Rule-Based Classification Approach for Railway Wagon Health Monitoring

G M Shafiullah, A B M Shawkat Ali, Adam Thompson, Peter J Wolfs

Abstract— **Modern machine learning techniques have encouraged interest in the development of vehicle health monitoring systems that ensure secure and reliable operations of rail vehicles. In an earlier study, an energy-efficient data acquisition method was investigated to develop a monitoring system for railway applications using modern machine learning techniques, more specific classification algorithms. A suitable classifier was proposed for railway monitoring based on relative weighted performance metrics. To improve the performance of the existing approach, a rule-based learning method using statistical analysis has been proposed in this paper to select a unique classifier for the same application. This selected algorithm works more efficiently and improves the overall performance of the railway monitoring systems. This study has been conducted using six classifiers, namely REPTree, J48, Decision Stump, IBK, PART and OneR, with twenty-five datasets. The Waikato Environment for Knowledge Analysis (WEKA) learning tool has been used in this study to develop the prediction models.**

Key Words – Railway wagons; classification algorithms; rule-based learning; WEKA.

I. Introduction

Classification is one of the most significant and popular machine learning areas, which is used to solve real world problems, especially in the fields of bioinformatics, medical diagnosis, vehicle and infrastructure monitoring, fraud detection, text classification and engineering fault detection. The goal of classification is to build a set of models with an input as a set of objects (i.e., training data), the classes which these objects belong to (i.e., dependent variables), and a set of independent variables. This algorithm always finds a rule or a set of rules to organise the data into classes [1-2].

Researchers already have proposed different types of classification algorithm, including decision tree induction, nearest-neighbour methods, error back propagation, rule-based learning, lazy learning and statistical learning [2-9]. However, it is really difficult to select a best suitable classifier for a specific application. The popular No Free Lunch (NFL) theorem [10] states that a more useful strategy is to gain an understanding of the dataset characteristics that enable different learning algorithms to perform well, and to use this knowledge to assist learning algorithm selection based on the characteristics of the datasets. Basically the established name of this process is called meta-learning. There were a number of proposals that have already been published to find out the most suitable classifier for a specific application. However, most of the comparative research on algorithms uses decision tree and neural network processes and places the emphasis on the percentage of correct classifications [11].

The statistical and logical learning algorithm (STATLOG) [12] project introduced a wide comparative analysis among classification algorithms on a large number of datasets with statistical analysis. The percentage of correct classifications and computational time has been considered in performing the analysis. Their analysis proves the basic idea of the NFL theorem that no algorithm is uniformly the most accurate [12]. Lim et al. [11] have considered twenty-two decision trees, nine statistical models and two neural network algorithms, and compared their performance on thirty-two datasets in terms of classification accuracy, training time and number of leaves (in the case of trees). They investigated the effect of adding independent noise attributes on the classification accuracy, and examined the scalability of some of the more promising algorithms as the sample size was increased.

With the increased demand for railway services, railway monitoring systems continue to advance at a remarkable pace to maintain reliable, safe and secure operations. The performance of rail vehicles running on tracks is limited by the lateral instability inherent in the design of the wagon's bogie steering system, and the response of the railway wagon to individual or combined irregularities [13-15]. Machine learning techniques have been introduced in different research projects to predict the typical dynamic behaviour of railway wagons running on the track [16-21]. Raw data collection, data pre-processing, and formatting are essential parts of developing any monitoring systems.

Li et al. [17] investigated a machine learning approach to automate the identification process of railway wheels defects using collected data from wheel inspections. Decision tree and Support Vector Machine (SVM) based classification schemes were used to analyse the railroad wheel inspection data. Li et al. [17] introduced a Bagging classification

Manuscript received February 5, 2010. This work was supported in part by CQUniversity, and in part by the Centre for Railway Engineering (CRE), CQUniversity.

GM Shafiullah is with the College of Engineering and Built Environment, Faculty of Sciences, Engineering and Health, CQUniversity, QLD 4702, Australia (g.shafiullah@cqu.edu.au).

