Safety Prediction of Rail Transit System Based on Deep Learning

Yan Zhang*, Jiazhen Han†, Jing Liu*§, Tingliang Zhou‡, Junfeng Sun‡§, and Juan Luo‡
*Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, 200062, China
†College of Computer Science and Technology, Jilin University, Changchun, China
‡CASCO SIGNAL LTD., Shanghai, China
§Corresponding author: Jing Liu, jliu@sei.ecnu.edu.cn, Junfeng Sun, sunjunfeng@casco.com.cn

***Abstract*—The safety prediction of rail transit system is a fundamental problem in rail transit modeling and management. In this paper, we propose a safety prediction model based on deep learning for rail transit safety, which has been implemented as a deep belief network (DBN). It can learn effective features for rail transit prediction in an unsupervised fashion, which has been examined and found to be effective for many areas such as image and audio classification. To increase the accuracy of prediction, we introduce user satisfaction and rare-event probability, the new input prediction factors, into safety prediction. The former takes account of human and the latter is computed by statistic model checking. To show proof of the model, a real-world subway data sets based on the Beijing Metro in China is presented to demonstrate the feasibility of the model. Experiments on data sets show good performance of our prediction. These positive results demonstrate that deep learning and new factors are promising in rail transit research.**

***Keywords*—rail transit, safety prediction, deep learning, statistic model checking.**

I. Introduction

Rail Transit (RT) has expanded and become a core transportation system in China. Once the failure or accident is happened, consequence is serious. The safety prediction of RT system aims at estimating the system safety level given historical data. Existing mainstream approaches include: fuzzy fault tree technique [1][2], comprehensive fuzzy evaluation model [3], a neural network approach for railway safety prediction [4], and fuzzy reasoning approach (FRA) and fuzzy analytical hierarchy process (FAHP) [5], etc.

Most methods possess some failings. The fault tree analysis can only analyze the (0,1) state events. The FAHP based on the expert investigation method can be limited by the depth and breadth of the expert knowledge, the possible data, the industry prejudice and the professional preference. The macro prediction model based on statistical regression analysis method can have a greater error and is difficult to produce satisfactory results.

The shallow neural network can more effectively predict the system safety. It only has 3 layers and one hidden layer, therefore, it is difficult to express complex functions. However, as a kind of new artificial intelligence learning method, deep learning can overcome these shortcoming and has better performance than shallow neural network with more hidden layer. Through learning a deep nonlinear network, it can realize complex function approximation, and characterize distributed representation of input data. It has powerful characteristic extraction ability in data.

In this paper, we attempt to use deep learning for the safety prediction of RT system that learns features with limited prior knowledge. This is achieved by training a Deep Belief Network (DBN) according to the work of Hinton *et al.* [6][7]. The DBN has been found to be very effective in learning the representative features from the data in an unsupervised fashion [8][9]. It has been used to predict the traffic flow prediction [10]. For a RT system, such as a typical complex system, the DBN can help us in learning and capturing effective features without prior knowledge. Moreover, to make more precise prediction, we make some improvements in prediction model factors. We introduce user satisfaction and rare-event probability as new factors.

The rest of this paper is structured as follows. Section II presents the background knowledge of safety prediction of RT system and DBN. In Section III we introduce the analysis about choice of prediction model factors, the computational method of rare-event probability and safety prediction model of RT system based on DBN. Section IV presents our experimental results and their analysis. Finally, the conclusion and future works are described in Section V.

II. Background

A. Safety prediction

With rapid railway construction in China, including high-speed railway and subway, comes an urgent need for more accuracy and reliability safety analysis methods.

Subway is a kind of transportation with complex technology and crowded passengers, which is also a typical RT system. It needs a structured feed-back process that ensures improvement of current safety analysis based on knowledge of past accidents. If one or more stations were disrupted by accidents, there would be a great impact not only on the individual subway line but also on the whole subway network. Communication system has a high failure rate and passenger accidents is the second.

At present, in the field of system safety, system safety is usually measured by the frequency or severity of the accident. In order to measure the impact of the accident on the operation,

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

this paper refers to the reliability theory, and puts forward the operational safety expression of the following RT system.

