

Study on Machine Learning Model for Comprehensive Evaluation of Rail Traffic Safety

Qiong Ren ^a, Hui Cheng ^b, Zhongyuan Xu

School of Mathematics and Computer Science, Jiangnan University, Wuhan, China

^aqren@163.com, ^bhuicheng2001@163.com

Abstract

In recent years, the rail transport has achieved rapid development. In assuming the responsibility of promoting the transport task of the economic development, it also faces greater challenges of the rapid growth of passenger volume and safety guarantee. To carry out research on rail traffic safety assessment has essential theoretical and practical significance in protecting public property and personal safety, promoting the development of city rail traffic, enhancing their level of service, and preventing and reducing accidents. The machine learning technology and security data index generated by traditional structured data are used to integrate, a comprehensive assessment model is adopted to study rail traffic safety state from a new perspective, which provides technical support for the comprehensive evaluation of railway transportation safety and provides decision support for rail transit operation.

Keywords

Rail Transit System; Safety Assessment; Machine Learning; Correlation Analysis.

1. Introduction

In recent years, with the continuous development of city construction and the continuous progress of the social economy, there is a sharp increase in the development of city population and the number of vehicles. Although the transportation industry has achieved great development, the contradiction between traffic capacity and traffic volume growth is still not relieved. In the exploration of solving this contradiction, it often takes measurements including increasing traffic construction, expanding the size of the network, making use of high technology and information technology tools to optimize the existing network transportation system and management system, as well as improving the management system, as shown in Figure 1 [1].

These measures promote the road transportation system information and intelligent development, but also bring the blowout of a lot of traffic information in the related industries system. Effective management of traffic information has become an important way to solve the contradiction [2]. As a result, in order to make large amount of basic data of the traffic management, traffic control, traffic planning, traffic safety and other aspects effectively and comprehensively utilized, thus the system information of the various departments play a greater role in the specific requirements of different applications, to provide decision support for different management departments and more convenient information for each people in the traffic system. In this paper, machine learning techniques is used to study rail traffic safety evaluation model, to obtain information that has strong target feature and high effectiveness from massive basic traffic data [3]. And on this basis, this paper studies rail traffic safety information fusion method and practice process, serving for rail transit operation and safety management.

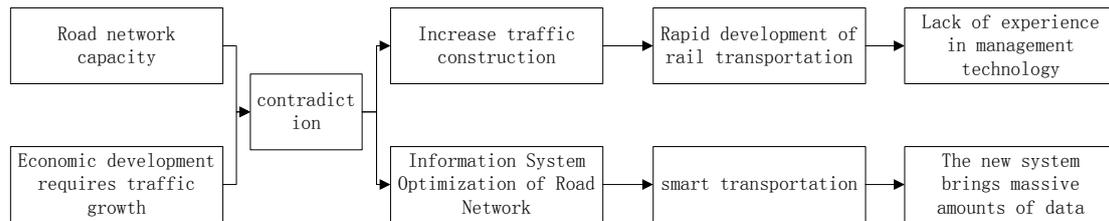


Figure 1. The contradiction between traffic capacity of road network and traffic growth

2. Machine Learning

Machine learning is generally defined as a process of self-improvement of a system. The subject of machine learning focuses on how the computer procedure automatically improves the performance with experience accumulated, so as to more accurately and efficiently complete the task of information processing. With the rapid development of information technology, the amount of information of various industries is growing exponentially, and thus how to obtain useful information from these massive and complex data becomes quite essential [4]. While the data processing under this data background, the traditional manual and computer processing mode has been stretched. How to let the computer have the ability to learn and to process new information data in the face of massive data has become a new topic in the field of information processing research. Based on this background, the machine learning technology was born and has become a hot research topic in recent years.

With the development of machine learning theory in recent years, researches on the application of machine learning in the field of production and life have gradually increased. It is mainly embodied in traffic volume prediction, traffic incident detection, traffic photo recognition, video processing and classification, traffic equipment production and so on [5-6]. Although machine learning techniques have been widely used in the field of transportation, the researches and applications of machine learning, especially support vector machine and association rule in rail traffic safety information processing, and safety comprehensive evaluation are still rare. To make use of this effective tool in the field of comprehensive evaluation of railway transportation safety has a certain research value.