A B M S. Ali is with the School of Computing Sciences, Faculty of Arts, Business, Informatics & Education, CQUniversity, QLD 4702, Australia.

A. Thompson is with the College of Engineering & Built Environment, Faculty of Sciences, Engineering & Health, CQUniversity, Rockhampton, QLD-4702, Australia.

P.J Wolfs is with Electrical and Computer Engineering, Curtin University of Technology, Perth, Western Westralia 6845, Australia.

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

ensemble approach especially for imbalanced data which boosted the prediction accuracy to 81 percent. The experimental results indicate that the proposed approach is very efficient, producing a classifier ensemble that has high sensitivity and specificity values during classification [17-18].

Duarte et al. [20] have analysed a data set extracted from a real-life vehicle tracking sensor network using popular classification algorithms. This data set has been extracted based on the sensor data collected during a real world wireless distributed sensor network (WDSN) experiment carried out at Twenty-nine Palms, CA. The WDSN vehicle classification problem comprises local classification and global decision fusion. Maximum Likelihood, *k*-Nearest Neighbour, and SVM algorithms were used in this experiment. It has been seen that, although the classification rates for the available modalities are only acceptable, methods used in multi-sensor networks such as data fusion will enhance the performance of these tasks.

In this study, the most suitable classifier has been proposed to develop a data acquisition model for railway using rule-based learning method. This model reduces power consumption of the railway monitoring systems as it needs only three sensor nodes instead of four required in an existing system to collect required data from railway wagons. Models have been developed with six popular classifiers and applied them to a unified platform. Initially the percentage of correct classification has been estimated. Later, rules have been generated with the help of statistical analysis to identify the best suitable classifier for this application. This paper is organised as follows: Section II discusses the background of the study. Section III presents an overview of the algorithms. The development of the model with different algorithms is discussed in Section IV. Results and analysis are described in Section V. Section VI concludes the article with future directions.

II. BACKGROUND OF THE STUDY

The "Health Card" system developed by a team of engineers at Central Queensland University [15, 22] aims to monitor every wagon in the fleet using low cost intelligent devices. Solid-state transducers including accelerometers and angular rate sensors with a coordinate transform to resolve car body motions into six degrees of freedom was used in the Health Card. An algorithm was developed to analyse signals from accelerometers mounted on wagon bodies to identify the dynamic interaction of the track and the rail vehicles. The algorithm has validated using collected field data including accelerations measured at strategic points on the wagon body and the bogies. Data was collected from ballast wagons, and dual axis accelerometers were fitted to each corner of the wagon bodies and each bogie side frame. The test run was a normal ballast laying operation, starting with a full load of ballast, travelling to the maintenance site, dropping the ballast on the track, and returning empty via the same route. A PC based data acquisition system was used to store data. The main purpose of the data acquisition was to provide real data that represented to the Health Card device. Data was to be used to validate and demonstrate the effectiveness of signal analysis techniques and finally develop a model to monitor typical dynamic behaviour and track irregularities.

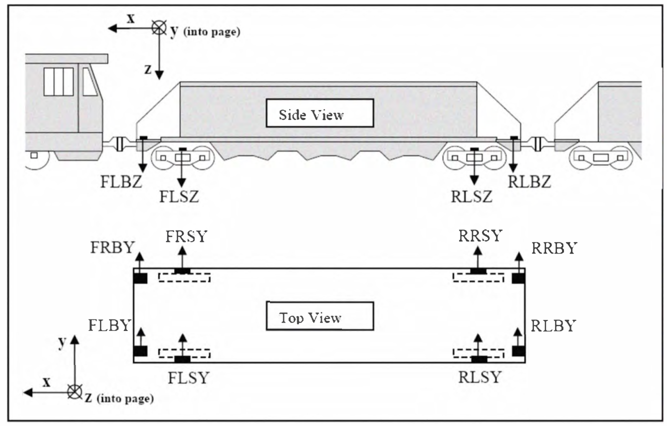

Figure 1: Accelerometer locations and axis naming convention [22]

