisy data.
2. 'Data integration': Data integration as a data pre-processing technique is responsible for combining multiple databases, data cubes, or files, and providing a unified view of these data.
3. 'Data reduction': Dimensionality reduction, numerosity reduction, and data compression are performed by data reduction module. Data reduction is mostly applied whenever a dataset may store terabytes of information and/or when a complex analysis may take a very long time to run on the complete dataset. Nowadays, there exist several methods for data reduction including

regression, data cube aggregation, data compression, histograms, clustering, and sampling techniques. The main advantages of data reduction can be summarized as avoiding the ‘curse of dimensionality’, helping eliminate irrelevant features and reducing noise, optimizing time and space required in data mining projects, and allowing easier presentation/visualization of data mining outcomes.

4. ‘Data transformation’: The data transformation module maps the entire set of values of a given attribute to a new set of replacement values so that each old value can be identified with one of the new values. Smoothing, attribute/feature generation, aggregation, normalization, and discretization are the most important methods widely applied for data transformation. Also, binning, histogram analysis, clustering analysis, decision-tree analysis, and correlation analysis are the most popular techniques for data discretization.

Principally, accuracy, completeness, consistency, timeliness, believability, and interpretability are the main indices utilized to measure the quality of pre-processed data before entering into the processing unit.

Processing is the most important unit in polymerization data mining projects. Pattern discovery, classification, clustering, outlier analysis, association and correlation are the main functions of the processing unit [10]. In other words, modeling and optimization of received datasets are handled in the processing unit. Modeling ‘cracks’ the complex interrelationships between input variables and corresponding responses/outputs in a given dataset, while optimization explores and returns the optimal solution(s) capable of satisfying predefined target(s). Over the last decades, a vast variety of classical deterministic and stochastic modelers and optimizers have been utilized to process the collected information in polymerization systems. However, the advent of novel polymerization mechanisms/systems along with new developments and applications of novel monitoring and characterization instruments/equipment to address the market’s increasing demands to tailor-make engineering polymeric products, have made it difficult, if not impossible, to appropriately process a large volume of data of different types continuously generated and collected. Undoubtedly, more powerful, robust, and versatile data processing techniques are required to effectively model and/or optimize experimental and theoretical datasets in novel polymerization systems.

In the past few years, effective modelers as the main data processors in almost all data mining projects have been mostly established based on the implementation and application of classification and clustering techniques. Classification is defined as a supervised learning, i.e. learning by example, while clustering is considered as an unsupervised learning or learning by observation. In the former case, the training datasets including observations, measurements, etc. are accompanied by labels indicating the class of the observations, while in the latter case the class labels of training datasets are unknown. A cluster is a collection of data objects that are similar to one another within the same group and dissimilar to the objects in other groups. Decision tree algorithms, rule-based and pattern-based classification methods, Bayesian classification methods, lazy learning and active learning techniques, and also support vector machines (SVM) are the most popular classifiers capable of extracting powerful models to describe important data classes [11]. The main clustering techniques consist of partitioning methods (e.g. K-means and K-medoids algorithms), hierarchical methods (especially BIRCH and CHAMELEON algorithms), density-based methods (e.g. DBSCAN, OPTICS, and DENCLU algorithms), and grid-based methods (e.g. STING and CLIQUE algorithms), which are capable of finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters [12].

Nowadays, Computational Intelligence techniques are emerging as serious competitors to well-established classical modeling approaches [13]. They have recently been implemented and applied in a large variety of disciplines as unique solutions. Computational Intelligence-based classifiers and clusterers are able to model/optimize all received datasets of any type, size and complexity with no need to take into account common simplification assumptions of classical modeling. All Artificial Intelligence techniques enjoy essential components of 'intelligence' including learning, generalization, and decision-making, for modeling and optimization of complex nonlinear problems. Artificial Neural Networks (ANNs) and fuzzy logic systems are very powerful intelligent modelers, while the most popular intelligent optimizers include swarm intelligence, simulated annealing, particle swarm optimization (PSO), and genetic algorithms (GAs).

Conceptual examples

As can be observed in Figure 1, the refined datasets can be properly processed with Artificial Intelligence-based classifiers/clusterers to establish intelligent modelers. Although the responses can be intelligently modeled with a single ANN or fuzzy logic system altogether, it is recommended to model each response with a separate intelligent modeler. The developed intelligent modeler(s) either can be directly sent to the post processing unit to decode/visualize the complex interrelationships between input variable and responses or can be hybridized with an appropriate optimizer. In the former case, the modeler can be used to either predict the outputs for any given set of input variables or represent the variations of responses via 2D graphs or 3D surfaces/contour plots. In the latter case, however, the modeler is in synergistic interplay with an appropriate optimizer in an attempt to find optimal solutions capable of satisfying preset target(s). To amalgamate the established intelligent modeler(s) with potential optimizers, a well-designed communicator capable of appropriately interconnecting the modeler and optimizer should be developed. In fact, the synergistic interplay between the modeler and optimizer is taking place through the communicator. Whenever the optimizer requires recalling the modeler, the encoded information is sent to the communicator first and then translated and forwarded to the modeler. Afterwards, the intelligent modeler handles the modeling process for the received information and returns the outcomes. The processed data are translated again by the communicator to be comprehended by the optimizer.