Definition 1: System operational safety is the probability of the system to avoid the occurrence of accidents in the specified conditions and within the stipulated time.

$$S(t) = 1 - \frac{\sum_{i=1}^{N} T_i \alpha_i}{t} \tag{1}$$

where t is evaluation cycle, N is the number of accidents occurred in the t period, T_i is the time of the ith accident, and α_i the deviation factor of the ith accident. System operational safety is the function of accidents frequency.

Definition 2: Operation accident is all kinds of accidents occurred in the system. Accidents can be expressed in the following form.

$$I = \mathit{\Gamma}\{H, F, P, E\} \tag{2}$$

where H indicates casualties, F indicates equipments damage, P indicates property loss, and E indicates environment disruption.

Another preliminary is the measurements for the safety prediction of RT system. We use mean absolute percentage error (MAPE) to indicate the measurements. MAPE is used as the basic measurement in our study when we determine the model architecture. It is computed as

$$\mathrm{MAPE}(s, s') = \frac{1}{n} \sum_{i=1}^{n} \frac{|s_i - s'_i|}{s_i} \tag{3}$$

B. DBN

Deep Learning was proposed by Hinton *et al.* [6] in Science in 2006. It is also a neural network with multi-layer, compared with a shallow neural network. It has been successfully used in classification, regression, image recognition and feature extraction [8][11].

The DBN is the most common and effective approach among all deep learning models. It is a stack of RBMs. RBM has one layer of binary stochastic hidden units and one layer of binary stochastic visible units, Typically, all visible units are connected to all hidden units. The weights on the connections and the biases of the individual units define a probability distribution over the binary state vectors $\boldsymbol{v}$ of the visible units via an energy function [12]. The energy of the joint configuration $(\boldsymbol{v}, \boldsymbol{h})$ is given by:

$$E(\boldsymbol{v}, \boldsymbol{h}; \theta) = -\sum_{i=1}^{|V|} \sum_{j=1}^{|H|} w_{ij} v_i h_j - \sum_{i=1}^{|V|} b_i v_i - \sum_{j=1}^{|H|} a_j h_j \tag{4}$$

where $\theta = (\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a})$ is the parameter set, w_{ij} is the symmetric weight between visible unit i and hidden unit j, and b_i and a_j are their bias. The number of visible and hidden units is represented as $|V|$ and $|H|$, respectively. The probability that the model assigns to a visible vector $\boldsymbol{v}$ is:

$$p(\boldsymbol{v}; \theta) = \frac{\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v}, \boldsymbol{h})}}{\sum_{\boldsymbol{u}} \sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v}, \boldsymbol{h})}} \tag{5}$$

This configuration makes it easy to compute the conditional probability distributions.

Then, we can stack several RBMs together into a DBN. Each layer of hidden units learns to represent features that capture higher order correlations in the original input data. The key idea behind training a deep belief net by training a sequence of RBMs is that the model parameters, θ, learned by an RBM define both $p(\boldsymbol{v}|\boldsymbol{h}, \theta)$ and prior distribution over hidden vectors, $p(\boldsymbol{h}|\theta)$. Therefore, the probability of generating visible variables can be written as

$$p(\boldsymbol{v}) = \sum_{\boldsymbol{h}} p(\boldsymbol{h}|\theta) p(v|h, \theta). \tag{6}$$

After θ is learned from an RBM, $p(\boldsymbol{v}|\boldsymbol{h}, \theta)$ is kept. In addition, $p(\boldsymbol{h}|\theta)$ can be replaced by a consecutive RBM, which treats the hidden layer of the previous RBM as visible data. This way, it can improve a variational lower bound on the probability of the training data, as introduced in [6]. The DBN can be used as an unsupervised feature learning method if no labels are provided.

III. Prediction Model

A. Prediction Model Framework

The prediction model framework is shown in Fig.1. Model is mainly divided into three modules. We consider a subway system in this framework.

First, we collect data sample from subway operating company and decide the prediction model factors. We use these factors as input of DBN. The input data structure determines the number of nodes in each layer. The collected data sample include all kinds of accidents, basic information of train, and operation information of company. We choose prediction factors through analyzing these collection data.

