Research Journal of Applied Sciences, Engineering and Technology 5(12): 3417-3422, 2013
DOI:10.19026/rjaset.5.4588
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 17, 2012 Accepted: November 23, 2012 Published: April 10, 2013

Research Article

A New Method for Traffic Forecasting Based on the Data Mining Technology with Artificial Intelligent Algorithms

[1]Wei He, [2]Tao Lu and [3]Enjun Wang
[1]Transportation Engineering Institute of Minjiang University, Fujian, 350108, China
[2]Hubei Province Key Laboratory of Intelligent Robot, College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, 430070, China
[3]Transportation Research Center, Wuhan Institute of Technology, 430073, Wuhan, China

Abstract: This study aims to investigate the traffic information forecasting based on the data mining technology. As well known, useful knowledge in traffic management system often hides in a large amount of traffic data. Generally, prior data pattern labels have been used to train the Artificial Neural Network (ANN) to identify the traffic conditions in the traffic information forecasting. The performance of the ANN models suffers from the prior information of the experts. To relieve this impact in the traffic information forecasting, a new ANN model is proposed based on the data mining technology in this study. The Self-Organized Feature Map (SOFM) is firstly employed to cluster the traffic data through an unsupervised learning and provide the labels for these data. Then the labeled data were used to train the GA-Chaos optimized RBF neural network. Herein, the GA-Chaos algorithm is used to train the RBF parameters. Experimental tests use practical data sets from the Intelligent Transportation Systems (ITS) were implemented to validate the performance of the proposed ANN model. The analyses results demonstrate that the proposed method can extract the potential patterns hidden in the traffic data and can accurately predict the future traffic state. The prediction accuracy is beyond 95%. Hence, the new data mining model can provide practical application for traffic information forecasting in the ITS system.

Keywords: Artificial neural network, data mining, optimization, traffic forecasting

INTRODUCTION

In recent years, there emerges a rapid development in the computer science and sensor technologies. As a result, there is a huge amount of data stored in the database ever than before (Zahra *et al*., 2010). The updating speed of data collection and storage in Intelligent Transportation Systems (ITS) is therefore very fast and a large amount of traffic data acquired by various sensors increases a lot of computer computation cost in the analysis of traffic information. Useful information has hidden in mass data. Using the data mining technology, it can find potential patterns of traffic activity and management to reduce the computation cost and enhance the traffic forecasting and control. It is therefore crucial to implement efficient data mining processing to discover important traffic rules and information to construct real-time and accurate traffic information system to help traffic status predicting and decision making.

In traffic forecasting and control, wireless sensor networks, cameras and high speed computers have been employed in current ITS systems (Nejad *et al*., 2009). The traffic volume, speed and occupancy data have been regarded as important features in traffic control and information management systems. Based on these traffic features, it is possible to develop models to predict and extrapolate the forthcoming traffic conditions (Wen and Lee, 2005). In general, the number of samples has great influence on the decision-makings. However, in real world the traffic data is extreme complex and the high dimensions of the data make classical statistical methods inefficient to provide a relatively good decision for the traffic forecasting and control. To overcome this problem, some new algorithms are imperative to analyze mass data and mine useful information. This procedure is the so called data mining technology. Lots of work has been done in traffic forecasting using data mining technology. Hauser and Scherer (2001) adopted clustering approach to manage urban traffic for the first time. Reasonable management scheme was obtained in their study. After that Park *et al*. (2003) employed Genetic Algorithm (GA) to solve the problem of unclean clusters and enhance the precision of the traffic forecasting. Following, the decision trees (Xu and Lin, 2009), Artificial Intelligent (AI) algorithms (Jia *et al*., 2006) etc., were introduced into the field of traffic forecasting management. However, most of the researches are limited for the purpose of accidents alarms. Very

Corresponding Author: Wei He, Transportation Engineering Institute of Minjiang University, Fujian, 350108, China

limited work has been done to connect the traffic features to the traffic conditions. However, the investigation on deep correlation of various traffic parameters is necessary for traffic forecasting management. A comprehensive understanding of potential traffic principals is important for correct traffic management decision-making. Although neural network models (Raahemi *et al*., 2008) were developed for digging the associated rules of the ITS database, the data was labeled in advance and the knowledge learning was under a supervised way. This is not realistic in practice because the classes of the data are difficult to determine before the data mining procedure (Li *et al*., 2010, 2011a, b, 2012a, b, c). More practical tools of finding the hidden knowledge in mass data stares us in the face.