As mentioned earlier, there exist several Artificial Intelligence-based optimization techniques to establish an effective intelligent optimizer. Contrary to classical optimizers, which mostly make use of random or exhaustive search strategies, intelligent optimizers are equipped with powerful heuristic evolutionary search algorithms. They are population-based optimizers utilizing stochastic intelligent exploitation and exploration operators. Considering the multi-dimensionality of almost all polymerization systems, professional intelligent optimizers should be inevitably implemented and applied to precisely handle multi-objective optimization problems. Among all proposed intelligent optimizers, Non-dominated Sorting Genetic Algorithm (NSGA-II), the multi-objective version of Genetic Algorithms, is one of the most popular and powerful intelligent processors to effectively handle a vast variety of optimization problems in polymerization data mining. Briefly, the input variables to be optimized are encoded into a chromosome-like structure first. Then, a preset number of chromosomes (as initial population) are generated in a stochastic manner. The degree of goodness of each chromosome is separately determined recalling the intelligent modeler(s). Having evaluated the fitness of each chromosome, the optimizer evolves the potential solutions towards the global optimum applying selection, mating, crossover, and mutation operators as intelligent genetic manipulators. Obviously, the modeler(s) is responsible for precisely translating the genotypes (i.e. the chromosomes transferring the encoded input variables) into phenotypes (i.e. the corresponding responses). The outcomes, mostly as tabulated optimal Pareto fronts, are then transferred to the post-processing unit for further evaluation, quantitative analyses, and decision making.

As mentioned above, the available equations, modelers, and simulators specifically proposed/developed to study polymerization systems and polymer characterization/processing can be employed in an offline mode to generate theoretical datasets required for polymerization data mining purposes. In fact, in the offline mode, theoretical datasets are produced first. Then, they are utilized for knowledge discovery applying the polymerization data mining techniques described above. Micromolecular landscape of olefin block copolymers, for instance, has been comprehensively patterned and reported, amalgamating a well-established Kinetic Monte Carlo (KMC) simulator and several well-trained Artificial Neural Networks (ANNs) [14]. To put this concept into practice, theoretical data on chain microstructure have been obtained by an in-house KMC simulator first. Then, the complex interrelationships between microstructure and polymerization recipes of chain shuttling copolymerization of ethylene with a-olefins have been disclosed constructing several ANNs in an offline mode. Furthermore, a new paradigm for inverse macromolecular engineering has recently been proposed developing a hybrid intelligent data processing technique [15]. In reference [15], the established molecular simulator has been applied to calculate/predict the microstructural features of all predefined virtual scenarios/experiments separately, and the results have subsequently been employed to train several intelligent modelers (ANNs). Obviously, the chain shuttling reaction simulator and the intelligent modelers interact in an off-line manner. Then, a well-designed communicator has been utilized to interconnect the trained intelligent modelers, as computationally cost-effective versions of the molecular simulator, and the developed heuristic optimizer (NSGA-II), in an attempt to discover optimal recipes capable of suggesting OBC chains having predefined micromolecular structures. It is worth mentioning that both 'algebraic/differential equations' and 'commercial/in-house simulators/modelers' can be utilized in an online mode as well. In this case, the available theoretical/empirical equations can be directly recalled by the intelligent optimizer. Recently, an intelligent search strategy based on the NSGA-II technique has been successfully implemented and examined to heuristically translate microstructural patterns to optimal copolymerization recipes/operating conditions in the case of metallocene-based copolymerization of ethylene with α-olefins containing multisite catalytic systems [16]. In fact, the proposed intelligent multi-objective optimizer is able to frequently recall well-known algebraic equations tracking molecular weight distribution and chemical composition distribution changes, all in an online mode to transform predefined microstructural profiles back to optimal copolymerization recipes. Also, commercial or in-house developed simulators/modelers can be directly amalgamated with the intelligent optimizer considering proper (1) image possessing/pattern reorganization, and (2) communication units to reformat and translate the outcomes of the simulators/modelers into understandable information for the intelligent optimizer and vice versa. For instance, Mohammadi et al. have developed a hybrid reciprocating technique, referred to as Optimulation algorithm, capable of simultaneously simulating/optimizing complicated chemical, biological, and macromolecular reaction engineering problems [17]. This makes use of an online communication of an in-house developed molecular simulator (a KMC-based chemical reaction simulator) with a heuristic multi-objective optimizer. Although the proposed computational tool has initially been successfully implemented to 'optimulate' the oxidative coupling of methane (OCM) as a complex chemical reaction case study, it can be effectively utilized to handle a wide range of multi-objective optimization problems for other complex reacting systems.