Second, we establish the DBN based on processed data. We determine the number of nodes in each layer and the number of hidden layer through some experiments.

Last, we use data sample to train the DBN model and use test sample to confirm the accuracy of prediction model.

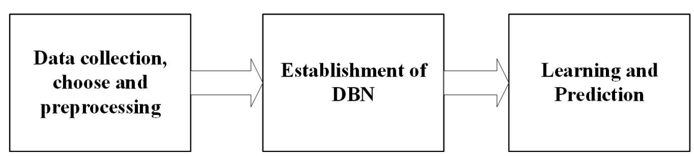

Fig. 1. Rail transit system prediction model framework.

B. Determination of Prediction Model Factors

Prediction factors are composed of many qualitative elements and complicated quantitative elements, and these are complex relationship. Therefore, human, equipment, environment and management are four major aspects in many analytic methods.

In "human" aspect, we mainly chooses average age of staff and user satisfaction as factors. According to survey, people in different ages has a different judgment on safety. There are obvious relationship between system safety and average age of staff. User satisfaction is further extended into the social environment. User satisfaction can be divided into functional satisfaction and non-functional satisfaction. The former indicates the users experience of system function, such as the capability of the candidate platform and the satisfaction of user and equipment interaction. The latter focuses on the objective factor of the system, such as passenger flow. Passenger flow in RT system refers to the sum of the passenger flow direction and number on RT line per unit time, which is the most significant feature reflecting passenger safety. We define user satisfaction as follows: $\Omega = \beta_c * \Omega_F + \beta_q * \Omega_o$ where Ω_F denotes the functional satisfaction, which includes platform satisfaction and the satisfaction of user and equipment interaction; β_c and β_q are weight coefficients; $\Omega_F = \Psi + I$, where Ψ represents the satisfaction of platform; and I denotes the satisfaction of interaction. Each corresponds to a value between 0 to 1. Regarding non-functional satisfaction, Ω_o is defined as follows: $\Omega_o = F$, where F denotes passenger flow.

In "equipment", we chooses rare-event probability in RT control system as a factor. Previous papers choose the failure rate of all part of system based on sample and survey data, rather than the simulation of the system. It can not analyze the security from the system itself, and the data from the outside will inevitably produce errors. Train control system mainly use The Communication Based Train Control System (CBTC). We simulate the movement authority (MA) in zone controller subsystem and get rare-event probability based on system operation randomly. We compute rare-event probability using statistical model checking method that introduced in Section III-C.

In "environment", we choose the number of nature disaster and man-made disaster in running section as factors. The former indicates the effect of natural environment, such as gale, rainstorm, flood and earthquake. They are also easy to cause secondary disasters. The latter focuses on the effect of human, such as many pessimistic people choose to suicide in the subway or create accidents to revenge society, or terrorist forces cause attacks by the explosion, shooting in the train.

In "management" we choose the rate of train punctual operation and staff's safety precautions training examination as factors. The former can be expressed by the ratio of the number of arriving on time and total within a year and we can observe the distribution of delayed trains in time and road sections through it. The latter is the key basic of safe management's implement. It can ensure effective handling of accidents and benefit to the safe operation of RT system.

C. Rare-event probability

CBTC is the main communication system in subway system. Zone controller (ZC) is the most important track-side subsystem. When a train moves into a ZCs control track, ZC protects the train by locking the zone avoiding other trains enter, and tell the train a distance it can run without dangerous. The zone that ZC keeps for the train called movement authority (MA). We consider a rare-event case in MA.

Rare event is the critical problem in RT safety system. Rare properties are often important but pose a particular challenge for simulation-based approaches. One of the main ways to solve this problem is statistical model checking (SMC) [13]. This technique blends randomized (*i.e.*, Monte Carlo) simulation, model checking, and statistical analysis, and it enjoys better scalability than other formal verification techniques. With statistical model checking one can compute approximations of the probability that a stochastic hybrid system satisfies a given temporal logic specification. However, estimating accurately rare-event probabilities using standard Monte Carlo techniques requires very high sample sizes, hence a key objective for SMC is to reduce the number and length of simulations. In the literature, two of techniques to cope with rare events is Importance sampling (IS) and The Cross-Entropy.