In order to mine useful information hidden in mass ITS data for the traffic information forecasting, a new hybrid intelligent data mining model is proposed in this study based on Self-Organizing Feature Map (SOFM) and GA-Chaos optimized RBF neural network. The SOFM was firstly used to label potential clusters hidden in the ITS data base through an unsupervised manner. Then the labeled clusters were treated as feature patterns to train the RBF neural network for traffic forecasting. To optimize the RBF model, the GA-Chaos algorithm was used to optimize the RBF parameters. Empirical study on the ITS data has prove that the new method is a useful tool for traffic forecasting and control.

DESCRIPTION OF THE PROPOSED PREDICTION MODEL

Data mining technology is a hottest topic in fields of database statistics. It aims to analyze and mine knowledge from mass data sets (Nejad *et al*., 2009). By data mining, some useful features associating traffic flow trend can be revealed from the ITS data warehouse. Thus, the traffic features can be transformed into readable information to enhance the traffic information forecasting and traffic control.

Figure 1 shows a typical Intelligent Transportation System (ITS). It includes ITS data source module, data warehouse module, data mining module and Decision Support System (DDS) module. Data mining is one of its key techniques in this traffic information system. It is the basic of the Decision Support System (DDS) module, which is respond to correct traffic information forecasting and traffic control. Hence, it is crucial to establish efficient data mining method for the ITS system. For this reason, the SOFM and RBF neural networks are applied for intelligent data mining for ITS system in this study.

Self-Organizing Feature Map (SOFM): SOFM is proposed by Kohonen (1990). It is a powerful tool for pattern recognition using unsupervised learning. Due to hidden patterns in the ITS data is unknown, the SOFM is very suitable in this case. The SOFM can find useful information contained in the ITS database to identify