In the online mode, not only there exist more flexibility and opportunities but also potential computational errors are considerably decreased as constructing and training intelligent modeler(s) inevitably lead to a built-in error. The offline mode, however, has the benefit of being computationally more cost-effective.

In general, data post-processing methods are divided into two main categories, including data visualization and data summarization. After knowledge extraction, it is necessary to visualize the discovered knowledge in such form so that the end user can gain perfect insight into processed data for better interpretation and decision making. Hence, pattern evaluation, pattern selection and interpretation, and pattern visualization are the most important responsibilities of a post-processing unit. In fact, not only does it provide a visual overview of derived computer representations but also simplifies searching for patterns, trends, potential irregularities, and relationships among processed data.

Even though there has been ample evidence with respect to applying data mining approaches in polymer science and technology [18-20], effective and comprehensive development/implementation of advanced data 'refinery'/processing techniques in macromolecular reaction engineering are still worth pursuing and/or improving. Recently, we have successfully developed, implemented and applied data mining techniques employing Artificial Intelligence-based modelers and multi-objective optimizers to appropriately handle several typical complex polymerization systems [14-17]. To achieve this, we have generated theoretical datasets applying available algebraic/differential equations and also in-house simulators. The generated datasets on polymerization kinetics and polymer microstructure have subsequently been effectively refined and processed applying data mining techniques, and intelligent modelers and optimizers followed suit. It has been clearly shown that the developed tools are powerful enough to (1) precisely monitor, control, and optimize complex polymerization reactors, (2) tailor-make the microstructure and architecture of produced macromolecules, and (3) simultaneously satisfy all predefined final properties.

The near future

All in all, polymerization data mining is a necessity in modern macromolecular reaction engineering to comprehensively analyze generated complex 'big data' and effectively 'crack' the recipe-microstructure-property interrelationship. Definitely, the development and implementation of computationally cost-effective (1) virtual polymer synthesizers, (2) intelligent modelers/optimizers, and also (3) virtual simulators for polymer characterization/processing are of paramount importance to guarantee the success of polymerization data mining projects. Although advanced virtual synthesizers and versatile intelligent modelers and optimizer have recently been developed and successfully put into practice [14-17] applying molecular simulation approaches and Artificial Intelligence techniques, the design and establishment of powerful simulators for characterization and processing of virtually synthesized macromolecules are open to future developments, being of paramount importance to both industry and academia.

References

1. J.-F. Lutz, J. M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mater.* **2016**, *1*, 1.
2. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1.
3. K. Matyjaszewski, J. Xia, *Chem. Rev.* **2001**, *101*, 2921.
4. K. Matyjaszewski, N. V. Tsarevsky, *J. Am. Chem. Soc.* **2014**, *136*, 6513.
5. A. Qin, J. W. Y. Lam, B. Z. Tang, *Macromolecules* **2010**, *43*, 8693.
6. D. J. Arriola, E. M. Carnahan, P. D. Hustad, R. L. Kuhlman, T. T. Wenzel, *Science* **2006**, *312*, 714.
7. S. García, J. Luengo, F. Herrera, *Data Preprocessing for Data Mining*, Springer International Publishing, Switzerland, **2015**.
8. J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, USA, **2012**.
9. D. T. Larose, C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Inc., USA, **2014**.
10. I. H. Witten, E. Frank, M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, USA, **2011**.
11. C. C. Aggarwal, *Data Classification: Algorithms and Applications*, CRC Press, USA, **2015**.
12. C. C. Aggarwal, C. K. Reddy, *Data Clustering: Algorithms and Applications*, CRC Press, USA, **2014**.
13. D. Greiner, B. Galván, J. Periaux, N. Gauger, K. C. Giannakoglou, G. Winter, *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*, Springer International Publishing, Switzerland **2015**.
14. M. R. Saeb, Y. Mohammadi, T. S. Kermaniyan, P. Zinck, F. J. Stadler, *Polymer* **2017**, *116*, 55.

15. Y. Mohammadi, M. R. Saeb, A. Penlidis, E. Jabbari, P. Zinck, F. J. Stadler, K. Matyjaszewski, *Macromol. Theory Simul.* **2018**, *27*, 1700106.

16. Y. Mohammadi, M. R. Saeb, A. Penlidis, *Macromol. Theory Simul.* **2018**, *27*, 1700088.

17. Y. Mohammadi, A. Penlidis, *Ind. Eng. Chem. Res.* **2018**, *57*, 8664.

18. N. Adams, U. S. Schubert, *J. Comb. Chem.* **2004**, *6*, 12.

19. O. AbuOmar, S. Nouranian, R. King, J. L. Bouvard, H. Toghiani, T. E. Lacy, C. U. Pittman Jr., *Adv. Eng. Inform.* **2013**, *27*, 615.