Importance sampling is based on the Monte Carlo Method, which can help us to reduce the variance of the Monte Carlo method. Here we just present the result of Importance Sampling. We define the Bernoulli random variable B that return 1 or 0. B has a certain probability p which we want to estimate whether $B = 1$ or not. Here we denote X as a random variable, meaning the value of $B = 1$ or not, and $I(\cdot)$ is an indicator function. For a given variable sequence $X_1, \ldots, X_N$, the crude Monte Carlo estimator $\hat{p}$ will converge to p as $N \to \infty$(with probability 1) by the strong law of large numbers. Here $\hat{p}$ is:

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} I_B(X_i) \tag{7}$$

and its variance is:

$$Var(\hat{p}) = \frac{Var(I_B(X_i))}{N} \tag{8}$$

For importance sampling to be efficient, it is thus crucial to find good importance sampling distributions without considering the entire state space. If we consider a more general situation combined with a random variable X, a measurable function $g : \mathbb{R} \to \mathbb{R}^{\geq 0}$ and an estimated variable $c = E[g(X)]$, the crude Monte Carlo estimator is

$$\hat{c} = \frac{1}{N}\sum_{i=1}^{N} g(X_i) \tag{9}$$

where $X_1, \ldots, X_N$ be random variables iid with density f. Its variance is

$$Var(\hat{c}) = \frac{1}{N}(E[g^2(X)] - c^2) \tag{10}$$

Suppose we had another distribution for X, with corresponding density f_*, such that the ratio $\frac{f}{f_*}$ is well-defined. We denote $W(x) = \frac{f}{f_*}$ as the weighting function or likelihood ratio. The Importance Sampling estimator is

$$\hat{c}_{IS} = \frac{1}{N}\sum_{i=1}^{N} g(X_i)W(X_i) \tag{11}$$

and its variance is

$$Var(\hat{c}_{IS}) = \frac{1}{N}(E_*[g^2(X)W^2(X)] - c^2) \tag{12}$$

Now that the problem becomes to find a density which can make $Var(\hat{c}_{IS})$ as small as possible. In fact, it's easy to verify that when the function g is non-negative the following optimal density results in a zero-variance estimator:

$$f_*(x) = \frac{g(x)f(x)}{c} \tag{13}$$

However, since the optimal one depends on $c = E[g(X)]$ which we are trying to estimate. Therefore, we should using a density close to the optimal one instead of the optimal one. This is the approach taken by the cross-entropy method.

The cross-entropy method was introduced in 1999 by Rubinstein. The method chooses the biasing density from the family such that the Kullback-Leibler divergence between the optimal biasing density and the chosen density is minimal. It has two basic steps:
1) find a density with minimal Kullback-Leibler divergence with respect to the optimal biasing density;
2) perform importance sampling with the biasing density computed in step 1 to estimate $E[g(X)]$.
For the limitation of the paper, we just show the result here. In the one-dimensional case we have that:

$$\hat{u}^* = \frac{E[g(X)X]}{E[g(X)]} \tag{14}$$

However, in statistical model checking $g(X_i)$ is either 1 or 0. Furthermore, in the rare event case it will be very unlikely to see a sample trace that satisfies the temporal logic property, which means that for reasonable sample sizes $\hat{u}^*$ would just give $\frac{0}{0}$. We can circumvent the problem by introducing the tilting parameter w, so the final result is:

$$\hat{u}^* = \frac{\sum_{i=1}^{N} g(X_i)W(X_i, w)X_i}{\sum_{i=1}^{N} g(X_i)W(X_i, w)} \tag{15}$$

D. Prediction Model Based On DBN

Many previous approaches of the safety prediction of RT system are shallow in architecture. Instead, we use deep learning method in this paper. For RT system such as a complex system, one single hidden layer usually would not be enough in describing the complicated relations between inputs and outputs. Deep learning can show its advantage in dealing with these complicated relations. In addition, it could learn features more profoundly.

Here, we employ a DBN for unsupervised feature learning for safety prediction of RT system. Recent work on deep learning has made training deep NNs more effective since Hintons breakthrough in 2006 [6]. Then, we use a sigmoid regression at the top layer in our approach so that we can perform supervised fine-tuning on the whole architecture easily.