Both the vertical and lateral conditions of the railway wagons have been measured by each accelerometer. The aim of the sensing arrangement was to capture roll, pitch, yaw, plus vertical and lateral accelerations of the wagon bodies. The ADXL202/10 dual-axis acceleration sensor measured 16 channels of acceleration data in *g* units, with 8 channels for each wagon body and 8 for its bogie side frames. Four sensor nodes were placed on each wagon body, and the locations of the sensors were front left body, front right body, rear left body and rear right body. Data collected from these four sensors are front left body vertical (FLBZ), front left body lateral (FLBY), front right body vertical (FRBZ), front right body lateral (FRBY), rear left body vertical (RLBZ), rear left body lateral (RLBY), rear right body vertical (RRBZ), and rear right body lateral (RRBY). Sensor locations and axis naming convention are illustrated in Figure 1.

Four sensor nodes were placed on each wagon's bogie side frames and the locations of the sensors were front left side frame, front right side frame, rear left side frame and rear right side frame. Data collected from these four sensors are front left side frame vertical (FLSZ), front left side frame lateral (FLSY), front right body vertical (FRSZ), front right side frame lateral (FRSY), rear left side frame vertical (RLSZ), rear left side frame lateral (RLSY), rear right side frame vertical (RRSZ), and rear right side frame lateral (RRSY). A field data acquisition model was developed for the wagon body and bogie side frames.

In an earlier work [23], a data acquisition model was proposed for the Health Card using machine learning techniques that improved the overall performances of the existing data acquisition system, in which the same amount of data was acquired using only three sensor nodes on each wagon body. The data of the sensor node located in the rear right corner of the wagon body was predicted in the study, i.e., RRBZ and RRBY, using the collected data referenced in [15, 22]. Therefore the model predicted the vertical and lateral conditions of the fourth sensor node, i.e., the sensor node located at the rear right corner of the wagon body. The best classifier was suggested based on relative weighted performance metrics. The prediction model replaced the use of the fourth sensor nodes in the rear right corner of the wagon bodies. From the experimental results stated in [23], it was observed that algorithm performance greatly depends on performance metrics. Finally, average weighted performance was estimated using classifier performance and computational complexity. From the final analyses, it was observed that computational complexity greatly affects the performance of the classifiers. Considering the overall situation and to improve the performance of the existing model, in this current study a rule-based approach has been introduced with the help of popular statistical analysis to select a unique classifier for developing prediction models for railway monitoring applications. Twenty five datasets have been selected considering track condition, wagon loaded and unloaded condition, data record etc. In addition to wagon body condition, this study also develops a model to predict the wagon side frame condition. This model improves the performance of the previous work, selecting a unique classifier to predict sensor data of railway wagons. This prediction model reduces power consumption of the existing application significantly as it reduces the requirement by one sensor node on each wagon body and one sensor node per wagon for bogie side frames. It is more energy-efficient than the existing data acquisition method developed by Central Queensland University, Australia [15].

III. ALGORITHM DESCRIPTIONS

Currently various classification and forecasting approaches are used to monitor railway operations to ensure safety and security. This section describes the popular classification algorithms used in this experiment to select a unique classifier for prediction of railway wagon condition. We have considered Tree-based learning reduced error pruning tree (REPTree), J48, and Decision Stump, Lazy-based learning IBK, rule-based learning PART, and OneR in this experiment [2 - 5].

PART: PART is a comparatively new algorithm for producing "decision lists", which are ordered sets of rules. It is developed by combining the C4.5 and RIPPER algorithms and is also called a partial decision tree algorithm. However, unlike C4.5 and RIPPER, PART does not have to perform global optimisation in order to generate rules. This algorithm works by forming pruned partial decision trees (built using C4.5's heuristics), and immediately converting them into a corresponding rule. It generates simple rules which are easily understandable [2, 5].

J48: J48 is a supervised learning algorithm developed by the developers of the WEKA package and is based on the widely-used C4.5 algorithm developed by J.R. Quinlan [3]. A decision tree is a tool for carrying out classification in three steps. First, the root node considers all the data instances as an input. Then each branch node generates the rules to do the classification task. Finally the leaf nodes introduce the class level.

