Expert Systems with Applications 41 (2014) 6524–6535

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Traffic conduction analysis model with time series rule mining ☆

Huiyu Zhou [a,*], Kotaro Hirasawa [b]

[a] SEM, Beijing Jiaotong University, Shangyuancun 3#, Haidian Ward, Beijing 100044, China
[b] IPS, Waseda University, 808-0135, Japan

ARTICLE INFO

Keywords:
Traffic conduction
Time related data mining
Time series rule mining
Traffic prediction

ABSTRACT

The traffic density situation in a traffic network, especially traffic congestion, exhibits characteristics similar to thermodynamic heat conduction, e.g., the traffic congestion in one section can be conducted to other adjacent sections of the traffic network sequentially. Analyzing this conduction facilitates the forecasting of future traffic situation; therefore, a navigation system can reduce traffic congestion and improve transportation mobility. This study describes a methodology for traffic conduction analysis modeling based on extracting important time-related conduction rules using a type of evolutionary algorithm named Genetic Network Programming (GNP). The extracted rules construct a useful model for forecasting future traffic situations and analyzing traffic conduction. The proposed methodology was implemented and experimentally evaluated using a large scale real-time traffic simulator, SOUND/4U.

1. Introduction

Traffic congestion in traffic networks exhibit conduction characteristics of radiation phenomena that are similar to the thermodynamic heat conduction effect observed in the natural world. This traffic congestion phenomenon has not been completely and systematically studied in previous studies. For example, suppose that if section A_1 has traffic congestion, and after 10 min, a section not far away from it, section A_2 also has a traffic jam, and, if this event sequence happens frequently enough, for example, over 80%, it becomes a rule. As a result, if section A again has congestion, we can predict that section B will also have traffic congestion 10 min later with a possibility higher than 80%. In such a case, this event sequence has formed a conduction rule. The objective of this study is to identify similar conduction rules in transport networks and apply them to the Intelligent Transport Systems (ITS) (Kaysi, Ben-Akiva, & Koutsopoulos, 1993).

A vast amount of traffic data is currently available. Satellite-based automatic vehicle location technologies such as Global Positioning System (GPS) and cellular phones can determine vehicle positions at small time intervals. These types of equipment collect information about the vehicle positions and speeds, which are archived in a large number of databases, enabling further analysis of the temporal data about traffic situations such as traffic density patterns. When and where traffic congestion or heavy traffic will occur in the traffic network should be predicted, and then, the navigation system can choose the route, avoiding the sections with potential congestion and giving higher priority to the sections with a potential low traffic density level.

The objective of this study is to establish a traffic conduction analysis model such that future traffic situations can be predicted and provided to navigation systems. The goal of analyzing conduction in a time series traffic database is to identify whether and how frequent a periodic pattern is repeated and transmitted within the time series of the traffic network. In other words, this problem is related to temporal and spatial regularity.

Time-related pattern mining is one of most interesting areas in data mining, which includes association rule mining (Agrawal & Srikant, 1994; Lin, Alvarez, & Ruiz, 2002; Zhou, Mabu, Wang, & Hirasawa, 2011), sequential pattern (Agrawal & Srikant, 1995; Merah, Samarah, Boukerche, & Mammeri, 2013; Srikant & Agrawal, 1996; Zaki, 2001, inter transaction pattern (Lu, Han, & Feng, 1998; Tung, Lu, Han, & Feng, 1999; Feng, Li, & Wong, 2001), periodic pattern mining (Cao, Mamoulis, & Cheung, 2007; Nishi, Ahmed, Samiullah, & Jeong, 2013) and episode mining (Laxman, Sastry, & Unnikrishnan, 2007; Tzima, Mitkas, Voukantsis, & Karatzas, 2013). However, most of the existing algorithms have a major limitation in mining interesting sequential patterns of user interest. That is, they only consider mining the patterns of specific length with all of the events sequentially one after another without concrete time period information, which is insufficient in traffic prediction systems. For example, the mining of these patterns only provides the order for most of the events (congestion, etc.) in traffic

☆ The Project is Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

* Corresponding author. Tel.: +86 01051688444.
E-mail address: hyzhou@bjtu.edu.cn (H. Zhou).

http://dx.doi.org/10.1016/j.eswa.2014.03.009
0957-4174/

networks. We cannot determine how long it will be until another traffic congestion event will occur after one existing traffic congestion event. Similarly, we cannot determine when the navigation system should choose another route to avoid predicted congestion.

Therefore, this study proposes a time series rule mining method to extract rules and represents them as a special type of association rule–conduction rules. The extracted rules will be analyzed to establish a traffic conduction analysis model and applied to traffic prediction systems. For example, a conduction rule may appear as follows: if section A_1 on the traffic map has traffic congestion at time $t = 0$ (current time), then section A_2 will also have traffic congestion at $t = 10$ (10 time units later).

