ITS를 위한 데이터 마이닝과 인공지능 기법 연구

Kaung Myat Sam, 이경현
부경대학교 IT융합응용공학과
e-mail : kmsam254@gmail.com, khrhee@pknu.ac.kr

Data Mining and Artificial Intelligence Approach for Intelligent Transportation System

Kaung Myat Sam, Kyung-Hyune Rhee
Dept. of IT Convergence and Application Engineering,
Pukyong National University.

Abstract

The speed of processes and the extremely large amount of data to be used in Intelligence Transportations System (ITS) cannot be handling by humans without considerable automation. However, it is difficult to develop software with conventional fixed algorithms (hard-wired logic on decision making level) for effectively manipulate dynamically evolving real time transportation environment. This situation can be resolved by applying methods of artificial intelligence and data mining that provide flexibility and learning capability. This paper presents a brief introduction of data mining and artificial intelligence (AI) applications in Intelligence Transportation System (ITS), analyzing the prospects of enhancing the capabilities by means of knowledge discovery and accumulating intelligence to support in decision making.

I. Introduction

Information technology (IT) has transformed many industries, from education to health care to government, and is now in the early stages of transforming transportation systems. While many think improving a country's transportation system solely means building new roads or repairing aging infrastructures, the future of transportation lies not only in concrete and steel, but also increasingly in using Intelligence Transportation System (ITS) [1]. The main contributing factor influencing the development of intelligent systems in transportation has been the demand for a more powerful yet flexible and robust technique. This is in order to cope with the extensive amount of data and knowledge stored in traffic, geological and financial databases, and more importantly, the complexity of interpretation. In most expert systems, information and knowledge are represented as a set of rules. These rules are usually easily understood by users; however generating and maintaining a large rule-based, knowledge-acquisition tool can be labor intensive [2]. This paper advances the understanding of the application of Artificial Intelligence (AI) and Data Mining tools to transportation related data by demonstrating the potential of these techniques in complex real time situations. For many classification and prediction modelling tasks obtaining large amounts of training data in order to build accurate models is often difficult due to technical and economic constraints. In some cases, it may be possible to acquire a large amount of data, but high quality training data is often quite limited. Hence, it is important to build accurate classifiers using advance machine learning, Artificial Intelligence (AI) and data mining techniques. In some cases it may be required to study the data in order to be able to identify suspect data and generate a subset of the data that has higher information content than the original data. Data mining, in other word deals with the analysis of large and complex databases in order to discover new, useful and interesting knowledge using techniques from machine learning and statistics.

II. Urgent requirement of integration in ITS

One of the key problems of ITS is the analysis and management of traffic information collected by enormous sensor. Some superficial information is obtained easily though traditional query operation from traffic data, but deep level information that hides in the traffic data is difficult to be discovered. It usually contains overall characteristics of data and forecast information of data development tendency. Integrating ITS improve the performance of a country's transportation system by maximizing the capacity of existing infrastructure, reducing to some degree the need to build additional highway capacity, increasing driver and pedestrian safety, enhancing mobility and convenience. ITS can be the centerpiece of efforts to reform surface transportation funding systems to hold transportation service providers more accountable for providing real results improving the operational performance of the transportation network. The following table shows today category of ITS applications.

<Table 1> Classifying ITS Applications Category

ITS CATEGORY	SPECIFIC ITS APPLICATIONS
Advanced Traveler Information Systems (ATIS)	Real-time Traffic Information Provision
	Route Guidance/Navigation Systems
	Parking Information
	Roadside Weather Information Systems
Advanced Transportation Management System	Traffic Operations Centers (TOCs)
	Adaptive Traffic Signal Control
	Dynamic Message Signs (or "Variable" Message Signs)
	Ramp Metering
ITS-Enabled Transportation Pricing Systems	Electronic Toll Collection (ETC)
	Congestion Pricing/Electronic Road Pricing (ERP)
	Fee-Based Express (HOT) Lanes
	Vehicle-Miles Traveled (VMT) Usage Fees
	Variable Parking Fees
Advanced Public Transportation Systems (APTS)	Real-time Status Information for Public Transit System (e.g. Bus, Subway, Rail)
	Automatic Vehicle Location (AVL)
	Electronic Fare Payment (for example, Smart Cards)
Vehicle-to-Infrastructure Integration (VII) and Vehicle-to-Vehicle Integration (V2V)	Cooperative Intersection Collision Avoidance System (CICAS)
	Intelligent Speed Adaptation (ISA)

