International Journal of Transportation Science and Technology xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Development of a global road safety performance function using deep neural networks

Guangyuan Pan [a], Liping Fu [a,b,*], Lalita Thakali [a]

[a] Department of Civil & Environmental Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[b] Intelligent Transportation Systems Research Centre, Wuhan University of Technology, Mailbox 125, No. 1040 Heping Road, Wuhan, Hubei 430063, China

ARTICLE INFO

Article history:
Received 28 February 2017
Accepted 20 July 2017
Available online xxxx

Keywords:
Road safety
Collision model
Global model
Deep learning

ABSTRACT

This paper explores the idea of applying a machine learning approach to develop a global road safety performance function (SFP) that can be used to predict the expected crash frequencies of different highways from different regions. A deep belief network (DBN) – one of the most popular deep learning models is introduced as an alternative to the traditional regression models for crash modelling. An extensive empirical study is conducted using three real world crash data sets covering six classes of highways as defined by location (urban vs. rural), number of lanes, access control, and region. The study involves a number of experiments aiming at addressing several critical questions pertaining to the relative performance of the DBN in terms of network structure, training method, data size, and generalization ability, as compared to the traditional regression models. The experimental results have shown that a DBN model could be trained with different crash datasets with prediction performance being at least comparable to that of the locally calibrated negative binomial (NB) model.

Introduction

Road safety studies, such as network screening of road network and evaluating safety effects of countermeasures, rely on collision prediction models or safety performance functions (SPF) to estimate the expected safety levels of specific road entities such as intersection and sections under specific conditions. Safety performance functions are commonly developed separately for different types of highways or entities and locally using data collected from the study area representing the specific highway types to be modelled. For example, SPFs are developed separately for freeways, two-lane rural highways, multilane rural highways and other highway types, depending on the context of the analysis. Such practices are well reflected in the Highway Safety Manual (HSM) that documents a number of example SPFs for various types of highways and intersections from different jurisdictions (https://www.fhwa.dot.gov/research/publications/technical, 2017; Highway Safety Manual, 2010).

This traditional approach of developing individual local models, e.g., based on highway types or geographical regions, is often time consuming as we need to go through the modelling approach for each case separately (Aguero-Valverde and Jovanis, 2008). Moreover, the commonly used parametric form of models (e.g., Negative Binomial (NB) model) require a ser-

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author at: Department of Civil & Environmental Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.
E-mail address: lfu@uwaterloo.ca (L. Fu).

http://dx.doi.org/10.1016/j.ijtst.2017.07.004
2046-0430/© 2017 Tongji University and Tongji University Press. Publishing Services by Elsevier B.V.

ies of trial and error process before arriving at the final model structure with a set of significant variables (Thakali et al., 2016; Shankar et al., 1995), while this model specification step is not required in the machine learning technique (e.g., deep belief network). In this research, we explore the feasibility of developing a modelling framework where datasets coming from multiple regions could be pooled together into a same modelling step and used to generate a single global model. Ideally, this modelling approach is superior in prediction performance as compared to the local models, is adaptive with ability to incorporate new data, and can be automated in model training and updating. However, the challenge lies in making sure that there is no loss of site-specific information during the generalization of the global model (Connors et al., 2013).

In this paper, we propose to apply the deep learning technique for developing a global SPF. Deep learning is one of the most recent and promising techniques developed in machine learning (Yann et al., 2015). It overcomes the shortcomings of traditional artificial neural network (ANN) and has been implemented successfully in many areas, such as pattern recognition, computer vision, and intelligent decision-support (Dahl et al., 2011; Yichuan, 2013; Taylor et al., 2006). An extensive empirical study is conducted to investigate the model performance in terms of prediction accuracy as compared to the traditional NB modelling approach which is the most widely used technique in road safety studies (Cheng et al., 2013; Usman et al., 2012). The paper hereafter is arranged as follows: Section 2 reviews some of the popular machine learning methods under the domain of artificial neural networks, including the deep neural networks proposed in this paper. Section 3 describes the deep belief network (DBN) model proposed for modelling road collision along with its training algorithm. Section "Case Studies" discusses the important parameters of the global DBN, and also presents a few case studies to demonstrate the idea of a global model. Finally, Section 5 concludes with the main findings.

Literature review

Machine learning (ML) has become a very popular topic of study in the field of computer science in recent years. This includes techniques such as artificial neural network (ANN) which is aimed to reproduce and simulate human behaviour and cognitive functions. ANN uses a network of nodes (often called "neurons") containing configurable weights that can be trained to produce a desired output (Miaou and Lord, 2003; Abdel-Aty and Haleem, 2011). These weights and layers can be configured to solve many kinds of pattern recognition problems, some ML models even have found their way into other fields, including road safety studies where they have been used for predicting collisions. For example, a recent study by Chang (2005) implemented an ANN to predict collisions in National Freeway in Taiwan. The model in their study accepts road condition features as input and provides collision numbers as output.

While ANN models are easy to understand, its weight solution space is non-convex, which means there are many local minimums, making it difficult to find the global optimum solution. Another drawback of traditional ANN is its supervised learning construct that necessitate availability of training data, limiting its applicability in real-world situations (Rumelhart et al., 1986; Bengio and Olivier, 2011). To address these problems, other improved versions of BP such as Bayesian regularization have been proposed (Xie et al., 2007). While Bayesian regularization shows improvements, it still requires training data that limits its applicability.

Deep learning (DL), or deep neural network (DNN), is a new machine learning technique that has been widely explored and successfully applied for a variety of problems such as in image and voice recognition and games (Hinton and Salakhutdinov, 2006; Hinton et al., 1207; Ren et al., 2014). A few variations of DNN have been developed, among which Deep Belief Network (DBN) is one of the most popular (Yann et al., 2015). What makes DBN different is its unique training method called greedy unsupervised training. Its development is based on mimicking the cognitive and knowledge reference processes of the human brain. By stacking several Restricted Boltzmann Machines (RBM, a kind of recursive ANN model that contains two layers, one input layer and one output layer), one upon another, DBN learns the features from inputs and obtains a better distributed representation of the input data, without requiring training data like back propagation. One of the most impressive characteristics of deep neural network is its ability to learn representation from multi-category data (Papandreou et al., 2015). For example, in Srivastava and Ruslan (2012), a Deep Belief Network architecture was shown to be able to learn the distribution over the space of multimodal inputs. The model is capable of creating a multimodal representation even when some image data modalities are missing.

Global deep belief network model

Model architecture

An illustration of global DBN model is shown in Fig. 1. It consists of three types of layers. The first is the visible layer (V) which receives the original feature data and acts as the input layer. This input layer is followed by several hidden layers, L1, L2 and more if necessary, with V and L1 forming the first RBM, and L1 and L2 the second RBM, and so on. The structure of each RBM is that of a two-way full connectivity between V and L as shown by double-sided arrows in Fig. 1. No connections exist between units of the same layer. In the training process, each hidden layer extracts the last layer's data information and features to form a better, although more abstract, distributed representation of input data. The last layer is the output layer which has only one unit, representing the predicted collision frequency. In our case study, we only illustrate a global DBN with two hidden layers, and more details will be discussed in Section "Case Studies".