REPTree: REPTree is a fast regression tree that uses information gain/variance reduction and prunes it using reduced-error pruning. It is also used as a classification tree. REPTree deals with missing values by splitting instances into pieces. Optimised for speed, it only sorts values for numeric attributes once. Pruning is used to find the best sub-tree of the initially grown tree with the minimum error for the test set [3].

IBK: Instance-based learning algorithms are derived from the nearest neighbour machine learning philosophy. IBK is an implementation of the k-nearest neighbour's algorithm. The number of nearest neighbours *(k)* can be set manually or determined automatically. Each unseen instance is always compared with existing ones using a distance metric. WEKA's default setting is $k = 1$. This algorithm performs well in application to artificial and real-world domains [3, 6].

Decision Stump: Decision Stump is a weak learning algorithm that consists of a decision tree with only a single branch. This learning algorithm builds simple binary decision "stumps" (1-level decision trees) for numeric and nominal classification problems. It deals with missing values by treating "missing" as a separate attribute value. Decision Stump is often used as components in ensemble learning techniques like bagging and boosting [5].

OneR: Based on a single attribute, OneR produces very simple rules. It is faster and useful in generating a baseline for classification performance. Real world databases contain very simple structured information about a domain as well, and these relationships can be parsimoniously detected and represented by OneR [2].

All these algorithms have been implemented in WEKA learning tools. The WEKA workbench [24] is a collection of state-of-the-art machine learning algorithms that is intended to make the application of machine learning techniques simpler and intuitive to a variety of real-world applications. WEKA is a very popular Java based set of machine learning tools. There are a large number of classification and regression algorithms built-in with WEKA that includes: Naive Bayes, Rule-based learning, Tree-based learning, Meta-based learning, Lazy-based learning, Statistical learning based algorithms, and Neural network based learning algorithms [1, 3, 24].

WEKA version 3.5.7 learning tools have been used in this study to evaluate the prediction accuracy. After pre-processing, the model has been developed using regression algorithms and measured performances using different attributes. The trained algorithm has been evaluated either with an additional test set or through k-fold cross validation, or by dividing the input data to a training and test set, considering records of data sets and experiment requirements

[3, 24]. In this study 10-fold cross validation test options have been used for experimental analysis.

IV. Experimental Setup

For experimental analysis J48, PART, REPTree, IBK, Decision Stump, and OneR classifiers are considered in this study. Twenty-five data sets were selected from the collected data in [22] to predict wagon body and bogie side frame condition. To cover a large experimental area, data sets were selected considering following metrics:

- train / track condition
- number of data records
- train location and time
- loaded and unloaded trains

Initially the percentage of correct classifications was measured for each of the algorithms with each of the twenty-five datasets, and the ranking performance for a given algorithm has been based on the percentage of correct classifications. Descriptive statistical analyses were conducted for each of the twenty-five datasets. Finally, rules have been generated with the help of ranking performance and statistical analysis.

Initially, with the help of WEKA [24] learning tools, the percentage of classification accuracy was estimated for each of the six algorithms for all of the twenty-five datasets. The configuration of the PC used in the experiments was a Pentium IV, 3.0 GHz Processor, 1GB RAM. The experiments demonstrated that the different algorithms predicted with minor to negligible errors.

The ranking performance for a given algorithm was measured based on the percentage of correct classifications. The best performing algorithm on each of these measures is assigned the rank of 1 and the worst is 0. Thus, the rank of the *j*th algorithm on the *i*th dataset is calculated as stated in [2]:

$$R_{ij} = 1 - \frac{e_{ij} - \max(e_i)}{\min(e_i) - \max(e_i)} \qquad (1)$$

where e_{ij} is the percentage of correct classification for the *j*th algorithm on dataset i, and e_i is a vector accuracy for dataset i. A detailed comparison of algorithm performance can be evaluated from this equation. The performances of all the algorithms were evaluated using the total number of best and worst performances.