Above ITS applications process a huge amount of traffic information every hour and its collected data seem to be generally unfeasible due to extremely large, noisy, and disorganized: unusual values or incomplete data with missing values. [3]

The following figure (1) shows a typical control and management framework of ITS. It includes ITS data source module which is collected by any median, data warehouse module which is stored in any server, data mining module and Decision Support System module. Whether the functions of these components can fully be realized depends on how data are collected and processed into useful information additional implementation should require as follows. Data mining is one of its key techniques in this traffic information system. It is the basic of the Decision Support System module (AI system), which is respond to correct traffic information forecasting and traffic control. Firstly, data mining can find a useful pattern of knowledge from a large data set and AI system can support for better evaluation and finding result. Hence, It is necessary to employ advanced approaches to excavate the hidden knowledge in the ITS database. For these reasons, data mining and AI have being suitable for applying in ITS.

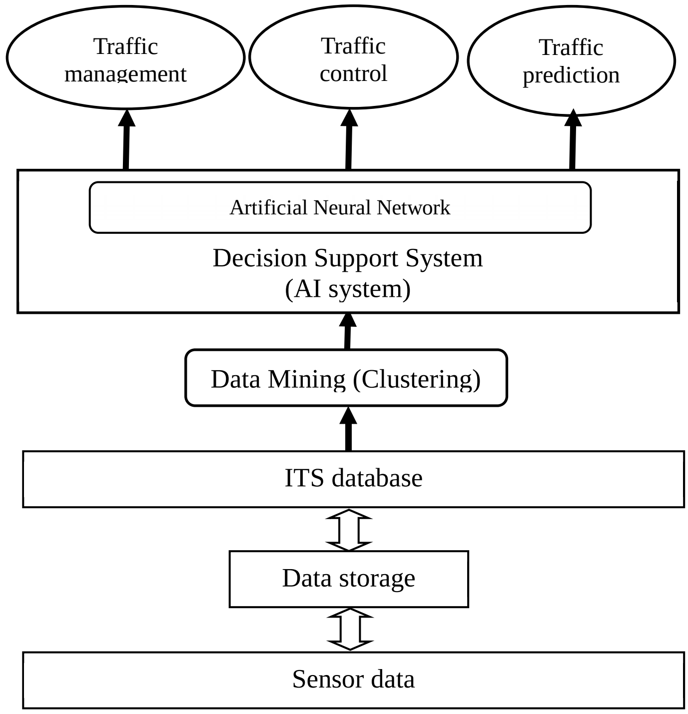

Fig. 1: Typical control and management framework of ITS

III. Data mining researches and applications in ITS

ITS have been implemented widely throughout the world, giving the access to large amounts of data. From the review of traffic-flow forecasting models done by Han and Song (2003) [4], data mining techniques represent one of the basic models currently adopted in this field. In particular Artificial Neural Network have been commonly used for the problem since 1990's, given the fact the appropriate neural network structure can approximate a given real-valued, continuous multi-variant function to any desired degree of accuracy (Hornik 1989)[5]. Gong and Liu (2003) [6] proposed an algorithm based on association rules mining and association analysis. Wang et al. (2006) [7] presented a dynamic traffic prediction model. The model deals with traffic flow data to convert them into traffic status. Cheng, Cui, and Cheng (2003) [8] employed a combination of neural networks and statistical analysis of historical data to forecast the inter-regional air traffic flow in China. Hayashi et al. (2005) [9] presented a detection method of driver's drowsiness with focus on analyzing individual differences in biological signals (pulse wave) and performance data (steering data). Based on these researches concern with data manipulating, decision-makers arrive at decision to solve respective problem. In real life the situation of transportation domain, diverse fields of data need to be collected to integrate and arrive at the solution. [10][11][12][13]