3. Rail Traffic Safety Comprehensive Evaluation Machine Learning Model

3.1 Rail Traffic Safety Information Processing Flow Based on Machine Learning

In the rail traffic safety information processing process, in order to solve the existing problems in the rail safety, a set of information fusion algorithm suitable for rail traffic safety information is put forward, and then serve the security evaluation model [7]. First of all, it is necessary to analyze the characteristics of rail traffic safety. rail traffic safety information is a typical multiple-source heterogeneous data. rail traffic operation involves in many disciplines, with a wide range of knowledge, and the system contains a large number of device systems and data sources are very rich. At the same time, there are many participants in rail traffic. For the participants and managers, there are channels of information dissemination, but also the recipient of information [8]. In the meanwhile, rail transport operations need more than many industries to cooperate, so the data from the third parties and other industries are also very huge in amount. The real-time of security information lies in that rail transit running speed is rapid, and the degree of automation equipment is high, the system needs to process the data that the system collected timely, and carry out comparative analysis of data, to grasp the situation and trend of rail transit operation, thus to ensure safe operation of rail traffic. The massive feature is that because the rail traffic information source is wide, the system is complex, which involves in disciplines, industries, equipment, and it is required to consider the real-time feature of data, the information collection, transmission, processing, analysis, and release of the whole system are very huge. Based on the above discussion, the processing flow of traffic safety information is shown in Figure 2 [9].