Our model architecture for safety prediction of RT system is summarized in Fig.2. Input space X is, generally, the raw data we collected. Prediction model has seven input data, including all prediction factors. Seven raw data compose a group of input data. Input vector X can be represented as

$$X = \{A, U, P, N, M, R, E\} \tag{16}$$

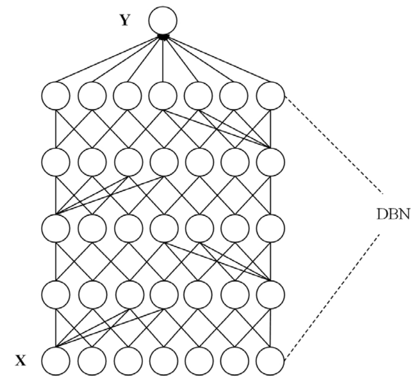

Fig. 2. Rail transit system prediction model on deep learning.

where A indicates average age of worker, U indicates user satisfaction, P indicates rare-event probability, N indicates numbers of nature disaster, M means numbers of man-made disaster, R means the rate of train punctual operation and E means staff's safety precautions training examination. So we set that the number of nodes in input layer is seven.

Unlike the binary RBM, as introduced in Section II-B, we replace it with real-valued units that follow normal distribution to model the RT system because the collected data is not 0 or 1. Moreover, since the RT system has its own model architectural feature, we have to introduce a correlation coefficient λ to take account into this feature. We observed the RT system model and produced λ from the result of lots of model simulation.

Energy function and conditional probability distributions are given as

$$E(\boldsymbol{v}, \boldsymbol{h}; \theta) = \sum_{i=1}^{|V|} \frac{(v_i - \lambda b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{|H|} \lambda a_j h_j - \sum_{i=1}^{|V|} \sum_{j=1}^{|H|} \frac{v_i}{\sigma_i} w_{ij} h_j \tag{17}$$

$$p(h_j|\boldsymbol{v}; \theta) = \sigma(\sum_{i=1}^{|V|} w_{ij} v_i + \lambda a_j) \tag{18}$$

$$p(v_i|\boldsymbol{h}; \theta) = N(\sigma_i \sum_{j=1}^{|H|} w_{ij} h_j + \lambda b_i, \sigma_i^2) \tag{19}$$

where σ is the standard deviation vector of normal distribution visible units, and $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ.

We can treat the DBN as a feature learning model. Each layer in the DBN is a process of nonlinear feature transformation. Features learned in the top layer of the DBN are the most representative features for modeling the data. The representative features are learned in an unsupervised way and can be used for various tasks, such as classification and regression. In our model architecture, the most representative features are used as the input vector for prediction (the top regression layer). Features learned from the DBN can be fine-tuned via error backpropagation on the whole structure by using labeled data for better prediction. It is also feasible if we

TABLE I
EFFECT OF THE NUMBERS OF LAYERS

Layers	MAPE	Times
1	0.823	0.09s
2	0.841	0.2s
3	0.878	0.34s
4	0.898	0.42s
5	0.931	0.57s
6	0.883	0.67s
7	0.854	0.84s

do not perform fine-tuning on the whole structure. Then, the DBN is the feature learning model, and the sigmoid regression layer is the prediction model.

IV. Experimental

In this section, we use the Beijing Metro data sample in 2014 to demonstrate the application of the safety prediction model to a real-world subway.

A. Experiment setting

Model Architecture: For all experiments, there are several parameters that we have to define for the RT system prediction model. For simplicity, the number of nodes in each layer is set to be the same. As the Section III-D said, input vector X is a seven tuple with seven factors. The input data structure determines the number of nodes in each layer. Therefore, we set nodes in layers = 7.

The number of hidden layer is key. The hidden layer's number parameter were optimized on the collected data set and then the best performing setting was used to compute the MAPE for the test set. We use MAPE to test the effect of the hidden layer's number parameter on our model architecture while keeping the other parameters fixed.

The issue of network size is one of the most typical problems for NN design. The learning time and the capability of the particular NN model are highly affected by the network size parameters. The result is reported in Table I.