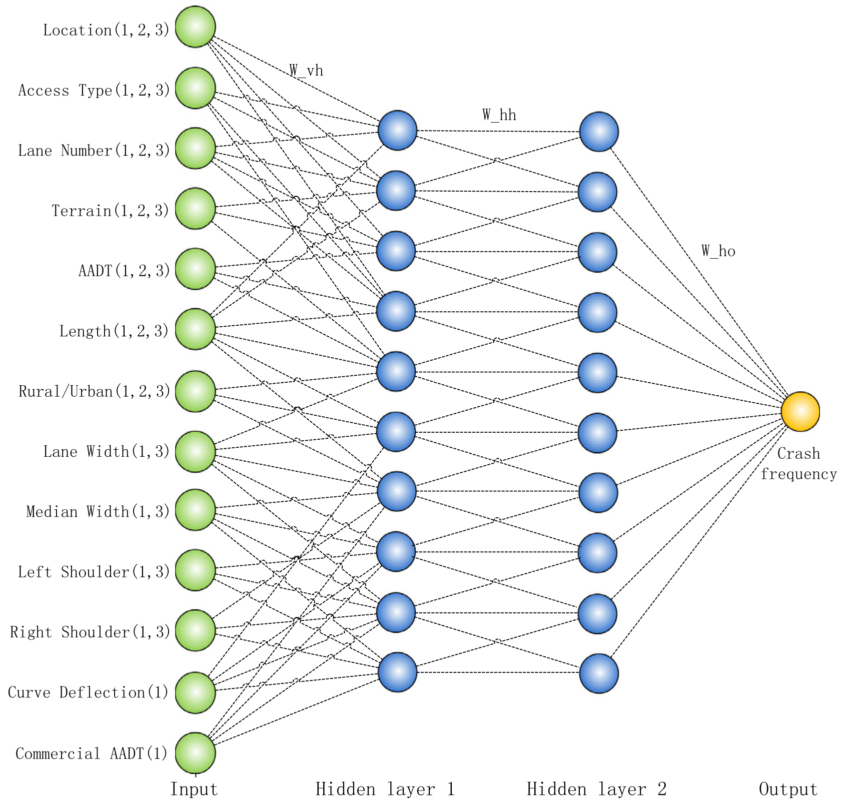


Fig. 1. Global DBN's architecture.

Traditional Deep Belief Networks (DBN) can only process binary signals (0 and 1). However, real world road safety related data are continuous values, which can be easily normalized to lie between 0 and 1. Hinton (Chen and Murray, 2003) improved the visible layer to receive continuous values, while the signal in the unsupervised training process will still be transformed to a binary value, and this limitation can be addressed by using an improved continuous version of the transfer function which is proposed by Bengio (Bengio, 2009). Based on his work, a continuous RBM can be used in continuous value prediction problem. For example, a continuous DBN is used in predicting a time series benchmark and good results are obtained (Qiao et al., 2015). Therefore, in our research, the continuous transfer function is employed. In the layer-wise greedy unsupervised training, once the first RBM is finished training, the output of it becomes the input of the second RBM. Furthermore, for a global DBN, the previous dataset may be used multiple times in training the model whenever a new dataset is added to update the model, thus the model may easily become overfitting. Therefore, in the proposed model, Bayesian regularization, instead of traditional back propagation, is implemented to raise prediction accuracy and, in the meantime, to reduce the overfitting problem (Pan et al., 2014).

Model algorithm

Global DBN is made of several RBMs, in which the first one is continuous RBM. They are trained separately before unfolding all the layers and using back propagation to fine-tune the weights. For the training of the first continuous RBM, the idea of knowledge learning and reasoning processes are shown in Eqs. (1) and (2).

$$p(h_j = 1) = \frac{1}{1 + e^{-b_j \sum_i v_i w_{ij}}} \tag{1}$$

$$s_i = \varphi_i\left(\sum_j w_{ij} s_j + \sigma \cdot N_i(0,1)\right)) \tag{2}$$

where

$$\varphi_i(x_i) = \theta_L + (\theta_H - \theta_L) \cdot \frac{1}{1 + e^{-a_i x_i}} \tag{3}$$

and v_i is the binary value of input neuron i of the visible (input) layer while s_i is the continuous value of it, h_j is the value of unit j in the hidden (output of the RBM) layer; b and c stand for the biases of the visible and hidden layers, respectively; w_{ij} is the weight between visible unit i and hidden unit j, $\mathrm{N}(0,1)$ is a Gaussian random variable with mean 0 and variance 1. σ is a constant, $\varphi(.)$ in (3) denotes the sigmoid function in equation (2) with asymptote of θ_H and θ_L. a is a variable that controls noise, which means it controls the gradient of the transfer function.

For the rest RBMs, the two processes are shown in Eqs. (4) and (5).

$$p(h_j = 1) = \frac{1}{1 + \mathrm{e}^{-b_j - \sum_i v_i w_{ij}}} \tag{4}$$

$$p(v_i = 1) = \frac{1}{1 + \mathrm{e}^{-c_i - \sum_j h_j w_{ji}}} \tag{5}$$

We then define $\theta = (W, b, c)$ according to a fast training method that is commonly used to estimate successive new weights and biases with a Markov chain, as follows (Hinton, 2002).

$$\theta^{(\tau+1)} = < h_j^0 v_i^0 > - < h_j^1 v_i^1 > \tag{6}$$

where $<>$ denotes the average over the sampled states, $h_j^0 v_i^0$ is the initial state of the visible layer multiplied by the hidden layer, and $h_j^1 v_i^1$ is the same product after a single iteration of a Markov chain.

Just like the general DBN model, a global DBN must also be fine-tuned in terms of model structure and learning rate. However, training a network based on limited samples is a loosely framed problem. That is to say, there are many potential models that could be used to train the model such that its output is very close to the expected output. In the proposed global DBN, this problem could be more severe, especially when a new dataset is added to an already well-trained model to retrain and update the weights. The retrained model can be easily over fitted, as the previously trained dataset will be reused. In order to avoid this problem, the model needs to be regularized. That is, additional conditions apart from the requirement that the response of the trained network must agree with the expected one need to be imposed. Therefore, the proposed global DBN is trained using Bayesian regularization supervised training algorithm.

Bayesian regularization is a technique commonly employed to achieve this objective. In Bayesian regularization, the additional term imposed ensures that the selected trained network not only minimizes a metric of the error but also achieves this with weights that are of as small a magnitude as possible. In this research, we propose Bayesian regularization instead of back propagation for the fine-tuning process after unsupervised training. The objective function employed is as follows,

$$F_W = \alpha P + \beta E_W \tag{7}$$

where, F_W is the new objective function in the process of supervised training, P is the original object function as per Eq. (8). E_W is the Bayesian regularization item, and α and β are performance parameters that need to be calculated in the iterations or be set before iteration. E_W has the form of mean square of the weights.

$$P = \frac{1}{T}\sum_{t=1}^{T}(O_t - O_r)^2 \tag{8}$$

$$\mathrm{E_W} = \frac{1}{\mathrm{m} \cdot \mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{w}_{\mathrm{ij}}^2 \tag{9}$$

where, T is the testing set, O_t means output at testing set t, O_r is the ideal output or teacher's signal. $\mathrm{w_{ij}}$ means the weight between layer i and j. If $\alpha \gg \beta$, then the first part of F_W dominates, which means that the objective of the training is to decrease the training error. Specifically, if $\alpha = 1, \beta = 0$, then $F_W = P$, and the Bayesian regularization becomes ordinary back propagation. On the other hand, if $\alpha \ll \beta$, the training will focus on decreasing the weights. Therefore, by introducing this regularization term, one can expect that weights that do not contribute to the response will be minimized ensuring that only parts of the network that have learned important features common to all the input patterns will remain. Therefore, an improvement in the response of the trained network to unknown test inputs is expected.

Bayesian regularization models are then trained by calculating values of α and β during the training process. During this process, the weights are treated as random variables, and assumes that the prior probabilities of P and E_W are Gaussian. α and β can then be obtained by using Bayes criterion.

For our study, we use Matlab as our computing tool with the codes being obtained from www.deeplearning.net (www.deeplearning.net, 2017) and modified for our modelling and experiments.

Case studies

To evaluate the feasibility of developing a single global safety performance function using DBN that can be used to predict the expected collision frequency on highways of different types from different regions, an empirical study is conducted using

Please cite this article in press as: Pan, G., et al. Development of a global road safety performance function using deep neural networks. International Journal of Transportation Science and Technology (2017), http://dx.doi.org/10.1016/j.ijtst.2017.07.004

collision data from three different regions, including, including Highway 401 of Ontario, highways of Colorado and Washington states in US, as shown in Fig. 1. A detailed description of the case study data is provided in the following section. We also developed negative binomial (NB) models, for benchmarking the performance of traditional approach of developing individual (local) model with the global DBN as well as with individual DBNs. It is noted that the NB model is the most widely used technique in road safety analysis (Lord and Mannering, 2010) which has also been adopted by HSM manual (Highway Safety Manual, 2010). More details on NB modelling can be found in Washington and Karlaftis (P. Washington et al., 2010). An attempt is made to include as many variables as possible in the model, e.g., traffic level, geometric features of roadways and other road environment related factors, depending on the data availability for each case study.