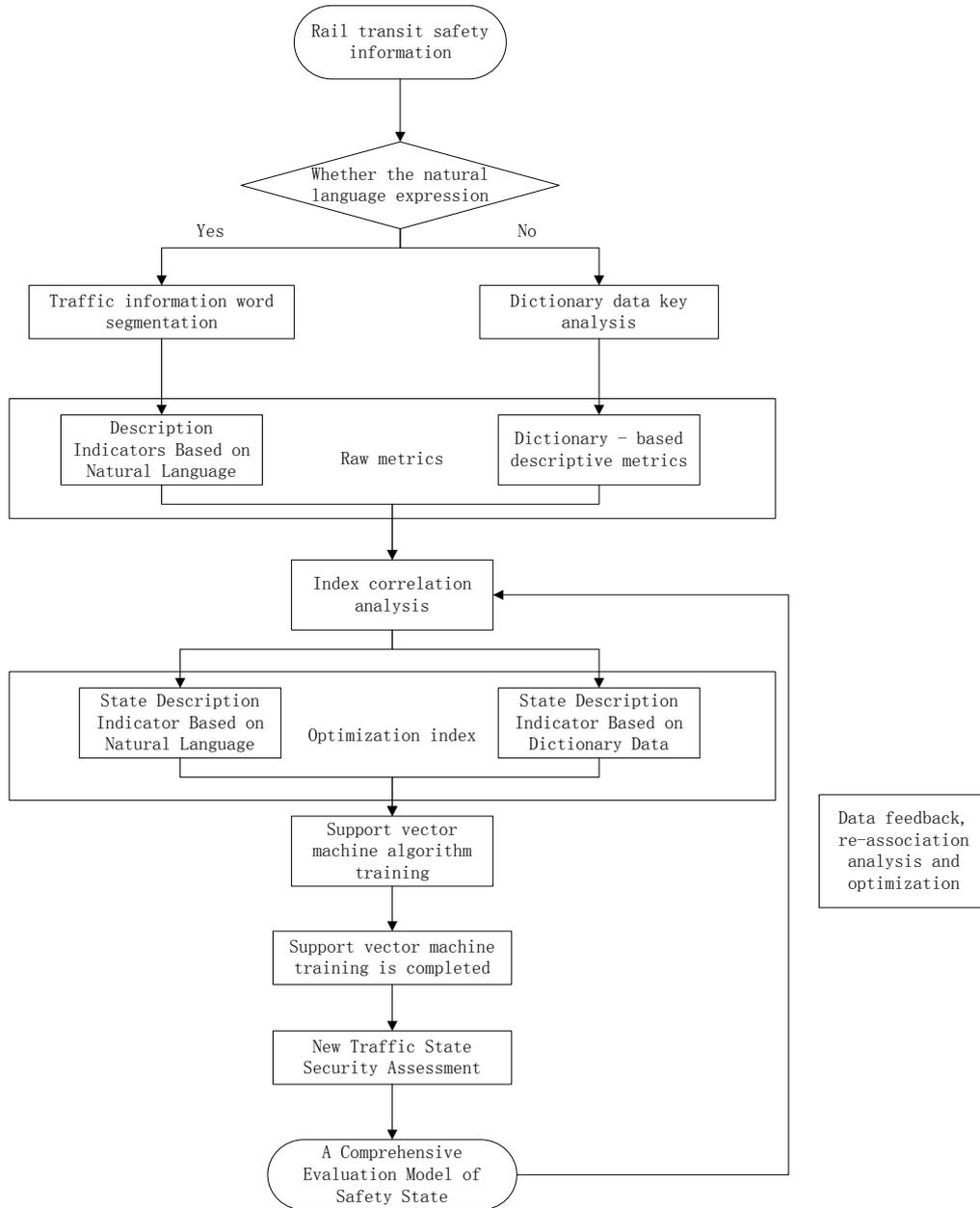


Figure 2. Traffic safety information processing flow

3.2 Construction of Comprehensive Evaluation Index System of Rail Traffic Safety

Through analysis of rail traffic safety information, two sets of rail traffic safety evaluation indexes are formed, as shown in Figure 3 [10]. One set is the index system formed for rail traffic information segmentation based on Chinese natural language, the other is a traditional data dictionary based on the existing dictionary data. These two sets of preliminary indicators will be the basis for the next rail traffic safety assessment.

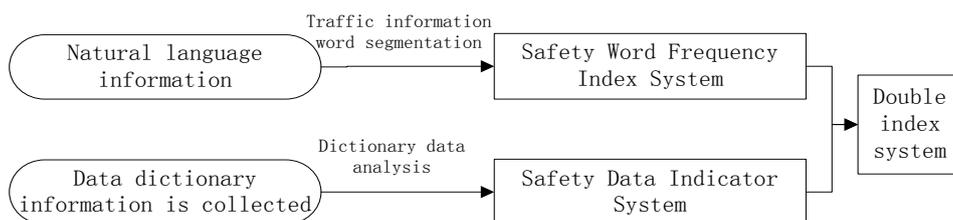


Figure 3. Dual index system for comprehensive evaluation of rail traffic safety

In the correlation analysis of index system of rail traffic safety comprehensive assessment completed based on association rules algorithm, the safety comprehensive security evaluation index system of a

given support degree is obtained, that is, safety words frequency index and safety data index [11]. For safety comprehensive security evaluation index system meeting the given threshold, in finite sample range, it is necessary to according to the existing data to evaluate the safety state. Here, it is the problem to be solved in the following to select a machine learning method that is suitable for finite sample, multidimensional cases, as well as has good generalization performance, so as to effectively carry out safety comprehensive evaluation of the rail traffic safety [12-13]. As previously described, support vector machine technology based on statistical learning theory, with the continuous development of machine learning disciplines itself, its importance and practicality are more important, and in many aspects of the industry application, it has a better performance. The safety index system after the connection of association rules processing, the support vector machine is applied to the comprehensive evaluation of rail traffic safety carried out according to the index system optimized [14]. This paper fully applies the global optimization, suitable for small sample, strong promotion and so on features of support vector machine to comprehensively evaluate the traffic safety state.