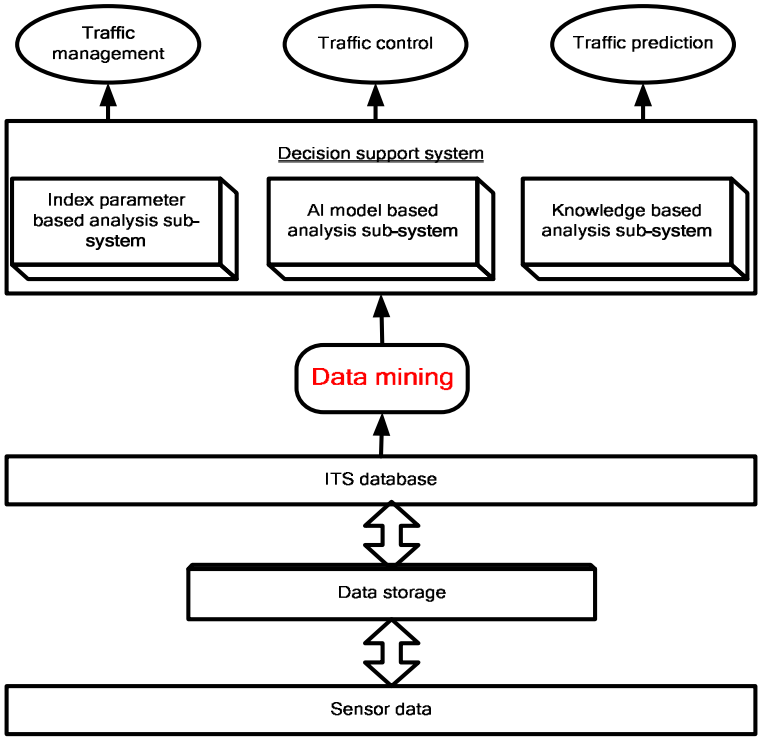

Fig. 1: Typical control and management framework of ITS

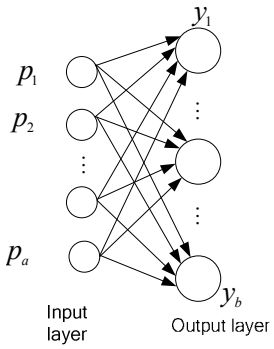

Fig. 2: The structure of SOFM neural network

potential clusters automatically (Jiang *et al*., 2010). By doing so, the labeled clusters can be used to instead of man-made labels and hence avoid the shortcomings of expert experience.

The theory of SOFM is fully discussed in Kohonen (1990). The SOFM is a two-layer neural network. The first layer is the input layer. The second layer is the output layer which contains neurons arranged in a rectangular pattern. By Kohonen learning (Kohonen, 1990), the SOFM can automatically find clusters in the input data if they exist. Figure 2 shows the structure of SOFM, where p_i (i=1, 2, ..., a) are the input variables and y_i (i=1, 2, ..., b) are the output neurons.

GA-Chaos optimized RBF neural network: The RBF neural network has good nonlinear mapping capability and hence is suitable for the traffic information forecasting. The performance of RBF network will be influenced by the hidden node number, the central values and the width of the base function. The Genetic Algorithm (GA) is used to optimize these parameters in this study. GA has three operators: selection, crossover and mutation. The goal of these operators is to pick out the new vitality strong fitness. However, the best fitness is not always easy to obtain. Sometimes GA may fall into local extreme, i.e., premature problem. In review of mechanism, the premature is mainly caused by lack of effective gene in offspring. In order to make the GA avoid premature, this study adopts chaos optimization technology to achieve this goal. Chaos optimization search is able to help GA in the search process to avoid local extreme. A common used chaos optimization principle is Logistic sequence (Krishna, 2012). It firstly the maps the chaotic variables into the solution space by Logistic. Secondly, it searches the characteristics of chaotic variables which are of ergodicity, randomicity and regularity. The initial population can be generated by chaotic sequences which in a certain extent can improve the searching efficiency of genetic algorithm. The mapping expression of Logistic is:

$$x_n = \mu x_{n-1}(1 - x_{n-1}) \quad (1)$$

where, μ is the control parameters and the system is in chaos situation when $\mu = 4$ and the chaos optimization process is as follows.

Firstly give any initial x_0 and the N chaotic variables, x $\{X_1, X_2, \ldots, X_n\}$, with different paths.

Secondly, the i chaotic variables are mapped into solution space by the first carrier:

$$y_{in} = c_i + d_i x_{in} \quad (2)$$

where, c_i and d_i are constants.

Then set the current best points y^*. If the optimal value is f^*, make $y^* = y_0$ and $f^* = f_0$. If f^* remains constant pass N iterations, the second carrier is:

$$y_{im} = x_i^* + \alpha_i (x_{im} - 0.5) \quad (3)$$

where,
m = The iterative step
α_i = A constant
x_{im} = Smaller chaotic variables in traversal area
y_{im} = The searching result

Until y_{im} satisfies the terminate qualification, it gives the optimal solution y^* and the optimal value f^* Through this process, the chaos algorithm can effectively find reasonable genetic operation parameters, help genetic algorithm jump out of local extreme.

Thus, the optimization process of chaos-genetic-RBF can be expressed as follows:

i. GA chromosomes are coded by hidden nodes number and the base function of central values and the width of the RBF networks
ii. Initialize chromosomes and set genetic operation parameters
iii. Calculate the corresponding fitness
iv. Do crossover and mutation
v. Decode newborn progeny populations to obtain the corresponding fitness
vi. The optimal individuals in groups are optimized by chaotic algorithm. If the searching result is bigger than the original fitness, substitute the individual
vii. If the results satisfy the termination conditions, stop to the end, Otherwise return to (iii) for iteration.

The principle of the proposed data mining method: Figure 3 shows the framework of the proposed data mining method for ITS system.