Case descriptions

Highway 401 in Ontario, Canada

This case is based on historical collisions and related data from Highway 401, a multilane access controlled highway in Ontario, Canada (hereafter, it is also referred as ON-M). This highway is one of the busiest highways in North America and connects Quebec in the east and the Windsor-Detroit international border in the west. The total length of the highway is 817.9 km of which approximately 800 km was selected for this study. According to 2008s traffic volume data, the annual average daily traffic (AADT) ranges from 14,500 to 442,900 indicating comparatively a very busy road corridor. The databases used in this case study include: 1) historical crash records for the period from 2000 to 2008 extracted from MTO's Accident Information System (AIS); 2) historical AADT data for the same years from MTO's Traffic Volume Inventory System (TVIS); and 3) road geometric features from MTO's Highway Inventory Management System (HIMS) database. Note that each record in this database is referenced to MTO's linear highway reference system (LHRS). LHRS is a one-dimensional spatial referencing system with a unique five-digit number representing a node/link on a particular highway. LHRS can be used to locate the position of features on a map using a Geographical Information System (GIS) tool.

Our first step in data processing was to generate a set of homogenous sections (HS) in which each HS section represents segments with similar characteristics such as number of lanes, shoulder width, the presence of median, curvatures, and other roadway features. As previously mentioned, all the data (crash, road geometry and traffic data), are spatially referenced to MTO's LHRS system. All features were geocoded in the GIS map through a multi-step procedure. First, the geometric features from the HIMS database, which was in a spreadsheet format, were geocoded in the GIS platform. There were a total of 244 records in the HIMS layer, each representing a road section with a set of uniform road geometry features. However, as the road curvature was missing in the database, further geoprocessing was needed to obtain the final set of HS sections. For this, curve sections were first demarcated on a map using a GIS tool, thus generating a curve layer. This tool automatically created an attribute table for the curve layer with detailed information such as LHRS number, start point, length and radius related to each curve. For a refined set of HS sections, the initial HIMS layer was split at the intersection of the curve layer, and the segmented HIMS layer was spatially joined to the curve layer in order to transfer all the curvature related information. As each road section's initial HIMS may have one or more curvature sections, these were disaggregated into smaller subsections thereby including an additional level of homogeneity. Note that the shortest length of HS section was 0.2 km. This selection of a certain lower threshold value complies with literature as it had been suggested that very short road segments might have higher uncertainty and lower exposure problems (Council and Stewart, 1999; Begum et al., 2009; Ahmed et al., 2011). Finally, these HSs, assigned with unique IDs, were used as the spatial unit for integrating crash and traffic volume data. There are a total of 418 unique HS sections with length ranging from 0.2 km to 12.7 km and covering 800 km, or 97.9%, of Highway 401.

After generating HS sections, historical crash data from 2000–2008 were geocoded. Similarly, traffic count data, consisting of AADT and average annual commercial vehicle counts for the period 2000–2008, were also geocoded. As each observation recorded LHRS and offset information, the traffic counts were spatially located using the linear referencing GIS tool. Each HS was then assigned the nearest traffic observation. Note that a total of 170 traffic counting stations were available for the 418 HSs. Approximately 85% of the HSs have traffic values assigned from its nearest count station within 2 km distance indicating that the traffic data was quite an extensive in this particular study area. To identify whether the HS segments located in rural or urban environment, we used google earth.

Finally, the processed crash and traffic data were integrated into a single dataset with HS and year as the mapping fields that resulted into total 3762 records. The input features included in this dataset are location, access type, lane number, terrain, AADT, segment length, city type (rural or urban), lane width, median width, left shield, right shield, curve deflection, commercial AADT. These set of variables relevant to Ontario's case study is indicated by "1" in Fig. 1. We also provide a summary description of continuous input features in Table 1, including the sample sizes for training and testing.

Two-lane two-way highways in Colorado, US

This dataset contains crash data from rural two-lane highways in the Colorado State, US (hereafter, it is also referred as CO-R). It represents a case of two-lane rural roads with observations from 1991 to 1998 (downloaded from http://extras.springer.com; (Hauer, 2015). Its AADT is ranging from 40 to 21,720, with an average approximately 2200, which is significantly lower than the first case described previously. The dataset covers a total of 4593 unique sections with length ranging from 0.21 to 31.76 km, which is much large in sample size of the Ontario case. The dataset is divided into two subsets: 1991–1996 set (27,558 observations) for training, and 1997–1998 (9186 observations) for testing. The input features included in

Table 1
Summary of the dataset (Highway 401, Ontario).

Variables	Mean	Max	Min	St. dev.	Sample size
Collisions (per year)	23.81	468	0	50.02	3762 Training:2926 Testing: 836
AADT (veh /day)	76633	442900	12000	91476	
Segment Length (km)	1.95	12.7	0.2	2.06	
No of lanes	5.44	12	4	2.42	
AADT- Commercial (veh/day)	13993	42076	0	6719	
Median width (m)	11.11	30.5	0.6	6.14	
Shoulder width-right (m)	3.14	4	2.6	0.28	
Curve deflection (per km)	0.19	1.86	0	0.35	
Shoulder width-left (m)	1.6	5.19	0	1.19	

this dataset are location, access type, lane number, terrain, AADT, segment length, city type (rural or urban). These variables relevant to Colorado's case study are indicated by "2" in Fig. 1. We also provide a summary description of continuous input features in Table 2, including the sample sizes for training and testing.

Highways in Washington State, US

This dataset was obtained from the Federal Highway Administration's (FHWA) Highway Safety Information System (HSIS), containing inventory and crash data for four types highways – rural two-lane two-way, rural multi-lanes, urban highways with two lanes, and urban highways with multi lanes, in Washington State, US. HSIS crash data has detailed information about the time, location and types of crashes. Each crash record is referenced to a geographical location by two fields, namely, route number and milepost. Our study focused on highway segments; therefore, we removed all the crashes that occurred at intersections, interchanges, ramps, driveways including those that are not related to the study scope. HSIS highway inventory data has detailed information about road geometry features (e.g., the number of lanes, shoulder width, median width and others) and traffic counts on a yearly basis. Each record in the file represented a homogenous (uniform) segment and is referenced to a geographical location by three fields – route number, beginning milepost and ending milepost. These locations related fields are important for merging inventory data with the crash data. Note that a separate inventory file is obtained for each year; therefore, any modifications in road geometric features over the years could have been included. Then, we excluded all the records with segment length less than 0.16 km and lane width less than 7.32 m, similar to the study by Vogt and Bared (1998). We also removed records that have missing traffic data. Finally, crashes were aggregated on an annual basis over individual homogenous segments in the inventory file using the common location related fields mentioned previously. Four types of highways are included in this dataset, rural highways with two lanes, rural highways with multi lanes, urban highways with two lanes, and urban highways with multi lanes, hereafter, they are also referred as WA-R2, WA-RM, WA-U2 and WA-UM, respectively. The input features included in this dataset are location, access type, lane number, terrain, AADT, segment length, city type (rural or urban), lane width, median width, left shoulder width, right shoulder width. These set of variables relevant to Washington's case study is indicated by "3" in Fig. 1. We also provide a summary description of continuous input features in Table 3, including the sample sizes for training and testing

Experimental design

Experiment 1: separate training

We first analyse the effect of the model performance with two case studies: Hwy401, Ontario and Colorado, US. The following bootstrapping process is followed:

1) Split the given dataset into two subsets: a training set and a testing set. The training set includes the first several years of data while the testing set includes the remaining data. In this experiment, we conduct HWY401 and Colorado dataset to evaluate the performance. For HWY401, the training set used year 2000–2006, and the testing set is 2007–2008. While in Colorado data, the training set designed is data 1991–1996 and testing is 1997–1998.
2) The whole data size is used to calibrate or train the candidate model, which is subsequently used to predict the collisions at the testing data set. The MAE and RMSE are then calculated as per Eqs. (10) and (11).
3) Repeat Step 2) for several times to reduce the error of affection of training set (100 times for HWY401 and 25 times for Colorado). After done, calculate the average, minimum and maximum MAE and RMSE of the repetitions.