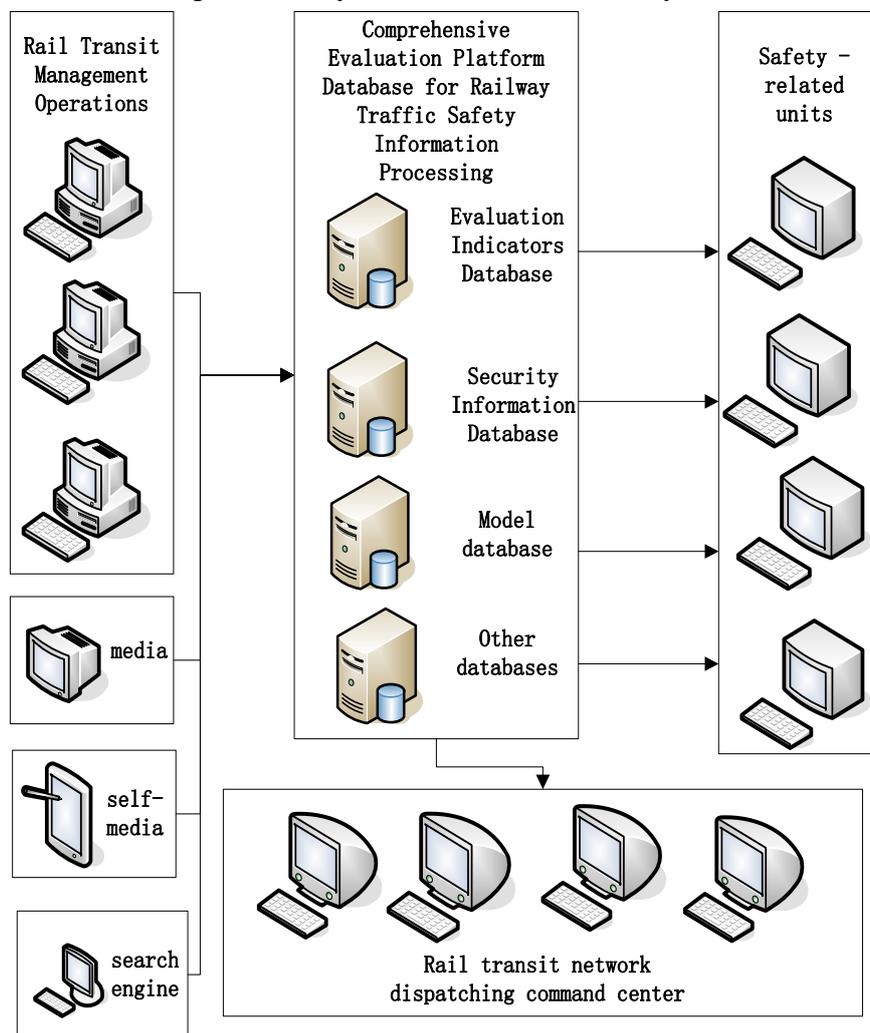


Figure 4. System structure

3.3 Research on Comprehensive Evaluation System of Rail Traffic Safety

The design of rail traffic safety comprehensive evaluation system based on B/S architecture, system in logic is divided into five layers from upper level to the lower level: presentation layer, application layer, security evaluation model layer, support layer, and basic data layer. Among them, the display layer uses imaged and easy to understand way to show the results of safety evaluation of rail transportation and the safety assessment of the interface to the user, which is directly responsible for the interaction with the user, to ensure that the interface is easy to use [15]. The application layer assumes the comprehensive

implementation of final function of the whole system, responsible for the final step processing of data, mainly including data maintenance and model maintenance. A comprehensive assessment is carried out on rail traffic safety combined with a variety of models of model layer, and system maintenance and other functions. The model layer is mainly responsible for the implementation of Chinese text segmentation model in text data researched in this paper, analysis of rail traffic safety index and safety index data word based on association rules, and the comprehensive evaluation model of railway transportation safety based on SVM [16]. The model layer is the core of the implementation of the algorithm in the system. The support layer of the system mainly contains to provide technical support for other layers, including the data preprocessing, cleaning, sorting, storage, providing interface and so on. The data layer includes the data of input system, including the national standard specification, rail traffic safety structure data of the input system, text data, as well as other possible input data. The system structure is shown in Figure 4 [17].

4. Conclusion

In the rail traffic safety information, the text data is widely used in unstructured data, such as text, image, video, and it has a high value of mining. But the current manual processing method is slow in speed, small in sending amount, and it cannot well combined with structured data. In this paper, through the research on the key technology of text mining, security data index that the traditional structured data forms are integrated by using machine learning technique [18]. And from a new perspective, it comprehensively estimates the rail traffic safety state. This paper establishes a comprehensive processing system of automatic comprehensive evaluation of rail traffic safety information based on machine learning, which can quickly and effectively process the rail traffic safety information, and it is unified in management and safe in storage, with certain theoretical innovations and practical application values, providing a theoretical basis for rail transportation management.