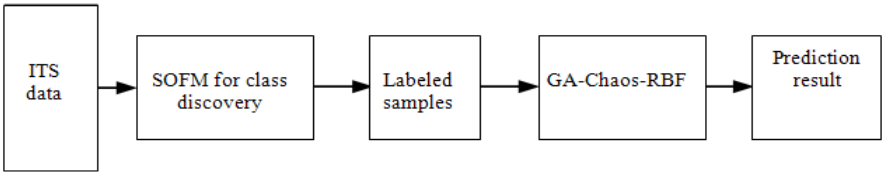

Fig. 3: Data mining method for traffic information prediction

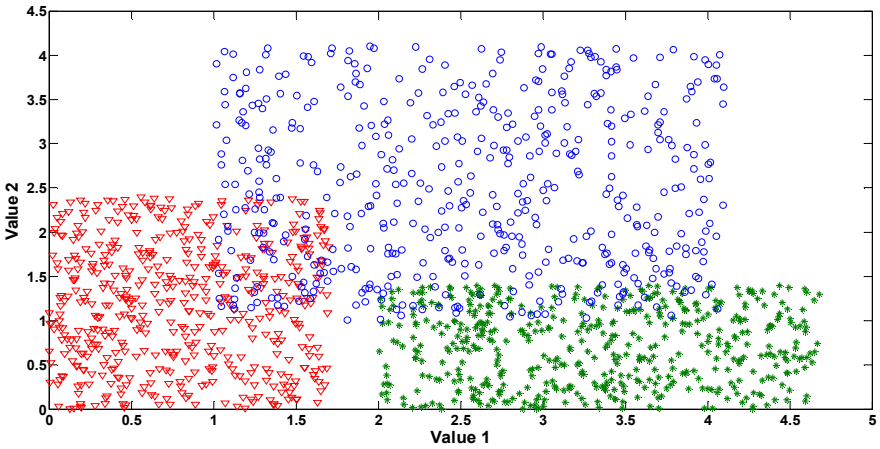

Fig. 4: Data mining of the ITS data using SOFM

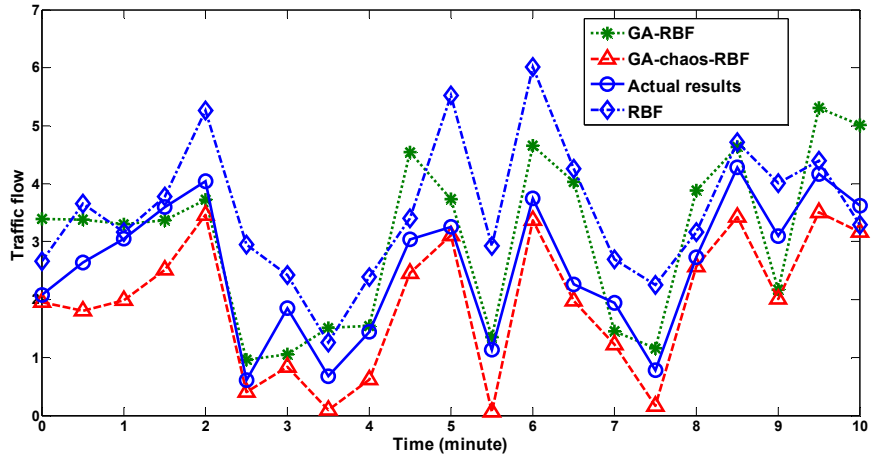

Fig. 5: Prediction results of small traffic flow

EXPERIMENTAL ANALYSIS

A set of ITS data has been used to validate the new method in this study. Here, 1500 data sets with unknown patterns were prepared for the traffic forecasting. The SOFM was firstly employed to cluster the ITS data. The input variables of the SOFM were traffic flow series and the output adopted 30 neurons. The data mining results are shown in Fig. 4. It can be seen in the figure that the ITS data can be clustered into 3 groups. Then we analyzed the ITS data and found that these three clusters represented small, middle and large traffic flow, respectively. This cluster result agrees well with the physical truth of the testing ITS data. The classification result indicates that the hidden patterns can be identified efficiently by the SOFM and hence a reliable ANN model can be constructed with those labeled groups. In this study, the three clusters are used to train the RBF models to predict the small, middle and large traffic flow, respectively.