Table I shows the MAPE and the training time with variations of the number of layers. In this experiment, the number of nodes in each layer is similarly fixed (7 here). The performance can be improved with the increase in layers from one to five. The training time show the spatiotemporal complexity of each model. They almost linearly increase with the increase in layers. Therefore, we set hidden layer size = 5.

Output we expect is the prediction value of RT system safety. Since, we employ sigmoid regression in the regression layer.

Data Sets: A data sample set is used in this study. A group of data X is collected in a runtime of the train. We collected 600 groups data and counted all kinds of accident, then the model output is a probability

The input (the feature vector X) in the model is 500 groups data. The output of the model is the prediction value of RT system safety with corresponding input. We use the data of 500 groups as the training set and 100 groups as the testing set.

TABLE II
RAIL TRANSIT SYSTEM SAFETY LEVEL

Level	explanation	Prediction Value
A	very safe	0.90-1.00
B	safe basiclly	0.80-0.89
C	not safe	0.75-0.79
D	dangerous	less than 0.75

We model and train the deep architecture in MATLAB2014a. Although deep learning costs a lot in storage and computation, most of the experiments can be finished in less than 1h in a single machine with Core i7 central processing unit, 8-GB memory,

B. Results of DBN

In this section, we investigate the learning and prediction capabilities of our model. We first give an illustration of the performance of our approach in Fig.3 The prediction of the safety and that of the real safety for the 100 groups data in the testing set are presented in Fig.3, We set the RT system safety level according to the different probability values. Our standard is in Table II.

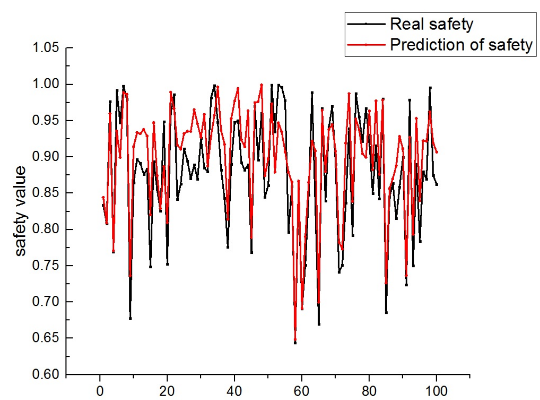

Fig. 3. Prediction of the safety of system and that of the real safety of system for the 100 groups data.

In Fig.3, Our approach almost follow the fluctuation of the real result, which means it can predict the safety of RT system in each case according to input accurately. The prediction of the RT system safety and that of the real RT system safety can match very well. It implies that the deep learning is useful in learning the patterns of a RT system. We computed the error of prediction and real is between 0.08 to -0.08 and almost of them are between 0.00 to -0.06. It is an acceptable error and can not cause to make wrong system safety level.

In conclusion, our prediction model is effective in safety prediction of RT system. It can imply the complex relationship of a RT system. The advantage of unsupervised feature learning would make this model easier for application. It is promising to apply deep learning into more RT research. Many related RT problems, such as RT induction and RT management, can employ the deep learning method for better results.

V. Related Work

Much research has been focused on the safety prediction method in recent years.

Our work is similar in spirit to a neural network approach for railway safety prediction [4]. In this paper, they have advocated the use of Artificial Neural Networks (ANNs) architecture to handle the prediction problem of the system malfunction or equivalently. They also have found out the best neural network structure and evaluated performances both in terms of mean squared error and correlation coefficient, but they only use ANN with two hidden layers due to the limit of artificial neural networks training algorithm. With the development of deep learning [6][7], many approaches are proposed based on it. Our approach provides a model based on deep learning.

In 2014, W. Huang *et al.* [10] have proposed a deep architecture that consists of two parts, i.e., a deep belief network (DBN) at the bottom and a multitask regression layer at the top. This is the first paper that applies the deep learning approach to transportation research. Our approach is inspired by it and has an application on RT safety prediction.

