INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3266 - 3271
RECEIVED : 15.03.2018 PUBLISHED : 16.04.2018
April 16, 2018

EFFICIENT AND FAULT RECOGNITION OF TRAIN PROGRESS IN DATA MINING

[1]Prof.N.Karthigavani M.E., [2]Ms.K.Dhivya M.E.,

1. Assistant Professor, Department of the Computer Science and Engineering, AVS Engineering College, Salem – 636003.

2. Scholar, Department of the Computer Science and Engineering, AVS Engineering College, Salem – 636003.

ABSTRACT

An immense quantity of text data is recorded in the forms of revamp accurately in railway safeguarding sectors. Efficient text mining of such persistence statistics acting a central role in detecting anomalies and humanizing fault diagnosis effectiveness. Nevertheless, unstructured to the letter, high-dimensional data, and imbalanced fault class distribution pose challenges for attribute selections and fault opinion. Folks proposition a recognition feature extraction-based text mining that integrates facial appearance extracted at both syntax and semantic levels in the midst of the endeavor to advance the fault arrangement routine. Project first performs an enhanced $\chi 2$ statistics based attribute selection at the sentence structure level to triumph over the erudition intricacy caused by an unprovoked data set. subsequently, propose perform a preceding concealed Dirichlet distribution based attribute variety at the semantic echelon to diminish the data set into a stumpy dimensional matter space. To finish, this fuse fault features consequent from both syntax and semantic levels through consecutive fusion. The proposed scheme uses fault features at diverse levels in addition to enhances the correctness of fault identification for all blunder program, predominantly underground ones. Its concert has been validating by means of a railway maintenance data set unruffled commencing some duration by a railway corporation. It elsewhere performs long-established approaches.

1. INTRODUCTION

Intelligent Transportation Systems

Intelligent transportation systems (ITS) are advanced applications which, without embodying intelligence as such, aim to provide innovative services relating to different modes of transport and traffic management and enable various users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks. Although ITS may refer to all modes of transport, defined ITS as systems in which information and communication technologies are applied in the field of road transport, including infrastructure, vehicles and users, and in traffic management and mobility management, as well as for interfaces with other modes of transport

Intelligent Transportation Technologies

Intelligent transport systems vary in technologies applied, from basic management systems such as car navigation; traffic signal control systems; container management systems; variable message

signs; automatic number plate recognition or speed cameras to monitor applications, such as security CCTV systems; and to more advanced applications that integrate live data and feedback from a number of other sources, such as parking guidance and information systems; weather information; bridge de-icing (US deicing) systems; and the like. Additionally, predictive techniques are being developed to allow advanced modelling and comparison with historical baseline data. Some of these technologies are described in the following section.

Traffic-flow measurement and automatic incident detection using video cameras is another form of vehicle detection. Since video detection systems such as those used in automatic number plate recognition do not involve installing any components directly into the road surface or roadbed, this type of system is known as a "non-intrusive" method of traffic detection. Video from cameras is fed into processors that analyse the changing characteristics of the video image as vehicles pass. The cameras are typically mounted on poles or structures above or adjacent to the roadway. Most video detection systems require some initial configuration to "teach" the processor the baseline background image. This usually involves inputting known measurements such as the distance between lane lines or the height of the camera above the roadway. A single video detection processor can detect traffic simultaneously from one to eight cameras, depending on the brand and model. The typical output from a video detection system is lane-by-lane vehicle speeds, counts, and lane occupancy readings. Some systems provide additional outputs including gap, headway, stopped-vehicle detection, and wrong-way vehicle alarms.