Table 2
Summary Description of Datasets for Colorado Highway.

Variables	Mean	Max	Min	St. dev.	Sample size
Collisions (per year)	0.9	54.0	0.0	2.2	36744
AADT (veh/day)	2217	21720	40	2534	Training: 27558 Testing: 9186
Length (km)	2.1	31.8	0.2	2.4	

Please cite this article in press as: Pan, G., et al. Development of a global road safety performance function using deep neural networks. International Journal of Transportation Science and Technology (2017), http://dx.doi.org/10.1016/j.ijtst.2017.07.004

Table 3
Summary description of datasets for washington highways.

Variable	Mean	Max	Min	St.dev	Sample size
Rural two-lane highways (WA-R2)					
Collisions (per year)	0.47	15.00	0	0.96	45005 Training: 28436 Testing: 16569
AADT (veh/day)	3424.00	27540	121	3431.54	
No of lanes	2	2	2	0.00	
Lane width (m)	3.53	7.32	2.44	0.33	
Segment length (km)	0.93	13.02	0.16	1.07	
Median width (m)	0.02	57.91	0.00	0.69	
Shoulder width- left (m)	1.43	11.28	0.00	0.81	
Shoulder width- right (m)	1.43	12.19	0.00	0.81	
Rural multi-lane highways (WA-RM)					
Crashes (per year)	1.28	54.00	0	2.35	6837 Training: 3381 Testing: 3456
AADT (veh/day)	21337.14	121311	3453	16628.81	
No of lanes	4	8	3	0.82	
Lane width (m)	3.68	7.32	3.35	0.20	
Segment length (km)	0.59	9.55	0.16	0.71	
Median width (m)	28.84	304.50	0.00	40.18	
Shoulder width- left (m)	1.20	7.92	0.00	0.64	
Shoulder width- right (m)	2.90	12.19	0.00	0.68	
Urban two-lane highways (WA-U2)					
Crashes (per year)	0.63	13.00	0	1.13	4704 Training: 3566 Testing: 1138
AADT (veh/day)	11146.32	66603	428	6292.23	
No of lanes	2	2	2	0.00	
Lane width (m)	3.72	7.32	3.05	0.69	
Segment length (km)	0.39	3.98	0.16	0.34	
Median width (m)	0.20	152.40	0.00	4.12	
Shoulder width- left (m)	1.62	9.14	0.00	0.89	
Shoulder width- right (m)	1.72	6.71	0.00	0.88	
Urban multi-lane highways (WA-UM)					
Crashes (per year)	1.28	54.00	0	2.35	7910 Training: 5377 Testing: 2533
AADT (veh/day)	21337.14	121311	3453	16628.81	
No of lanes	4	8	3	0.82	
Lane width (m)	3.68	7.32	3.35	0.20	
Segment length (km)	0.59	9.55	0.16	0.71	
Median width (m)	28.84	304.50	0.00	40.18	
Shoulder width- left (m)	1.20	7.92	0.00	0.64	
Shoulder width- right (m)	2.90	12.19	0.00	0.68	

Model performances mentioned in Step 2 are estimated on the basis of mean absolute error (MAE) and root mean square error (RMSE) in these case studies, as defined in Eqs. (10) and (11). These measures quantify the average deviation of the estimated crash frequencies from the observed values. Therefore, smaller the magnitude of these measures better is their performance level.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{11}$$

where, P_i is the predicted collision frequency for ith observation, O_i is the ith observed collision frequency, and n is the total number of observations.

Note that the difference between these two performance measures, MAE and RMSE, is how the residuals are weighted. In MAE, equal weights are given to the residuals from the observed points, whereas in RMSE larger residuals are given greater weights by squaring the deviation. Hence, the smaller the magnitude is, the better the performance level we achieve.

Experiment 2: simultaneous training

In this experiment, we analyze the performance of the DBN trained using all of the three data sets described previously in Section "Case Descriptions". As illustrated in Fig. 2, these datasets are collected from different kinds of highways and with different features. Therefore, we need to process some of the input features that indicate the three kinds before training. The training process is as follows.

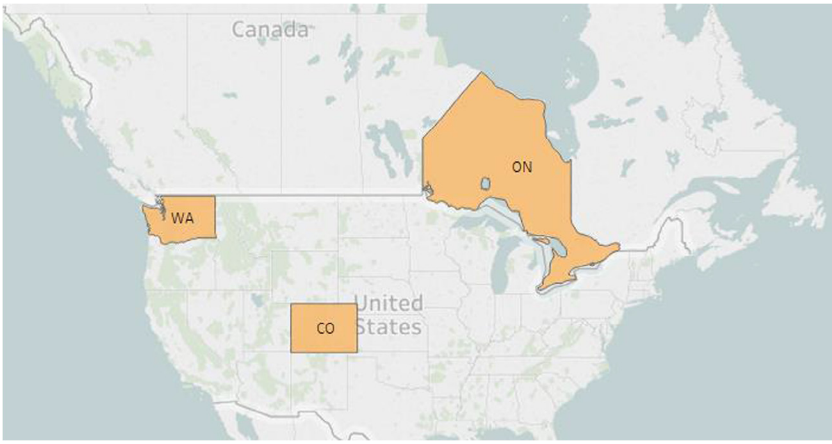


Fig. 2. Location map of case studies in three different regions: Ontario (ON) province, Colorado (CO) state and Washington (WA) state.

1) Three input neurons are used to indicate the different datasets. The first one is to indicate locations. We use a constant number for each study location: 1 for Hwy 401, 2 for Colorado highways and 3 for Washington highways. However, the dataset of Washington contains four road types, which are rural or urban, two-lane or multi-lane. Therefore, we design the second and the third input neurons to indicate road type and access type. For road type, a constant input is used to distinguish all rural or all urban roads, and similar method is conducted to distinguish limited or controlled access type, and for where the feature information is missing, a constant 0 will be used instead.
2) Normalize the full dataset to [0,1], so that the global DBN can uniformly regard the inputs of the three cases we designed from step 1.
3) The full size training data is then used to train the candidate model, which is subsequently used to predict the collisions at the testing data set separately. The MAE and RMSE are then calculated.
4) Repeat Step 3) for 10 times to reduce the affection of initialization. After done, calculate the average, minimum and maximum MAE and RMSE of the repetitions.
5) Change model size and repeat step 3) to 4) again.

Experiment 3: sequential training

This experiment is designed to testify the generalization of the global DBN. As in course of time, road conditions could change and more collision data could become available. This experiment is to show whether or not a DBN model, once trained, can be further trained with a completely new data set without significant loss in prediction performance. For this experiment, a DBN is first trained using the first two datasets and then further trained using the dataset of Washington. The experimental process is as follows.

1) Train a global DBN using the datasets of HWY401 and Colorado as in Experiment 2.
2) Combine the three training datasets together and normalize them. Fine tune the model using the updated dataset with only supervised learning approach.
3) Repeat the process for 10 times to reduce the effect of initialization.

Model settings

Parameters that affect the effect of DBN are various, and among those parameters, several are the most important, and sensitivity analysis is necessary in model setting (Hinton, 2010). To experiments in this paper, the key parameters include network structure, including layers and neuron numbers, learning rate, and input features.

The structure pertaining to a DBN model is defined by the number of hidden layers and number of units (or nodes) in each layer, which could have a significant effect on the performance of the model. Too many hidden layers and hidden units can increase the computation and training time as well as lead to overfitting problem, while too few hidden layers and hidden units may lead to poor feature learning and under-fitting situation. The significance of these parameters on the model performance means that these values must be chosen carefully. Previous studies have shown that the most effective method for model size selection is an empirical one. A structure with two layers was chosen due to the fact that traditional DBN employed in a pattern recognition task often only include two layers. This structure strikes a balance between the need for a more powerful model and the desire to keep computation times lower. It offers more predictive power than a single layer model, but is significantly less time-consuming than a three layer model (Pan et al., 2014).