Acknowledgments

Hubei Provincial Department of Education Guidance Project of Scientific Research Program (No. B2016281).

References

- [1] Naghiyev A, Sharples S, Ryan B, et al. Real world verbal protocol data analysis of European Rail Traffic Management System train driving and conventional train driving[C]//Intelligent Rail Transportation (ICIRT), 2016 IEEE International Conference on. IEEE, 2016: 191-196.
- [2] Grabowska-Bujna B, Wiczorek S, Mikulski J. Rail Traffic Remote Control Systems Within the Areas Affected by the Occurrence of Mining Damage and Railway Safety [C] //International Conference on Transport Systems Telematics. Springer International Publishing, 2015: 264-273.
- [3] Podvezko V, Sivilevičius H. The use of AHP and rank correlation methods for determining the significance of the interaction between the elements of a transport system having a strong influence on traffic safety[J]. Transport, 2013, 28(4): 389-403.
- [4] Stolarski M, Żyłkowska J. The Analysis of the Collisions of Rail Vehicles with Animals with Regard to the Railway Traffic Safety[J]. Logistics and Transport, 2014, 22(2): 49-56.
- [5] Atchley P, Shi J, Yamamoto T. Cultural foundations of safety culture: A comparison of traffic safety culture in China, Japan and the United States[J]. Transportation research part F: traffic psychology and behaviour, 2014, 26: 317-325.
- [6] Liu J, Wang X, Khattak A J, et al. How big data serves for freight safety management at highway-rail grade crossings? A spatial approach fused with path analysis[J]. Neurocomputing, 2016, 181: 38-52.
- [7] Liu J, Bartnik B, Richards S H, et al. Driver behavior at highway rail grade crossings with passive traffic controls: a driving simulator study[J]. Journal of Transportation Safety & Security, 2015 (just-accepted): 00-00.

-
- [8] Xi X, Wei Z, Xiaoguang R, et al. A comprehensive evaluation of air pollution prediction improvement by a machine learning method [C] //Service Operations And Logistics, And Informatics (SOLI), 2015 IEEE International Conference on. IEEE, 2015: 176-181.
- [9] Fu Y, Zhang H, Bai Y, et al. An Under-sampling Method: Based on Principal Component Analysis and Comprehensive Evaluation Model[C]//Software Quality, Reliability and Security Companion (QRS-C), 2016 IEEE International Conference on. IEEE, 2016: 414-415.
- [10] Shrivastava V K, Londhe N D, Sonawane R S, et al. Reliable and accurate psoriasis disease classification in dermatology images using comprehensive feature space in machine learning paradigm [J]. Expert Systems with Applications, 2015, 42(15): 6184-6195.
- [11] Wang S, Chen S. A Comprehensive Evaluation of Well Completion and Production Performance in Bakken Shale Using Data-Driven Approaches[C]//SPE Asia Pacific Hydraulic Fracturing Conference. Society of Petroleum Engineers, 2016.
- [12] Khosravian M, Kazemi Faramarzi F, Mohammad Beigi M, et al. Predicting antibacterial peptides by the concept of Chou's pseudo-amino acid composition and machine learning methods[J]. Protein and Peptide Letters, 2013, 20(2): 180-186.
- [13] Velpuri N M, Senay G B, Singh R K, et al. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET[J]. Remote Sensing of Environment, 2013, 139: 35-49.
- [14] Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties[J]. Journal of the American Medical Informatics Association, 2014, 21(e2): e278-e286.
- [15] Wei Z, Wang W, Bradfield J, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease[J]. The American Journal of Human Genetics, 2013, 92(6): 1008-1012.
- [16] Srivastava P K, Han D, Ramirez M R, et al. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application[J]. Water resources management, 2013, 27(8): 3127-3144.
- [17] Meredig B, Agrawal A, Kirklin S, et al. Combinatorial screening for new materials in unconstrained composition space with machine learning[J]. Physical Review B, 2014, 89(9): 094104.
- [18] Mohabatkar H, Mohammad Beigi M, Abdolahi K, et al. Prediction of allergenic proteins by means of the concept of Chou's pseudo amino acid composition and a machine learning approach[J]. Medicinal Chemistry, 2013, 9(1): 133-137.