To make the decision about traffic management, control and prediction, one should know traffic volume for the entire year. Hence, the data mining and AI approach provide accurate estimation without much incurs cost and man power. ITS based devices collected traffic data for daily, monthly and yearly basis. Thus, the required data could be available through ITS database. We use one of the data mining techniques: Model-based clustering methods to find the basic traffic pattern. This method showed slightly better performances compared to the other methods and had a robust mathematical structure. To calculate, we should define some basic word; Annual Average Daily Traffic (AADT) is the traffic count of average day in a year recorded by ITS devices for a given place section. Permanent Traffic Counts (PTCs) are counts the traffic in longer term of time-of-day, day-of-week, and seasonal travel pattern precisely, while Short Period Traffic Counts (SPTCs) are taken for short basic of 24 hours to some weeks). Since road sites k for the *i*-th day of the week of the *j*-th month is

$$f_{ij}^{k} = \frac{AADT_k}{ADT_{ijk}} \tag{1}$$

Where $AADT_k$ is the AADT for the *k*-th road site, ADT_{ijk} is the average daily traffic recorded in the *i*-th day of week of the month in the *k*-th road site. Above equation is for average traffic flow of one site. To get the combination (*c*) of number of road sites (*n*) can be calculated by:

$$f_{ijc} = \frac{1}{n}\sum_{k=1}^{n}\frac{AADT_k}{ADT_{ijk}} = \frac{1}{n}\sum_{k=1}^{n} f_{ij}^{k} \tag{2}$$

Above equation is for average traffic flow for multiple combination of the average daily traffic recorded in *i*-th day of week of the *j*-month in the I site. From this value, we can estimated the AADT by multiplying daily traffic count DT_{ij}, obtained for the *i*-th day of the week of the *j*-th month, by the resulted f_{ijc}:

$$AADT_{Estimate} = DT_{ij} \,.\, f_{ijc} \tag{3}$$

DT is the 24 hour volume obtained from SPTC; if SPTC is for more than 24 hours, then DT is the average 24 hour volumes for the duration of SPTC. Finally, we can calculate the exact value of the $AADT_k$, which require for above two equations.

$$AADT_k = \frac{1}{7}\sum_{i=1}^{n}\left[\frac{1}{12}\sum_{j=1}^{12}\left(\frac{1}{n}\sum_{l=1}^{12}\left(DT_{ijk}\right)\right)\right] \tag{4}$$

where DT_{ijk} is the daily traffic count taken from *i*-th day of a week, *j*-th month in the *k* road site. The index *l* represents the occurrence of given *i*-th day of a week, *j*-th month, and *n* is the number of times with same occurrence (usually between 1 to 5, depending on the calendar and the number of missing day). To be better result, we should use For each AADT estimate $AADT_{i,Estimate}$ the absolute percent error was calculated by the Mean of the Absolute percent error (MAE), as the expected daily traffic volume on a "typical day":

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{AADT_{i,Estimate} - AADT_{i,Actual}}{AADT_{i,Actual}}\right| \times 100\right) \tag{5}$$

Error patterns showed worse results in AADT estimation in some specific cases: such as road group, week day and week end, period of year and vehicle type, Hence, Artificial Intelligence (AI) usages are not avoidable in ITS in order to enhance the data mining technique.

IV. AI applications and techniques in ITS

AI role in ITS is the decision support system to evaluate the pattern and outcomes from data mining. It attempts to understand the substance of intelligence, and produce a new intelligent machine could make reactions similar to the human intelligence, Artificial Intelligence [14][15] in the network information retrieval applications, mainly in ITS. As the network knowledge and information include both the regularity knowledge, such as the concept of general principles, and a large number of empirical knowledge. Such empirical knowledge will inevitably carry with fuzziness, randomness, non-reliability and other uncertain factor, reasoning it needed to use artificial intelligence research.