Beside the model structure, selection of an appropriate learning rate is also very important. For a given DBN, learning rates for both unsupervised and supervised component must be specified, as they control the weights of the connections between layers in each iteration. If the learning rate is too large, the reconstruction error (the index to evaluate unsupervised learning) and weights usually increase dramatically. On the other hand, if learning rate is too small, more epochs are needed, which not only increases the learning time but may also limit its chance to find the global optimal solutions.

In addition to the learning rate and model structure, the input features themselves are also important to the performance. Normally, we want to include as many good features as possible, however, some features may contain too much noise and their inclusion may reduce the accuracy of the model. To assess these effects, this paper also explored the relationship between features and performance. In order to test how road features can affect the final modelling result, we compare the results of HWY401 and Colorado in Experiments 1 and 2 by using different training features as have discussed previously.

For Experiment 1, we used a network structure of 13–10–10–1 (13 units for the input layer - 10 units for the hidden layer 1–10 units for the hidden layer and 2–1 units for the output layer) for the model for case HWY401. Previous researches have showed that in certain range, with neurons increased, the accuracy increased, however when the network size became too large, the accuracy fell. And the best network size was also related to input dataset. So for the three experiments, it is necessary to test and try, to determine the most effective size. After deciding network size, the same method was used on choosing learning rate. Basis of our previous work (Qiao et al., 2015), we found that when the training dataset is around 3000 two hidden layers with no more than 10 are the most effective. The same structure is used because the effective input for case Colorado is much fewer. Also, different learning rates varying from 0.1 to 0.5 were tried, with the learning rate of 0.1 being the most effective. In terms of training, we used a total of 20 unsupervised iterations and 1000 fine-tuning iterations to ensure model convergence. In experiment 2, the training dataset is much larger than the first one; as a result, the size of hidden layers should be larger. Model structures which contain hidden neurons from 8 to 14 were tested, of which the structure with more hidden neurons was found to outperform the other structures for case Washington. However, fewer hidden neurons are still better for the first two cases. So the structure of the model was set to be 13–12–12–1 at last. Accordingly, a total of 50 unsupervised iterations and 1000 fine-tuning iterations were adopted. A learning rate of 0.1 is also selected after trying out a range of values from 0.1 to 0.5. For experiment 3, we use the structure of 13–12–12–1 although at this moment we assume there is no dataset of Washington, we still want the model can be more tolerant in case there is a new database. After the new dataset has come, we conduct the supervised training with another 1000 iterations. As the supervised training has been implemented for the pervious data, the training error has already met the setting value, so the new training error should set to be much smaller otherwise the added supervised training will stop soon.

Results

NB models

A NB model is developed for each highway type and geographical location, as this is a common practice of the existing road safety studies. Unlike the DBN, a NB model requires pre-specification of the relation of crash frequency and a set of input features (explanatory variables). We specify this relation using exponential function as shown in Eq. (12).

$$\mu = e^{\beta_0 + \beta X} \tag{12}$$

where, μ is expected crash frequency, X is a vector of features (e.g., exposure, lane width, shoulder width etc.), β is a vector of model coefficients for the corresponding input features and β_0 is intercept.

Note that, for the exposure related features, namely, AADT and segment length, we used their log forms so that their transformations lead to the case of zero crash for zero values. We also tested these features in two forms. First, we considered their combined form by multiplying AADT and length to represent the total exposure as a single feature, as shown in Case Study 1 in Table 4. Second, we considered these features individually as in case study 2 and 3 (Tables 4 and 5). Decision on which exposure form to retain was made based on performance measures, MAE and RMSE, which were evaluated using the testing dataset. In addition, like in any other parametric model, we exclude the features that have a high correlation with any

Table 4
NB model results for Case 1 (Ontario) and Case 2 (Colorado).

Highway 401, Ontario		Rural Two-lane Two-way Highways, Colorado	
Variable	Coefficient estimate	Variable	Coefficient estimate
(Intercept)	−1.070 (<0.000)	(Intercept)	−7.93 (<0.000)
log(AADT*Length*365/10^6)	0.827 (<0.000)	log(AADT)	0.87 (<0.000)
AADT-commercial	0.000 (<0.000)	log(Length)	0.97 (<0.000)
Median width (m)	−0.014 (<0.000)	Terrain-Mountainous	0.96 (<0.000)
Shoulder width-left (m)	-0.104 (<0.000)	Terrain- Rolling	0.50 (<0.000)
Shoulder width-right (m)	0.138 (<0.000)	Terrain- Flat	0.00
Curve deflection (per km)	-0.132 (<0.000)		
Lane width (m)	ns		

Note: ns means not significant; value in the parenthesis represents p-value.

Table 5
NB model results for Case 3 (Washington).

Variable	Coefficient estimates			
	Rural two-lane	Urban two-lane	Rural multilane	Urban multilane
(Intercept)	−6.321 (<0.000)	−5.400 (<0.000)	−6.672 (<0.000)	−12.087 (<0.000)
log(AADT)	0.827 (<0.000)	0.773 (<0.000)	0.776 (<0.000)	1.319 (<0.000)
log(Length)	0.951 (<0.000)	0.933 (<0.000)	0.999 (<0.000)	0.854 (<0.000)
Median width (m)	0.047 (<0.000)	ns	ns	−0.008 (<0.000)
Shoulder width-left (m)	−0.090 (<0.000)	−0.199 (<0.000)	−0.124 (<0.000)	ns
Shoulder width- right (m)	−0.112 (<0.000)	−0.113 (<0.010)	−0.112 (<0.000)	−0.193 (<0.000)
Lane width (m)	−0.185 (<0.000)	−0.210 (<0.000)	−0.129 (<0.000)	ns
Terrain: Mountainous	0.473 (<0.000)	/	1.025 (<0.000)	/
Terrain: Rolling	0.149 (<0.000)	ns	0.276 (<0.000)	0.144 (<0.023)
Terrain: Level	0.00	ns	0.00	0.00

Note: ns means not significant; value in the parenthesis represents p-value; / means no data.

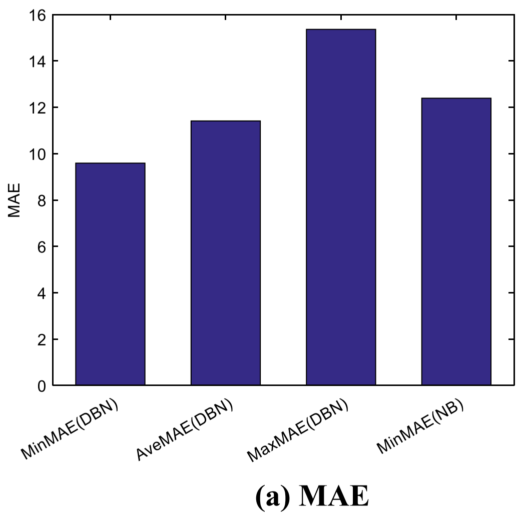
other important features. An example in our study is the existence of a high correlation between AADT and lane number. Therefore, we included only the AADT feature.

The NB model coefficients are estimated using the maximum likelihood method on R statistical platform. A backward selection approach is employed to include only those features that are significant at 5% level of significance. For the first case study, we did include the road type feature (Rural or Urban) in the model; however, this caused other road geometry related features (median width and shoulder width-left) to be insignificant. As previously mentioned, information regarding the road type – being rural or urban was obtained using google map which could be relatively less accurate compared to the information related to road geometric features obtained from the inventory file. We also confirmed that there was minimum difference in their model performance measures. Therefore, for this particular case study, we retained the model based on road geometric features only. A summary of model results for all case studies are presented in Table 4 and 5.

Experiment 1: local DBN model

This experiment focuses on comparing the relative performance of DBN versus NB model when applied for modelling crashes at a specific region. Case 1 (Ontario) and Case 2 (Colorado) were used for this analysis. Fig. 3 shows the validation results for the Ontario case, including the minimum, average, maximum MAE and RMSE of the two models. Table 6 shows the values of testing MAE (minimum, average, and maximum) and RMSE (minimum, average, and maximum) of DBN and NB. The final testing MAE is 9.59 and RMSE is 19.58 using global DBN, which are all better than the other one. The improvements are 22.60% and 32.34% respectively, comparing to NB.