Information fusion from multiple traffic sensing modalities

The data from the different sensing technologies can be combined in intelligent ways to determine the traffic state accurately. A Data fusion based approach that utilizes the road side collected acoustic, image and sensor data has been shown to combine the advantages of the different individual methods

2. RELATED WORKS

Technological advances in telecommunications and information technology, coupled with ultramodern/state-of-the-art microchip, RFID (Radio Frequency Identification), and inexpensive intelligent beacon sensing technologies, have enhanced the technical capabilities that will facilitate motorist safety benefits for intelligent transportation systems globally. Sensing systems for ITS are vehicle- and infrastructure-based networked systems, i.e., Intelligent vehicle technologies. Infrastructure sensors are indestructible (such as in-road reflectors) devices that are installed or embedded in the road or surrounding the road (e.g., on buildings, posts, and signs), as required, and may be manually disseminated during preventive road construction maintenance or by sensor injection machinery for rapid deployment. Vehicle-sensing systems include deployment of infrastructure-to-vehicle and vehicle-to-infrastructure electronic beacons for identification communications and may also employ video automatic number plate recognition or vehicle magnetic signature detection technologies at desired intervals to increase sustained monitoring of vehicles operating in critical zones.

Inductive loop detection

Inductive loops can be placed in a roadbed to detect vehicles as they pass through the loop's magnetic field. The simplest detectors simply count the number of vehicles during a unit of time (typically 60

seconds in the United States) that pass over the loop, while more sophisticated sensors estimate the speed, length, and class of vehicles and the distance between them. Loops can be placed in a single lane or across multiple lanes, and they work with very slow or stopped vehicles as well as vehicles moving at high speed.

Video vehicle detection

Traffic-flow measurement and automatic incident detection using video cameras is another form of vehicle detection. Since video detection systems such as those used in automatic number plate recognition do not involve installing any components directly into the road surface or roadbed, this type of system is known as a "non-intrusive" method of traffic detection. Video from cameras is fed into processors that analyse the changing characteristics of the video image as vehicles pass. The cameras are typically mounted on poles or structures above or adjacent to the roadway. Most video detection systems require some initial configuration to "teach" the processor the baseline background image. This usually involves inputting known measurements such as the distance between lane lines or the height of the camera above the roadway. A single video detection processor can detect traffic simultaneously from one to eight cameras, depending on the brand and model. The typical output from a video detection system is lane-by-lane vehicle speeds, counts, and lane occupancy readings. Some systems provide additional outputs including gap, headway, stopped-vehicle detection, and wrong-way vehicle alarms.

3. EXISTING SYSTEM

At the semantic level, we borrow the idea from and propose an LDA with prior knowledge (ab. PLDA) to perform the feature extraction. By representing documents in topics rather than word space, we are able to provide more feature extraction at the semantic level to compensate those extracted at the syntax level. The integration of prior knowledge with the basic LDA is based on the fact that LDA, as an unsupervised model, cannot deal with such issues as selecting topic counts and reducing the adverse effect of common words, which may not produce topics that conform to a user's existing knowledge. Prior knowledge helps us guide topic mining in basic LDA.

Disadvantage of the Existing System

- Unstructured verbatim, high-dimensional data, and imbalanced fault class distribution pose challenges for feature selections and fault diagnosis.
- The learning difficulty caused by an imbalanced data set.

4. PROPOSED SYSTEM

At the syntax level, we propose an improved $\chi 2$ statistics (ICHI) to cope with the feature selection of imbalanced data set. First, we overcome the negative effect of imbalanced data set by adjusting the feature weight of minority and majority classes. This makes minority classes relatively far away from the majority ones. Second, we consider the Hellinger distance as a decision criterion for feature selection, which is shown to be imbalance-insensitive. The proposed ICHI can be regarded as feature selections at the syntax level because it mainly uses the document-word matrix.

Advantage of Proposed System

- A prior latent Dirichlet allocation-based feature selection at the semantic level to reduce the data set into a low-dimensional topic space.
- Enhances the precision of fault diagnosis for all fault classes, particularly minority ones.

5. SYSTEM MODULES

A module is a part of a program. Programs are composed of one or more independently developed modules that are not combined until the program is linked. A single module can contain one or several routines.

Our project modules are given below:

1) User
2) Admin

Module Description

Generate Accident Report

This paper integrates methods for safety analysis with accident report data and text mining to uncover contributors to rail accidents. This section describes related work in rail and, more generally, transportation safety and also introduces the relevant data and text mining techniques.