In order to forecast the traffic flow, we use the three labeled clusters to construct three RBF models to predict small, middle and large traffic flow, respectively. The prediction results are shown in Fig. 5 to 7. The comparison of the RBF, GA- RBF and GA-chaos-RBF has been implemented in the traffic flow forecasting. The prediction rate of the RBF model is 83.5%, the prediction rate of the GA-RBF model is

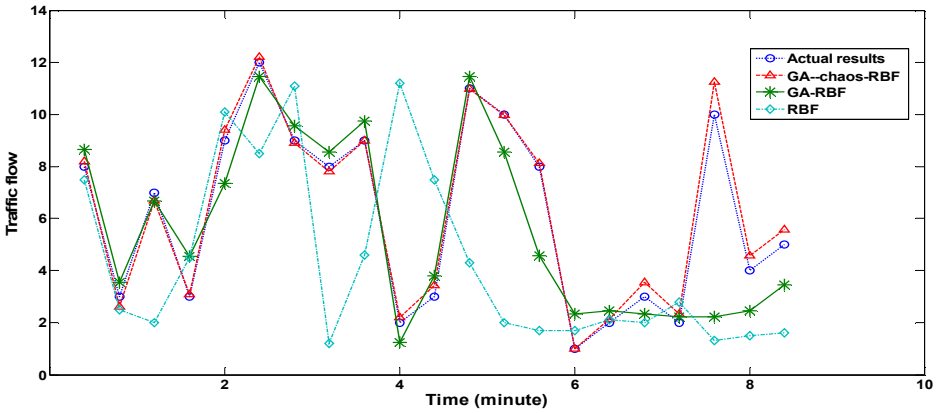

Fig. 6: Prediction results of middle traffic flow

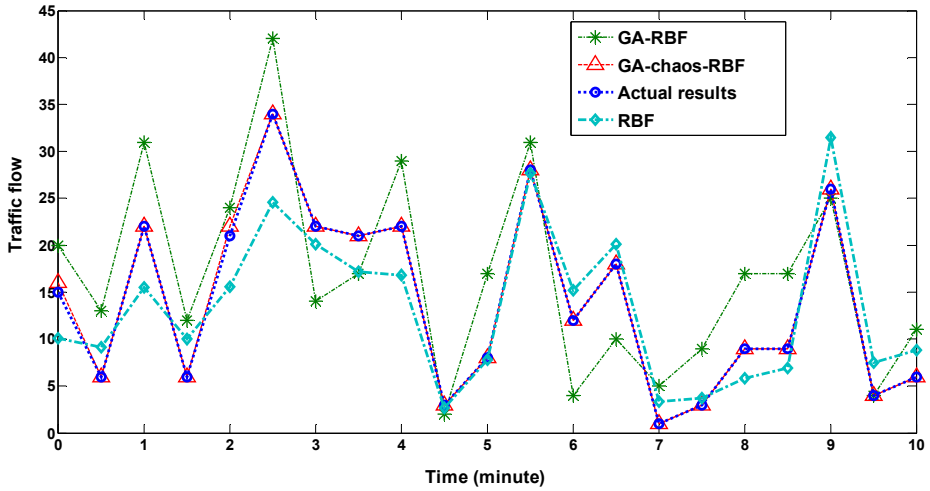

Fig. 7: Prediction results of large traffic flow

90.5%, while the GA-chaos-RBF is 95%. Hence, the new GA-chaos-RBF model is the best one among these approaches in the traffic flow forecasting. With the proposed SOFM-GA-chaos-RBF data mining model, accurate traffic flow can be forecasted and optimized traffic management decision can be provided.