For the Colorado case, three features (AADT, length, terrain) are used for both models (NB and DBN). The training and testing results are shown from Fig. 4. The results of using different models are summarized in Table 7. After trying different structures, we conclude that a DBN with 10 hidden neurons has the best prediction power for this case, having the lowest MAE of 0.81, comparing to 0.83 by NB. Similarly, the best RMSE is 1.48, comparing to 1.67 by NB. Although more features may mean more uncertainty and sometimes unwanted features, and larger model sizes need more training time and data and may learn less important features, the result has turned out the be very similar to the one using only three feature in



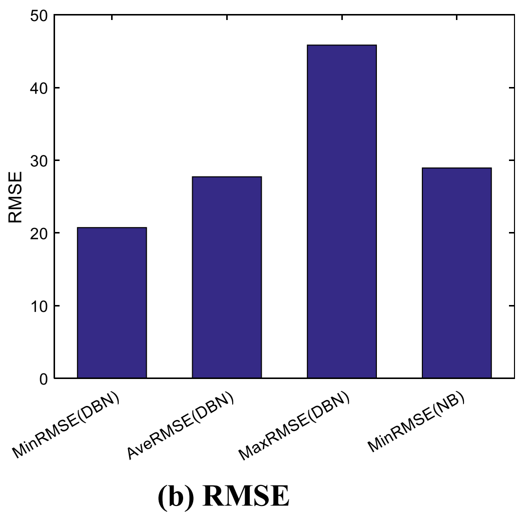


Fig. 3. Performance of DBN and NB (The Ontario HWY401 case).

Table 6
Comparison of models in training HWY401.

Models	NB (Base)	Bayesian ANN	DBN	
			Error	%Improvement
MAE	12.39	11.61	9.59	22.60
RMSE	28.94	26.81	19.58	32.34

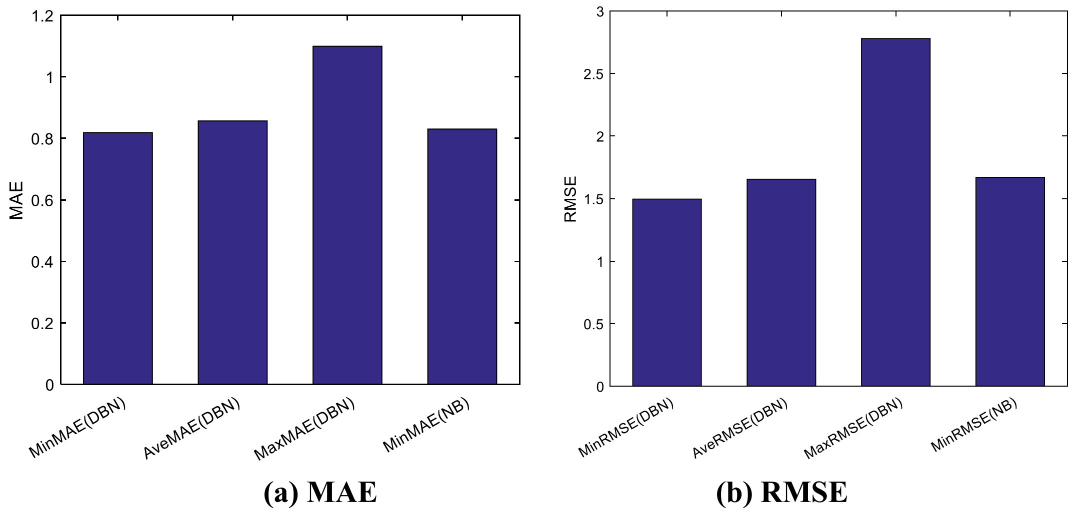


Fig. 4. Performance of global DBN and NB in Colorado highways.

Table 7
Comparison of models in training Colorado highways.

Models	NB (Base)	Bayesian ANN	Global DBN	
			Error	%Improvement
MAE	0.83	0.83	0.81	2.40
RMSE	1.67	1.60	1.48	11.38

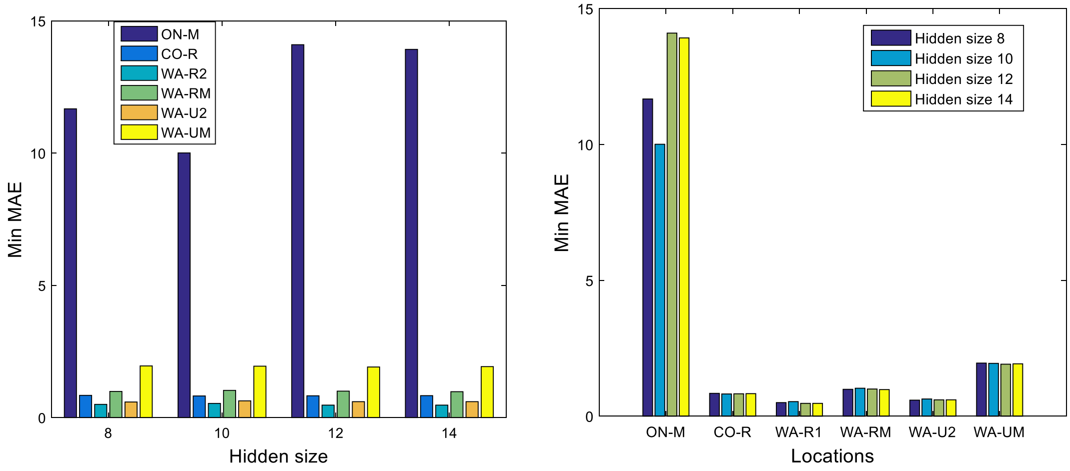
a normal DBN, which shows that the DBN is capable of learning the correct knowledge from the different input patterns. Besides, different unsupervised learning iterations from 50 to 80 were tested for this case, and the results of testing MAE and RMSE didn't change much, this proves that the DBN also has good generalization ability that is capable of reducing over-fitting when there is fewer input data.

Experiment 2: simultaneously trained global DBN model

In this experiment, data from all three cases described in Section "Case Descriptions" are used to train a global DBN, which is then compared to the local site specific NB models. As shown in Table 1, the Ontario case (ON-M) contains 2926 samples for training set and 836 for testing set. Similarly, Colorado highway (CO-R2) has 27,558 samples for training set and 9186 for testing set. In case of Washington highways, we categorised them into four groups (WA-R2, WA-RM, WA-U2 and WA-UM), and they together contained 40,760 samples for training set and 23,696 for testing set. Therefore, the total training data set for this global DBN includes 71,244 observations, with each including 13 features. Comparing to the Experiment 1, this experiment is to show if a DBN can be trained globally with multiple datasets from different regions and "act" locally. The network size with 8–14 neurons and unsupervised learning iterations of 50 and 80 were tested separately.

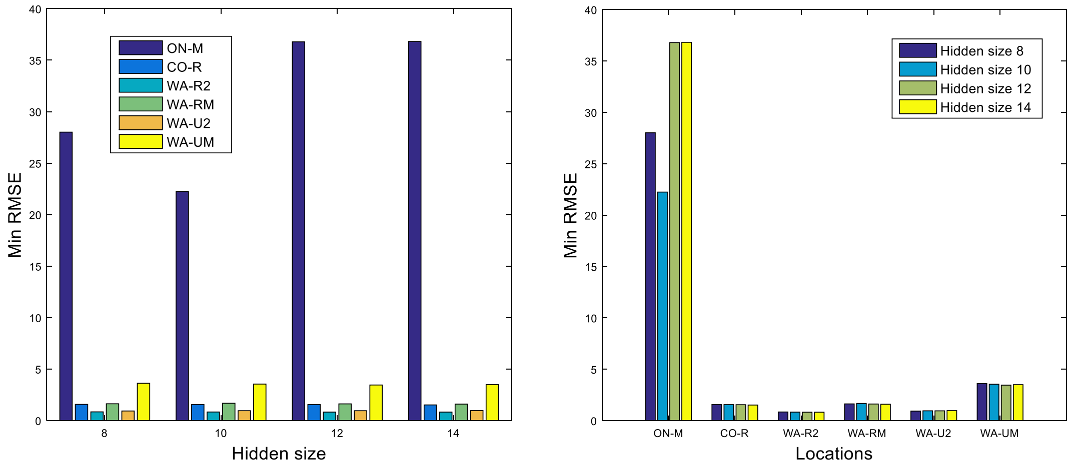
Fig. 5(a) shows the minimum testing MAE on the six highway types. It is obvious that for HWY401 (mostly rural highways), a global DBN with 10 hidden neurons has the best performance while for the other 5 locations the number of hidden neurons makes little differences. As shown in Fig. 5(b), a model with fewer hidden neurons has better performance for the Ontario and Colorado cases. In contrast, for highways in Washington State, better performance can be achieved when using more hidden neurons, especially for WA-R2 and WA-U2. Similar results can also be found in Fig. 6(a) and (b).