To select the best classifier for railway wagon health monitoring, a data matrix has been constructed with the help of statistical descriptive analysis and ranking performance. Descriptive statistical analysis involved collecting information for each of the twenty-five classification problems. Twelve statistical measures are considered for descriptive analyses. Descriptive statistics are used to summarise the relevant characteristics of any large dataset. The considered descriptive statistics are stated in Table 1. Details about the above mentioned descriptive statistical terms are available in statistical books and MATLAB statistics toolbox [25].

Table 1: Descriptive statistics for characterisation of each dataset

Statistical Name	Symbolic Name
Geometric mean	*geomean*
Harmonic mean	*harmmean*
Trim mean	*trimmean*
Mean	*mean*
Median	*median*
Inter quartile range	*iqr*
Mad	*mad*
Range	*range*
Standard deviation	*std*
Variance	*var*
Kurtosis	*k*
Skewness	*s*

Finally, rules have been generated from the data matrix using the popular rule-based PART [5] algorithm which is built into the WEKA learning tools. Finally, a unique classifier was proposed for this railway application from the generated rules. Using similar procedures, models have been developed to predict the bogie side frame condition. Experimental results show that these models predicted similarly both for the wagon body and bogie side frame. Therefore, for experimental analysis in this paper, only the wagon body condition has been focused on.

V. Results and Analysis

In an earlier work [21], a prediction model was developed to predict rear right wagon body lateral and vertical conditions. Ranking performance, average accuracy and average weighted performance was evaluated to select a suitable algorithm for railway application. From different analyses the experimental results showed that no individual algorithm performs best for all of the performance metrics and has closely related to each other with minor to negligible error. To improve the performances of this analysis, in this study a rule-base learning approach was introduced to select the best suitable classifier for the same application. The proposed algorithms with a 10-fold cross validation approach were used to predict the rear right body condition of a railway ballast wagon. Models have been developed using the WEKA learning tools [24] with the six selected classifiers and twenty-five data sets. The percentage of correct classifications has been measured for each of the algorithms for the selected twenty-five datasets. Average percentages of classification accuracy for the twenty-five datasets are shown in Figure 2. From that figure it is seen that the average prediction accuracy is comparatively low as data sets with large variability were selected for the simulation. However, most of the datasets were predicted with higher accuracy as per the example which is shown in Figure 3. From Figures 2 and 3 it is observed that the J48 algorithm predicted the datasets with the highest accuracy.

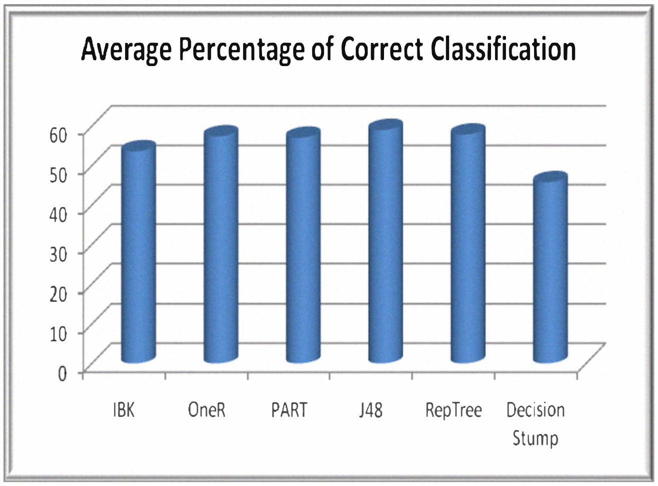

Figure 2: Average percentage of correct clasification for the selected twenty-five datasets