CONCLUSION

Intelligent Transportation Systems (ITS) processes a large amount of traffic information every hour. It is necessary to employ advanced data mining approaches to excavate the hidden knowledge in the ITS database. This study presents a new hybrid intelligent data mining model for the traffic information forecasting. This new method combines the advantages of the unsupervised learning of SOFM and supervised learning of RBF network to mine distinct and potential patterns of the traffic data. Moreover, the GA-chaos algorithm is adopted to optimize the RBF parameters. The experimental test results show that the presented data mining approach is feasible and efficient for potential knowledge extraction of ITS data. The prediction rate of the proposed SOFM-GA-chaos-RBF model is 95% and much better than the model without optimization algorithm. The proposed forecasting system in this work may provide practical utilities for ITS data mining. Further research can extend the proposed method to other complex information mining system.

ACKNOWLEDGMENT

This study is sponsored by the National Natural Science Foundation of China (No. 51208394) and National Science Foundation of Hubei Province of China (No. 2012FFA099).

REFERENCES

Hauser, T. and W. Scherer, 2001. Data mining tools for real time traffic signal decision support and maintenance. Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, 3: 1471-1477.

Jia, L., L. Yang, Q. Kong and S. Lin, 2006. Study of artificial immune clustering algorithm and its applications to urban traffic control. Int. J. Inform. Technol., 12: 1-9.

Jiang, Y., Z. Li and Y. Geng, 2010. Research on AR modeling method with SOFM-based classifier applied to gear multi-faults diagnosis. Proceeding of International Asia Conference on Informatics in Control, Automation and Robotics, 2: 488-491.

Kohonen, T., 1990. Derivation of a class of training algorithms. IEEE T. Neural Networks, 1: 229-232.

Krishna, B., 2012. Binary phase coded sequence generation using fractional order logistic equation. Circ. Syst. Signal Process, 31(1): 401-411.

Li, Z., X. Yan, C. Yuan, J. Zhao and Z. Peng, 2010. The fault diagnosis approach for gears using multidimensional features and intelligent classifier. Imeche. Sem. Worldwide, 41: 76-86.

Li, Z., X. Yan, C. Yuan, J. Zhao and Z. Peng, 2011a. Fault detection and diagnosis of the gearbox in marine propulsion system based on bispectrum analysis and artificial neural networks. J. Mar. Sci. Appl., 10: 17-24.

Li, Z., X. Yan, C. Yuan, Z. Peng and L. Li, 2011b. Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mech. Syst. Signal Pr., 25: 2589-2607.

Li, Z., X. Yan, Y. Jiang, L. Qin and J. Wu, 2012a. A new data mining approach for gear crack level identification based on manifold learning. Mechanika, 18: 29-34.

Li, Z., X. Yan, Z. Guo, P. Liu, C. Yuan and Z. Peng, 2012b. A new intelligent fusion method of multi-dimensional sensors and its application to tribo-system fault diagnosis of marine diesel engines. Tribol. Lett., 47: 1-15.

Li, Z., X. Yan, C. Yuan and Z. Peng, 2012c. Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. J. Mech. Sci. Technol., 26(8): 2413-2423.

Nejad, S., F. Seifi, H. Ahmadi and N. Seifi, 2009. Applying data mining in prediction and classification of urban traffic. Proceeding of the WRI World Congress on Computer Science and Information Engineering, 3: 674-678.

Park, B., D. Lee and H. Yun, 2003. Enhancement of time of day based traffic signal control. Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, 4: 3619-3624.

Raahemi, B., A. Kouznetsov, A. Hayajneh and P. Rabinovitch, 2008. Classification of peer-to-peer traffic using incremental neural networks (fuzzy ARTMAP). Proceeding of IEEE Canadian Conference on Electrical and Computer Engineering, pp: 719-724.

Wen, Y. and T. Lee, 2005. Fuzzy data mining and grey recurrent neural network forecasting for traffic information systems. Proceeding of the IEEE International Conference on Information Reuse and Integration, pp: 356-361.

Xu, P. and S. Lin, 2009. Internet traffic classification using C4.5 decision tree. J. Softw., 20(10): 2692-2704.

Zahra, Z., P. Mahmoud and S. Hossein, 2010. Application of data mining in traffic management: Case of city of Isfahan. Proceeding of the International Conference on Electronic Computer Technology, pp: 102-106.