Table 8 compares the results of different models and training methods. The best performance was achieved when training separate DBN models; however, the globally trained model is still able to outperform the traditional NB model in most cases.



(a) (b)

Fig. 5. Testing MAE by changing hidden sizes and locations.



(a) (b)

Fig. 6. Testing RMSE by changing hidden sizes and locations.

Table 8
Comparison of models.

Models	NB (Base)	DBN training each case individually	Global DBN training All cases globally	
			Error	%Improvement
MAE ON-M	12.39	9.59	10.00	19.29
RMSE ON-M	28.94	19.58	22.24	23.15
MAE CO-R	0.83	0.81	0.82	1.20
RMSE CO-R	1.67	1.48	1.52	8.98
MAE WA-R2	0.50	/	0.47	6.00
RMSE WA-R2	0.79	/	0.82	−3.80
MAE WA-RM	0.98	/	0.98	0.00
RMSE WA-RM	1.56	/	1.59	−1.92
MAE WA-U2	0.66	/	0.59	10.61
RMSE WA-U2	0.93	/	0.93	0.00
MAE WA-UM	2.04	/	1.91	6.37
RMSE WA-UM	3.81	/	3.45	9.45

Note: For notations, see Section 4.1. "/" means no data.

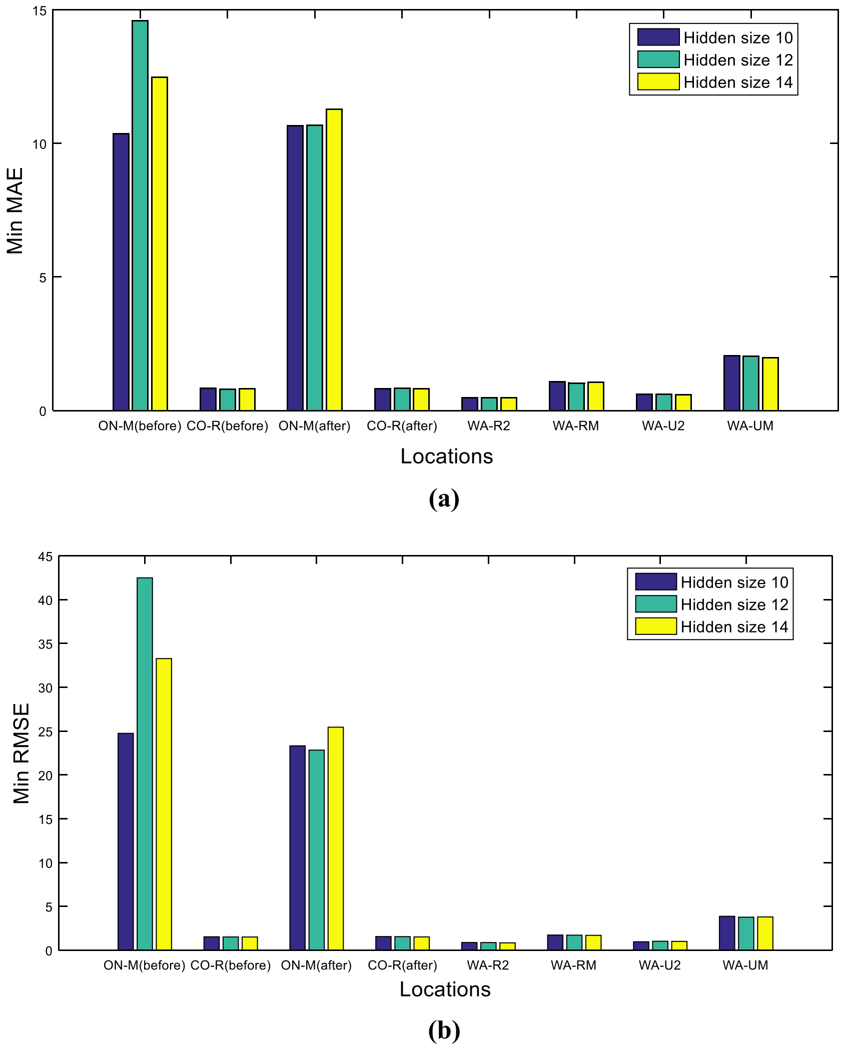


Fig. 7. Testing MAE and RMSE _ comparison of locations.

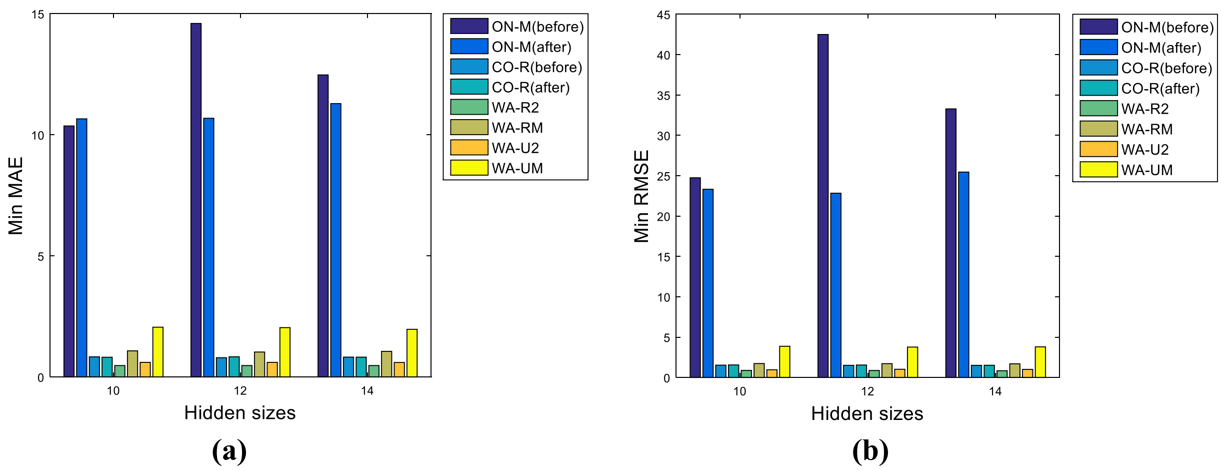


Fig. 8. Testing MAE and RMSE _ comparison of hidden sizes.

Table 9
Comparison of before and after training new database.

Models		NB (Base)	Global DBN Training Globally Before new data	Global DBN Training Globally After new data	
				Error	%Improvement
MAE ON-M		12.39	14.60	10.68	13.80
RMSE ON-M		28.94	42.49	22.24	23.15
MAE CO-R		0.83	0.80	0.82	1.20
RMSE CO-R		1.67	1.51	1.51	9.58
MAE WA-R2	New added data	0.50	/	0.47	6.00
RMSE WA-R2		0.79	/	0.83	−5.06
MAE WA-RM		0.98	/	1.00	−2.04
RMSE WA-RM		1.56	/	1.65	−5.77
MAE WA-U2		0.66	/	0.60	9.09
RMSE WA-U2		0.93	/	0.96	−3.22
MAE WA-UM		2.04	/	1.94	4.90
RMSE WA-UM		3.81	/	3.61	5.25

Note: For notations, see Section 4.1. "/" means no data.