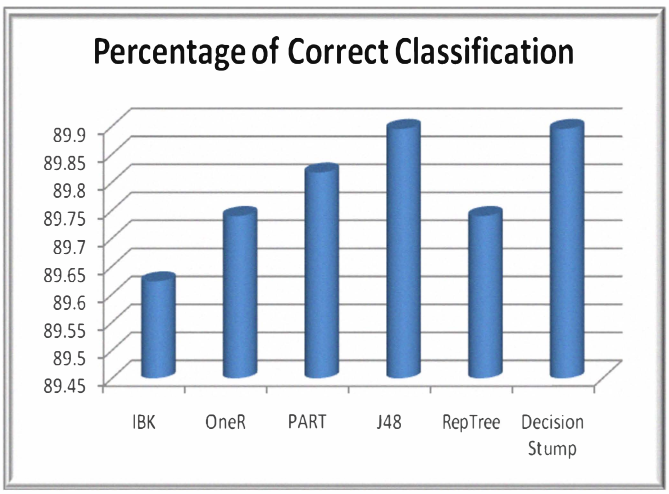

Figure 3: Observed percentage of correct classification for the 23rd dataset

The ranking performance for a given algorithm has been estimated using equation (1) for each of the data sets. Ranked algorithm performances for each of the datasets for different classifiers are represented in Table 2. Based on ranked performance the algorithms have classified into six classes that are presented in Figure 4. The algorithm that achieved rank 1 for the maximum number of data sets is classified as 1, and so on. For this experiment, J48 achieved rank 1 (best performance) for a maximum 12 datasets, and so J48 is classified as class 1. RepTree has rank 1 for 5 datasets and rank 0 (worst performance) for 4 datasets. On the other hand, OneR has rank 1 for 4 datasets and rank 0 for 1 dataset. Therefore, RepTree and OneR are classified respectively as class 2 and class 3. PART, IBK and Decision Stump are classified as class 4, class 5 and class 6 respectively. For the 23rd dataset, J48 and Decision Stump performed equally the best and are therefore both ranked 1.

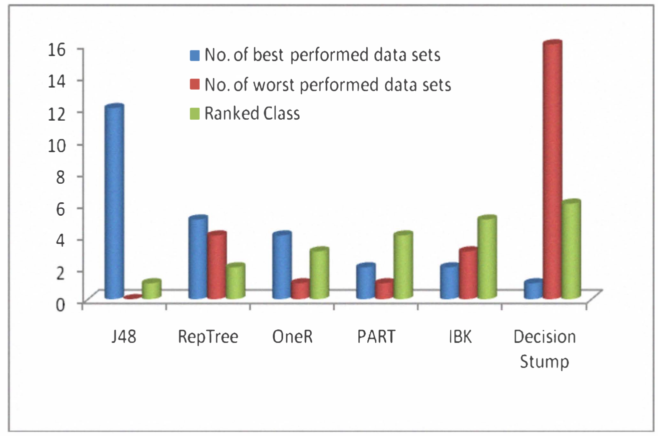

Figure 4: Classifier performance with number of best and worst performed data sets for each algorithm.

Table 2: Ranked algorithm performance for the six selected algorithms on each dataset

	IBK	**OneR**	**PART**	**J48**	**Rep Tree**	**Decision Stump**
DT1	0.8358	0.7164	0.8358	1	0.7164	0
DT2	0.7126	0.9655	0.7943	0.9425	1	0
DT3	0.5705	0.7724	0.6314	1	0.8077	0
DT4	0.4802	0.8885	0.7021	1	0.777	0
DT5	0.4703	0.8757	0.7622	0.9405	1	0
DT6	0.6457	0.5433	0.7953	1	0.7402	0
DT7	0.7043	1	0.9494	0.9961	0.9455	0
DT8	0.4859	1	0.8732	0.9648	0.9577	0
DT9	0.5646	0.988	0.8828	0.9713	1	0
DT10	0.5732	0.9071	0.8646	1	0.9465	0
DT11	0.4887	0.8576	0.8835	1	0.9903	0
DT12	0.6772	0.8418	0.7975	1	0.8987	0
DT13	0.4516	0.8802	0.788	1	0.8571	0
DT14	0.6482	0.9571	1	0.9929	0.9911	0
DT15	0.5392	0.9085	0.9314	1	0.9412	0
DT16	0.78	0.04	0.88	1	0.8467	0
DT17	1	0.3333	0.4167	0.25	0	0.3333
DT18	0	1	0.6155	0.5385	0.5385	0.4615
DT19	0.4285	1	0.4285	0.1428	0	0.2857
DT20	1	0.7499	0.5	0.3751	0	0.3751
DT21	0.0625	0	0.3125	0.6875	1	0.125
DT22	0	0.9328	0.6134	1	0.5798	0.8151
DT23	0	0.4285	0.7141	1	0.4285	1
DT24	0.15	0.9	0	0.85	1	0.85
DT25	0.5942	0.3768	1	0.5217	0	0.3913