On a different line of research, H. Z. Huang *et al.* [2] have adopted fuzzy fault tree technique in railway traffic system safety analysis. In addition, in the fuzzy fault-tree model, the possibility of failure. However, fault tree only can relate to the basic and top events in each unit with (0,1) state. Jin, L *et al.* [3] have proposed a two-grade fuzzy synthetically evaluation model of RT safety system to evaluate the running safety of RT. They analyzed the various factors that affected the safety of RT. Then the indexes of safety evaluation system of RT were chosen by a conjoint way of integrating the qualitative and quantitative methods from four parts.

In the area of safety analysis in RT system, Analytical Hierarchy Process (AHP) is a classic method in safety analysis engineering. S Huang *et al.* [5] have presented a new safety risk assessment methodology for conducting systematic safety risk assessment using fuzzy reasoning approach (FRA) and fuzzy analytical hierarchy process (FAHP). This method can evaluate both qualitative and quantitative risk data and information associated with railway operation efficiently and effectively. Z Youpeng *et al.* [14] have presented a new risk assessment method for railway signal system using Failure Mode, Effects and Criticality Analysis (FMECA) and Fuzzy Analytical Hierarchy Process (FAHP). Although these work has good result, they are based on subjective judgment, so they can not guarantee that the calculated value is the objective value of each factor of the complex system. Unlike the traditional method, our proposed methodology attempt to prediction safety from the perspective of objective data analysis, excluding subjective factors and combining with the methods of deep learning.

VI. Conclusion

In this paper, we have presented a prediction model based on deep learning for RT safety prediction, which has been implemented as a DBN. DBN is effective for unsupervised feature learning. Meanwhile, we have some improvements in input prediction factors, which are user satisfaction and rare-event probability.

From experiments on data sets, we demonstrated that our prediction model has a high accuracy in RT system safety prediction. Moreover, the two factors we introduced can more effectively reflect the system safety and improve the prediction accuracy. It implies that deep learning is not only effective in neural-related areas such as image and audio recognition. For other complex systems, it is also useful.

In order to build a better safety prediction model, the train data sample can be increased and the applicaton in larger-scale system needs a further study.

Acknowledgment

This paper is partially supported by the projects funded by the NSFC Key Project 61332008 and NSFC 61572195. The Shanghai Trustworthy Computing Key Lab is supported by Shanghai Project SHEITC160306. The authors would like to thank the students in our laboratory and the engineers of CASCO SIGNAL LTD. for helpful discussions and the referees for greatly improving the manuscript.

References

[1] N. R. Storey, *Safety Critical Computer Systems*, 1996.

[2] H. Z. Huang, X. Yuan, and X. S. Yao, "Fuzzy fault tree analysis of railway traffic safety," in *International Conference on Transportation and Traffic Studies*, 2000, pp. 107–112.

[3] L. I. Jin, R. Song, and J. L. Jiang, "Safety evaluation of rail transit based on comprehensive fuzzy evaluation model," *Journal of Transport Science & Engineering*, 2011.

[4] S. Nefti and M. Oussalah, "A neural network approach for railway safety prediction," in *IEEE International Conference on Systems, Man and Cybernetics*, 2004, pp. 3915–3920 vol.4.

[5] S. Huang, M. An, Y. Chen, and C. Baker, "Railway safety risk assessment using fra and fahp approaches - a case study on risk analysis of shunting at waterloo depot," pp. 181–186, 2007.

[6] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets." *Neural Computation*, vol. 18, no. 7, p. 1527, 2006.

[7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." *Science*, vol. 313, no. 5786, pp. 504–7, 2006.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[9] C. Ekanadham, "Sparse deep belief net models for visual area v2," *Advances in Neural Information Processing Systems*, vol. vol 20, pp. 873–880, 2008.

[10] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[11] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks." *Journal of Machine Learning Research*, vol. 10, no. 10, pp. 1–40, 2009.

[12] Y. W. Teh and G. E. Hinton, "Rate-coded restricted boltzmann machines for face recognition," *Advances in Neural Information Processing Systems*, pp. 908–914, 2001.

[13] E. M. Clarke and P. Zuliani, "Statistical model checking for cyber-physical systems," in *International Conference on Automated Technology for Verification and Analysis*, 2011, pp. 1–12.

[14] Y.-P. Zhang, Z.-J. Xu, and H.-S. Su, "Risk assessment on railway signal system based on fuzzy-fmeca method," 2013.