Characteristics of Accident Report

This report has a number of fields that include characteristics of the train or trains, the personnel on the trains operational conditions (e.g., speed at the time of accident, highest speed before the accident, number of cars, and weight), and the primary cause of the accident.

This field has become increasingly important because of the large amounts of data available in documents, news articles, research papers, and accident reports.

Stored In databases:

Text databases are semi structured because in addition to the free text they also contain structured fields that have the titles, authors, dates, and other Meta data. The accident reports used in this paper are semi structured.

Step by Step Process:

User:

User Register the Accident details and casualty details.

All the details stored in the Database.

Admin:

Admin can verify the Accident details.

Predict the accident and casualty details.

CONCLUSION

Text mining of repair verbatim for fault diagnosis of railway systems poses a big challenge due to unstructured verbatim, high-dimension data, and imbalanced fault classes. In this paper, to improve the fault diagnosis performance, especially on minority fault classes, we have proposed a bi-level feature extraction-based text mining method. Propose first adjust the exclusive feature weights of various fault classes based on $\chi 2$ statistics and their distributions. Then we reselect the common features according to both relevance and Hellinger distance. This can be categorized as feature selection at the syntax level. Next, we extract semantic features by using a prior LDA model to make up for the limitation of fault terms derived from the syntax level. Finally, we fuse fault term sets derived from the syntax level with those from the semantic level by serial fusion. The proposed bi-level feature extraction method has been evaluated by RTP /RFP and F1-measure with a real data set collected by a railway company in China. The experiments show that the diagnosis results of the proposed feature fusion method, especially for minority fault classes, are much better than those of the traditional ones, such as $\chi 2$ statistics and information gain. Efficient feature fusion methods play an important role in feature extraction. Therefore, such powerful methods as parallel feature fusion should be further researched to improve the proposed method's performance Other merging learning methods should also be explored for better imbalanced classification.

REFERENCES

[1] D. G. Rajpathak, "An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain," Comput. Ind., vol. 64, no. 5, pp. 565–580, Jun. 2013.

[2] W. Wang, H. Xu, and X. Huang, "Implicit feature detection via a constrained topic model and SVM," in Proc. Conf. Empirical Methods Natural Lang. Process., Seattle, WA, USA, 2013, pp. 903–907.

[3] L. Yin, Y. Ge, K. Xiao, X. Wang, and X. Quan, "Feature selection for high-dimensional imbalanced data," Neurocomputing, vol. 105, pp. 3–11, Apr. 2013.

[4] Z. Zhai, B. Liu, H. Xu, and P. Jia, "Constrained LDA for grouping product features in opinion mining," in Proc. 15th Pacific-Asia Conf. Adv. Knowl. Discov. Data Mining, Shenzhen, China, 2011, vol. 1, pp. 448–459.

[5] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," J. Sound Vibration, vol. 335, pp. 367–383, Jan. 2015.

[6] L. Huang and Y. L. Murphey, "Text mining with application to engineering diagnostics," in Proc. 19th Int. Conf. IEA/AIE, Annecy, France, 2006, pp. 1309–1317.

[7] J. Silmon and C. Roberts, "Improving switch reliability with innovative condition monitoring techniques," Proc. IMechE, F C J. Rail Rapid Transit, vol. 224, no. 4, pp. 293–302, 2010.

[8] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.

[9] J. Chang, J. Boyd-Graber, C.Wang, S. Gerrish, and D. Blei, "Reading tea leaves: How humans interpret topic models," Neural Inf. Process. Syst., vol. 22, pp. 288–296, 2009.

[10] D. A. Cieslak and N. V. Chawla, "Learning decision trees for unbalanced data," in Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases-Part I. Berlin, Germany: Springer-Verlag, 2008, pp. 241–256.

[11] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Commun. Technol., vol. 15, no. 1, pp. 52–60, Feb. 1967.

[12] J. Yang, J. Yang, D. Zhang, and J. Lu, "Feature fusion: Parallel strategy vs. serial strategy," Pattern Recognit., vol. 36, no. 6, pp. 1369–1381, Jun. 2003.