Artificial Neural Networks (ANNs) are an effective technique to assign an input to an output when the causal relation is not well-understood. First of all, we have to find the input layer of groups which are often fuzzy.

The input layer includes a node of SPTC (short period traffic count) containing the period of year through hour's factor (h_1), is defined as

$$h_1 = \frac{HT_l}{DT} \tag{6}$$

Where l = 0, ..., 23 hour of a day, HT_l is the hourly traffic volume for hour *l* and *DT* is the daily traffic.

The output layer composed of road group and nodes for all sets, such as Groups (1 or 2), Groups (2 or 3 or 5)

The hidden layer has SPTCs considered nodes of variable number.

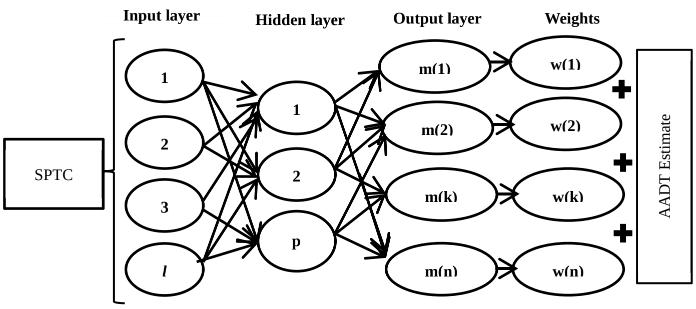

Fig. 2: Sample ANN for given road group

Firstly, we should take data from the group of road sites using previous data mining technique in order to find the basis of seasonal adjustment factors of individual road sites. Secondly, assign the SPTC to one or more predefined road groups as input layer. ANN calculates the measures of uncertainty associated with specific road groups. The last step is to calculate AADT for each given road section when SPTC are available for respective group and then combine all resulted group to calculate the final AADT estimation.

$$AADT_1 = w(1).SPTC.f_{ij1} + w(2).SPTC.f_{ij2} \tag{7}$$

where f_{ij1} and f_{ij2} were calculated from equation 2 and the final AADT estimate is:

$$AADT_{Final} = \frac{AADT_1 + AADT_2 + AADT_3}{3} \tag{8}$$

where each AADT is the AADT estimated using the 24hr volume data for each of the d monitoring days. Consequently,

If we get the most realistic traffic data about the given location of daily basic to annual, we will easily make managing, prediction and control of ITS data. The following Fig.3 illustrated the mean absolute error for given data set of 24hr and 72hr. It can be clearly see that increase in the duration of the SPTCs results in a decrease in the errors in AADT estimates, particularly when the counting duration increases from 48 to 72 hours.

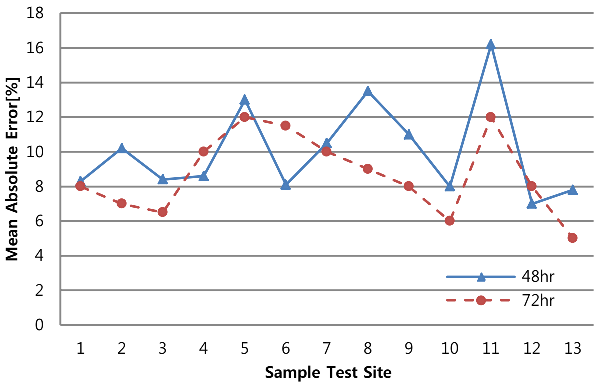

Fig. 3: Mean Absolute Error [%] form Gecchele et al

Some criticism of AI is that the human analyst needs to make some very important design decisions, including the topology of the network to use, the number of hidden layers, the number of neurons to use in the hidden layer, the type of trigger functions to use in the neurons. And another limitation or a challenge related to the issue of potential liability.