The two exceptions are the results from the rural highway cases – WA-R2 and WA-RM, in which the global DNB was slightly outperformed by the local NB model.

Experiment 3: sequentially trained global DBN

This experiment is designed to test the generalization ability of a global DBN when trained sequentially with new datasets. We started the experiment by first training a global DBN with data from the Ontario and Colorado cases (Cases ON-M and CO-R). We then re-trained the model using a new dataset – data from the Washington case, includes four classes of highways: rural two-lane, rural multi-lane and urban multi-lane, and urban multilane. In this experiment, we also explored several retraining methods.

Fig. 7(a) shows the minimum testing MAE on the six types, in which the results of before retraining and after retraining for ON-M and CO-R are shown separately. It can be seen that a global DBN with 10 hidden neurons still has the best performance for ON-M (both before and after retraining), so is for CO-R and WA-R2. While for the rest kinds, more hidden neurons could be slightly better. Fig. 7(b) shows a little different pattern of testing RMSE. As can be observed, when a new database is added into the model, literally more training iterations are used in training the model, a larger model could imply better performance. Therefore, for this instance, the final network structure is 13–12–12–1 after testing different sizes. Similar results could also be observed in Fig. 8. Fig. 8(a) and (b) also show that when using a structure that is larger than 10 hidden neurons, the performance for ON-M is not satisfactory at first, this may be a sign of overfitting. However, it becomes much better after retraining, because after combined with the new database, the problem of overfitting is reduced.

Table 9 compares the results of different models and training methods. Large progresses are achieved in case of ON-M and ON-R. Besides, significant progresses are also achieved for WA-U2 and WA-UM, comparing to NB. There are also some decent in terms of testing RMSE of WA-R2 and WA-U2, and testing MAE and RMAE of WA-R2, which are the same highway types as ON-M and ON-R, so this may suggest that those areas are affected too much by the previous learned knowledge.

Conclusions

In this paper, we have investigated the potential of applying a deep belief network (DBN) – one of the most popular deep learning models for developing a global road safety performance function (SPF) for highways of different types from different regions. An extensive empirical investigation was conducted using three large real world crash data sets from vastly different highways and regions. It was shown that a single DBN could be trained globally with multiple datasets to predict the expected crash frequencies with a performance at least comparable to the traditional NB model.

This research finding suggests that, as compared to traditional approach which involves developing several local models separately, the proposed framework of developing a global model using DBN technique significantly reduces the modelling steps. Furthermore, the DBN technique was shown to have the flexibility to make use of newly available data for improved model accuracy, adaptation and automation, which would otherwise have been a very tedious process if the traditional NB modelling technique were to apply.

This research represents an initiating effort with several unsolved questions that need to be investigated in the future. For example, the robustness of the model performance to network structure and size needs to be further studied so that methods for determining the best network configuration could be devised for specific problems. The second issue is related to the need of adapting network structure to increased data size. Our preliminary analysis has shown the performance advantage of a fixed global DBN model as data size increases. There may be further performance gain with adjusted model structure. Lastly, we have not explored the potential of applying deep neural networks for addressing road safety problems by using the raw crash data.

Acknowledgement

This research is supported by National Science and Engineering Research Council of Canada (NSERC). Authors would like to thank Federal Highway Administration (FHWA) and Ministry of Transportation Ontario (MTO) for providing the datasets used in this study.

References

https://www.fhwa.dot.gov/research/publications/technical, 2017.
Highway Safety Manual (HSM), American Association of State Highway and Transportation Officials (AASHTO), 2010.
Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. J. Transport. Res. Board 2061 (1), 55–63.
Thakali, L., Fu, L., Chen, T., 2016. Model-based versus data-driven approach for road safety analysis: do more data help? In: Transportation Research Board 95th Annual Meeting. pp. 3516–3531.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.
Connors, R.D., Maher, M., Wood, A., Mountain, L., 2013. Methodology for fitting and updating predictive accident models with trend. Accid. Anal. Prev. 56, 82–94.
Yann, L., Bengio, Y., Hinton, G.E., 2015. Deep learning. Nature 521 (7553), 436–444.
Dahl, G.E., Dong, Y., Deng, L., 2011. Large vocabulary continuous speech recognition with context-dependent DBN-HMMS. In: Proc of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4688–4691.
Yichuan, T., 2013. Deep learning using linear support vector machines, arXiv preprint arXiv:1306.0239.
Taylor, G.M., Hinton, G.E., Roweis, S., 2006. Modeling human motion using binary latent variables, In: Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference, pp. 1345–1352.
Cheng, L., Geedipally, S.R., Lord, D., 2013. The poisson-weibull generalized linear model for analysing motor vehicle crash data. Saf. Sci. 54, 38–42.
Usman, T., Fu, L., Miranda-Moreno, L.F., 2012. A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level. Accid. Anal. Prev. 48, 368–378.
Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes methods. J. Transport. Res. Board 1840, 31–40.
Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accid. Anal. Prev. 43 (1), 461–470.
Chang, L.Y., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Saf. Sci. 43 (8), 541–557.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representation by back-propagating errors. Nature 232, 533–536.
Bengio, Y., Olivier, D., 2011. On the expressive power of deep architectures. Algorithmic Learning Theory. Springer, Berlin Heidelberg, pp. 18–36.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accid. Anal. Prev. 39 (5), 922–933.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.
Hinton, G.E., Srivastava, N., Krizhevsky, A., 2012. Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv: 1207.0580.
Ren, Z., Shen, F., Zhao, J., 2014. A model with fuzzy granulation and deep belief networks for exchange rate forecasting, IJCNN2014, Beijing, pp. 366–373.
Papandreou, G., Iasonas, K., Pierre-André, S., 2015. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 390–399.
Srivastava, N., Ruslan, S., 2012. Learning representations for multimodal data with deep belief nets. In: International conference on machine learning workshop.
Chen, H., Murray, A.F., 2003. A continuous restricted Boltzmann machine with an implementable training algorithm. IEEE Proc Vision Image Signal Process 3 (150), 153–158.
Bengio, Y., 2009. Learning Deep Architectures for AI. Foundations & Trends in Machine Learning 2 (1), 1–127.
Qiao, J., Pan, G., Han, H., 2015. The design and application of continuous deep belief network. Acta Automatica Sinica 12 (41), 2138–2146.
Pan, G., Qiao, J., Chai, W., Dimopoulos, N., 2014. An improved RBM based on bayesian regularization. In: 2014 IEEE International Joint Conference on Neural Networks (IJCNN), pp. 2935–2939.
Hinton, G.E., 2002. Training products of experts by minimizing contrastive divergence. Neural Comput. 14, 1771–1800.
www.deeplearning.net [Online], 2017.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44 (5), 291–305.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC.
Vogt, A., Bared, J.G., 1998. Accident Models for Two-lane Rural Roads: Segment and Intersections. Transport. Res. Rec.: J. Transport. Res. Board 1635, 18–29.
Hauer, E., 2015. The art of regression modelling in road safety. Springer International Publishing, Cham.
Council, F.M., Stewart, J.R., 1999. Safety Effects Of The Conversion Of Rural Two-Lane To Four-Lane Roadways Based On Cross-Sectional Models. Transport. Res. Rec.: J. Transport. Res. Board 1665 (1), 35–43.
Begum, M., Persaud, B., Nichol, S., 2009. Safety Performance of Ontario Road Segments. In: Proc 19th Canadian Multidisciplinary Road Safety Conference, pp. 8–10.
Ahmed, M., Huang, H., Abdel-Aty, M., 2011. Exploring a Bayesian Hierarchical Approach for Developing Safety Performance Functions for a Mountainous Freeway. Accid. Anal. Prev. 43 (4), 1581–1589.
Hinton, G., 2010. A practical guide to training restricted Boltzmann machines. Momentum 9 (1), 926.

Please cite this article in press as: Pan, G., et al. Development of a global road safety performance function using deep neural networks. International Journal of Transportation Science and Technology (2017), http://dx.doi.org/10.1016/j.ijtst.2017.07.004