20. X. Zhao, D. W. Rosen, *J. Manuf. Syst.* **2017**, *43*, 271.

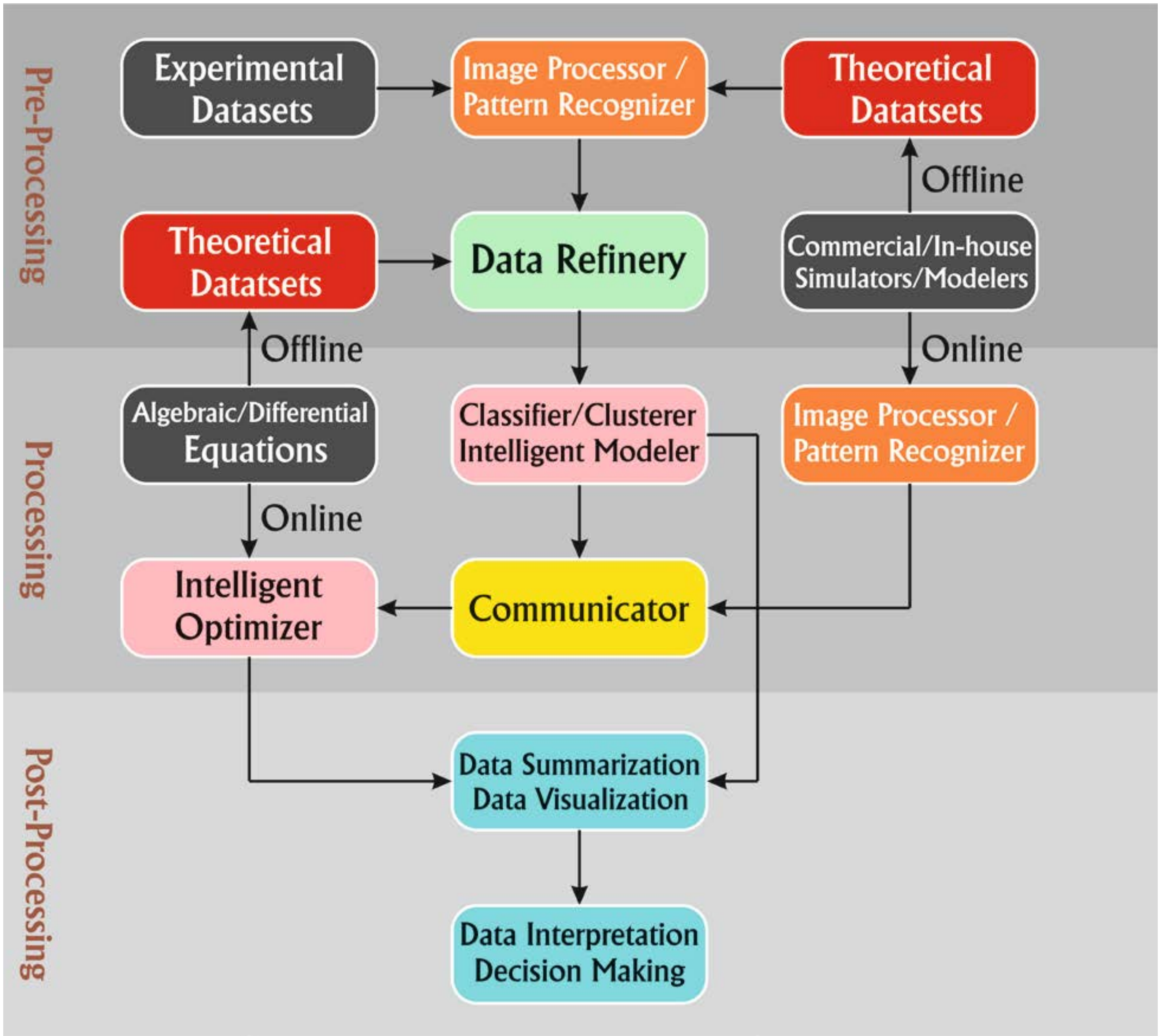
Pre-Processing
Processing
Post-Processing
Experimental Datasets
Image Processor / Pattern Recognizer
Theoretical Datatsets
Offline
Theoretical Datatsets
Data Refinery
Commercial/In-house Simulators/Modelers
Offline
Online
Algebraic/Differential Equations
Classifier/Clusterer Intelligent Modeler
Image Processor / Pattern Recognizer
Online
Intelligent Optimizer
Communicator
Data Summarization Data Visualization
Data Interpretation Decision Making

Figure 1. Polymerization Data Mining.