Then, a data matrix has been constructed with the results of the statistical analysis and ranking of classifiers. Finally, using the same dataset for training and testing, rules have been generated to select the best classifier for this application. Rules have been generated using the PART algorithm which is built into the WEKA learning tools. PART has two significant parameters; confidence factor and minimum number of objects. The confidence factor is used for pruning the tree. The smaller values of confidence factor affect more pruning and higher values affect less pruning. The minimum number of objects represents the minimum number of instances per rule. The default values used in WEKA for confidence factor and minimum number of object are 0.25 and 2 respectively. The default parameters have tuned to select a suitable classifier for this railway application.

Accuracy of the classifier has been evaluated based on a confusion matrix. The generated rules and percentage of rule accuracy are summarised in Table 3. Experimental results show that the percentage of rule accuracy for J48 was 100%, with 80% for RepTree and only 57.4% for OneR.

Table 3: Generated Rule-Set

J48 Classifier:

IF k <= 1.169 AND var > 0.0027, THEN select J48

IF trimmean > 2.4288 AND s > 0.136, THEN select J48

OR,

IF k <= 1.169 AND var > 0.0027 OR IF trimmean > 2.4288 AND s > 0.136, THEN select J48

Rule Accuracy 100%

Rules for RepTree Classifier:

IF median > 2.4171 THEN select RepTree

Rule Accuracy 80%

OneR Classifier:

IF s <= 0.1433 AND k > 1.1639 AND k <= 1.1672

THEN select OneR

Rule Accuracy 57.4%

The default classifier for this problem is J48. If any data sets do not satisfy any of the algorithms, by default it uses the J48 algorithm as it performs better than any of the other algorithms.

VI. CONCLUSION

Classification algorithms play a key role to solve real world problems. Selection of an application specific classifier is an emerging research area. In this paper, an energy-efficient data acquisition method for railway monitoring has been investigated using popular classifiers. A prediction model has been developed with the help of WEKA learning tools to predict rear right wagon body and bogie side frame lateral and vertical conditions. Initially, the percentage of correct classifications has been measured in which J48 predicted most of the datasets with the highest accuracy. Later, ranking performance has been estimated to select a suitable algorithm for this application. The ranking performance has shown that J48 performs the best and Decision Stump performs the worst for the selected twenty-five datasets. However, no individual algorithm performs the best for all of the classifier problems. Therefore, a rule-based learning approach has been proposed for selection of a unique classifier. Rules have been generated based on statistical analysis and average ranking performance. This data acquisition method reduces power consumption of the existing application significantly as it reduces the requirement by one sensor node on each wagon and one sensor node on each bogie side frame. This also reduces computational complexity, and development and maintenance costs both in terms of hardware and human inspection.

Modern machine learning techniques are a new research topic, especially in railway monitoring and communication areas which still require further investigation that focuses on some specific areas including:

-analyses the performances of the models using Sensitivity, Specificity, *g*Means and ROC curves.

-introduction of bagging techniques to improve the performance of the model; and

-extension of the research with more problems from different domains with different classifiers.

REFERENCES

[1] A. B. M. Shawkat Ali and Saleh A. Wasimi, *Data Mining: Methods and Techniques*. Victoria, Australia: Thomson, 2007.

[2] Shawkat Ali and Kate A. Smith, "On learning algorithm selection for classification", *Journal on Applied Soft Computing, ELSEVIER,* vol-6, pp. 119-138, 2006

[3] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tool and Technique with Java Implementation", Morgan Kaufmann, San Francisco, 2000.

[4] G.H. John and R. Kohavi, "Wrappers for feature subset selection", *Artificial Intelligence,* Volume 97, Issue 1-2, December 1997.

[5] S. Jo Cunningham and G. Holmes, "Developing innovative applications in agriculture using data mining", Tech Report, Dept. of Computer Science, University of Waikato, New Zealand.