V. Conclusion

In this paper, we introduce the framework of how to find basic traffic knowledge through ITS, data mining and AI. We can use the large data from ITS database which is collected by sensors of ITS devices for finding the basic knowledge about the traffic. Our data mining tools firstly find the daily, monthly and yearly traffic pattern yet this have some limitation in some specific location hence fueled by ANN to improve accuracy and better estimation. AI based data mining technique have been successfully used in a number of transportation environments. The economic and social and environmental will benefit from accurately predicted model and outcome. We have learned that the use of data mining and AI for traffic related applications can be very effective and successful However, it is necessary to develop standards, which can be followed when developing such computerized programs to ensure that there is some consistency within informatics programs. Currently, many real time traffic based cases are being carried out to ensure that the implementation of such techniques improves and benefits transportation sector and bring together traffic related fields and Information Technology and make them a better success. In the longer-term, this integration will hope to provide decision making rules based on traffic data, which can subsequently be incorporated into computer driven “intelligent systems”. Finally, utilizing these applications and methodologies of data mining and artificial intelligence approach to intelligence transportation system can uncover unique patterns in diagnostic datasets that will enhance the predictive power of various planning for decisive optimal control policies. In contrast, this intended not to develop a general-purpose problem solver, but rather to address real transportation problems that have defied solution using classical solution methods.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2013R1A1A4A01009848)

References

[1] Stephen Ezell “Explaining International IT Application Leadership: Intelligent Transportation Systems” January 2010, pp. 3-10.

[2] Fariba Shadabi and Dharmendra Sharma “Artificial Intelligence and Data Mining Techniques in Medicine – Success Stories” 2008 International Conference on Biomedical Engineering and Informatics. pp 235 – 239.

[3] Wei He, Tao Lu and Enjun Wang “A New Method for Traffic Forecasting Based on the Data Mining Technology with Artificial Intelligent Algorithms” Research Journal of Applied Sciences, Engineering and Technology 5(12), 2013. pp. 3417-3422.

[4] Han, C. and S. Song (2003). “A Review of Some Main Models for Traffic Flow Forecasting”. In: 2003 IEEE International Conference on Intelligent Transportation Systems. Ed. by IEEE. Shanghai, China, pp. 216–219.

[5] Hornik, K. (1989). “Multilayer Feed forward Networks Are Universal Approximators”. In: Neural Networks 2, pp. 359–366.

[6] Gong, X. and X. Liu (2003). “A Data Mining Based Algorithm for Traffic Network Flow Forecasting”. In: KIMAS 2003. Ed. by IEEE. Boston, MA, USA, pp. 243–248.

[7] Wang, Y. et al. (2006). “Dynamic Traffic Prediction Based on Traffic Flow Mining”. In: 6thWorld Congress on Intelligent Control and Automation. Ed.by IEEE. Dalian, China, pp. 6078–6081.

[8] Cheng, T., D. Cui, and P. Cheng (2003). “Data Mining for Air Traffic Flow Forecasting: A Hybrid Model of Neural-Network and Statistical Analysis”. In: 2003 IEEE International Conference on Intelligent Transportation Systems. Ed. by IEEE. Shanghai, China, pp. 211–215.

[9] Katsuki HAYASHI, “Individualized Drowsiness Detection during Driving by Pulse Wave Analysis with Neural Network”, Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems Vienna, Austria, September 13-16, 2005.pp 901-906.

[10] Sudhir Kumar Barai “Data mining application in transportation engineering”. Transport- vol xviii, no 5- 6. pp. 216–223.

[11] Gregorio Gecchele and Iginio Marson “Transportation data analysis; advances in data mining and uncertainty treatment”. Anno Accademico 2011/2012. pp. 67-72.

[12] Fariba Shadabi and Dharmendra Sharma, "Application Analysis of Data Mining in Intelligent Transport System," Computer engineering, vol.31, 2005, pp.94-96

[13] Ranjan,J(2009): Data Mining in Pharma Sector : Benefit. Ghaziabad, India: Emerald. Vol 11, pp. 82-93.

[14] Song Ning & Ma Yan. “Discussion on Research and Development of Artificial Intelligence”, 2010 IEEE. pp 110 - 112

[15] Transportation Research Board of The National Academies, “Artificial Intelligence Applications to Critical Transportation Issues”, Transportation research circular, November 2012. pp. 17-25.