[6] D. Aha, "Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms", *Int'l Journal on Man-Machine Studies*, Vol. 36, pp. 267-287, 1992.

[7] "Linear Regression," GraphPad Software, Inc. San Diego, USA, Tech. Rep. [Online]. Available:

http://www.curvefit.com/linear_regression.htm as at 25th January 2008.
[8] A. O. Sykes, "An introduction to regression analysis", Chicago working Paper in Law and Economics, Tech. Rep. [Online]. Available: http://www.law.uchicago.edu/Lawecon/WkngPprs_01-25/20.Sykes.Regression.pdf as at 5th January 2008..
[9] G. D. Magoulas, V. P. Plagianakos and M. N. Vrahatis, "Neural network-based colonoscopic diagnosis using on-line learning and differential evolution", *Appl. Soft Computer*, vol-4, pp. 369-379, 2004.
[10] D.H. Wolpert and W.G. Macready, "No Free Lunch Theorems for search", Technical Report SFI-TR-05-010, Santa Fe Institute, Santa Fe, NM, 1995.
[11] T. S. Lim, W. E. Loh and Y. S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms", Machine Learning, v. 40, pp. 203-229, 2000.
[12] D. Miche, D. J. Spiegelhalter and C. C. Taylor, "Machine Learning, Neural and Statistical Classification", Ellis Horwood, New York, 1994.
[13] J. Smith, S. Russel, and M Looi, "Security as a safety issue in rail communications", in *Proc. of 8th Aus. Workshop on Safety Critical System and Software (SCS'03),* Canberra, Australia, pp. 79-88, 2003.
[14] V. K. Garg and R. V. Dukkipati, *Dynamics of railway vehicle systems*. Academic Press, 1984.
[15] P. J. Wolfs, S. Bleakley, S. T. Senini, and P. Thomas, "An autonomous, low cost, distributed method for observing vehicle track interactions", in *Rail Conf. 2006*, Atlanta, USA, April 2006.
[16] S. Nefti and M. Oussalah, "A neural network approach for railway safety prediction", in *2004 IEEE Int'l Conf. on Systems, Man and Cybernetics,* pp. 3915-3920, October 2004.
[17] C. Li., B. Stratman, and S. Mahadevan, "Improving railroad wheel inspection planning using classification methods", in *Proc. of the 25th IASTED Int'l Multi-Conf.*, Innsbruck, Austria, pp. 366-371, February 2007.
[18] C. Li, "Classifying imbalanced data using a bagging ensemble variation (BEV)", in *Proc. of the 45th annual Southeast Regional Conf.*, USA, pp. 203-208, March 2007.
[19] S. Kaewunruen and A. M. Remennikov, "Response and prediction of dynamic characteristics of worn rail pads under static preloads", in *Proc. of 14th Int'l Congress on Sound Vibration*, Cairns, Australia, July 2007.
[20] M. F. Duarte and Y. H. Hu, "Vehicle classification in distributed sensor networks", *Journal of Parallel and Distributed Computing,* vol.64, pp.826-838, 2004.
[21] GM Shafiullah, S. Simson, A. Thompson, P. Wolfs, and S. Ali, "Monitoring Vertical Acceleration of Railway Wagon Using Machine Learning Techniques", *Proceedings of the 2008 International Conference on Machine Learning; Models, Technologies and Applications (MLMTA'08)*, pp. 770-775, Las Vegas, Nevada, USA, 2008.
[22] Steven S. Bleakley, "Time Frequency analysis of railway wagon body accelerations for a low-power autonomous device", Master's thesis, Faculty of Engineering and Physical Systems, Central Queensland University, Australia, October 2006.
[23] GM Shafiullah, A. Thompson, P J Wolfs, S Ali, "Reduction of power consumption in sensor network applications using Machine Learning Techniques", TENCON 2008, Hyderabad, India, November, 2008.
[24] "Weka 3", The University of Waikato, New Zealand, Tech. Rep. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka as at 9th March 2008.
[25] "The Math Works", The Math Works, Inc. [Online]. Available: http://www.mathworks.com/ as at 9th January 2008.