1.1. Background and related works in traffic prediction

Many traffic prediction methods are based on traffic congestion prediction using regression analysis. Regression models have the advantage of accurate prediction, especially when model assumptions are thoroughly examined. Traditional statistical methods (Smith & Dermetsky, 1997) for traffic prediction include historical average modeling, Kalman filtering modeling, Markov prediction and Maximum likelihood formulation models, among others. Although they can solve the problem of traffic flow changes for different times and durations to a certain extent, they cannot solve the problem for unconventional and unexpected traffic conditions, such as accidents.

Regression analysis is a forecasting method based on obtaining the traffic information of multiple sections, establishing linear regression equations among different section parameters and predicting the future traffic situation, but, when the data are limited, this method cannot work well, and when the traffic network is very complex, the equations for regression analysis become complicated (Dia, 2001).

Time series models (Voort, Dougherty, & Watson, 1996) are a common way to describe the statistical characteristics of time series and thus can also be used in traffic prediction because they use parameterized models to process dynamic random data in practical ways. Their models enjoy high forecast accuracy but require complex parameter estimation, and the calculated parameters cannot be transplanted.

Kalman filter theory (Okutani & Stephamedes, 1984) adopts state and observation equations to create a linear stochastic system state-space model to describe the filter. According to the linear unbiased minimum mean square error estimation criteria, the method uses a recursive algorithm for the state variable filter to achieve the best estimate, which seeks to filter out noises. However, this method is a linear model; therefore, when predicting a non-linear situation or traffic flow with uncertainty, the model performance decreases. In each calculation, the weights have to be re-adjusted. Therefore, when the calculation is huge, the predicted output values can sometimes be delayed for several time units.

In general, prediction methods based on traditional statistical theories are simple and easy to understand, but as most of the models are based on linear assumptions, when the prediction interval is less than 5 min, because of the strengthening of the randomness and nonlinearity of traffic changes, the model's performance deteriorates. In addition, the prediction of the section uses only the historical data, without considering the impact of the adjacent sections, which affects the accuracy of the prediction.

Soft computing-based methods such as neural networks, can also be used for traffic congestion prediction (Lingras & Osborne P., 1997), where the self-evolved parameters are used and also robustness under noisy environments is obtained. In addition, they have good adaptability because the parameters can be adjusted automatically as the OD (Origin/Destination) changes in the road networks.

Dynamic Traffic Assignment (DTA) has also attracted much attention, as it is capable of processing the time-varying properties of traffic flows in guidance generation. Real-time OD estimation models (Bottom, Ben-akiva, Bierlaire, Koutsopoulos, & Yang, 1999) update the actual road network demand, capturing drivers' traveling patterns. A behavior-consistent approach has also been studied by Paz and Peeta (2007), who developed models that simultaneously consider network flow interactions and behavior realism for meaningful information based network control strategies. Generally speaking, DTA models require the time-dependent OD (origin and destination) data in the prediction process.

In addition, some methods have been already proposed to study the effect of traveler's information on reducing traveling times and improving information reliability (Balakrishna, Koutsopoulos, Ben-Akiva, Fernandez-Ruiz, & Mehta, 2005). Studies have verified that driver overreaction from the dissemination of information can be eliminated through prediction-based route guidance that uses forecasts of network states. These methods focus on detailed simulation-based study for the objective and independent evaluation of advanced traveler information systems and analysis of the impacts of various design parameters (e.g., logit values) and modeling errors on the quality of the generated guidance. The proposed method is focused more on the analysis of the detailed data mining method for rule extraction, and the parameter settings will be briefly discussed considering the methodology of the proposed model.

First, in real-world applications, it is difficult for the traffic guidance systems to obtain historical OD demand data. Thus, the proposed method is not based on the traffic OD information. Instead, it establishes the prediction model from a new perspective by adopting a data mining-based approach using traffic density information, which only requires the traffic network time-dependent traffic density information.

Second, the proposed mechanism uses an evolutionary-based method (Holland, 1975) to extract interesting conduction rules, and these rules are stored and used to construct a classifier model that predicts the traffic of the road network. As a result, the proposed method can address the traffic prediction of most of the sections in the road networks under various traffic situations with noisy environments and displays good traffic prediction performances in noisy environments.

Data mining is generally a mechanism involving statistics that extracts important rules from complex databases. In the proposed method, time-related conduction rules for all of the sections on the map can be obtained and used for traffic prediction. Once enough rules are extracted, they can be used on-line for the traffic analysis of the whole traffic networks, whenever and wherever future traffic information is necessary during real-time route guidance.

1.2. Time series data mining

As the data naturally have the form of time sequences in transportation applications, conventional association rules are insufficient to predict the future traffic density, which means that association rules should be time related. Therefore, the proposed method is more interested in how the traffic density situation (for example, congestion) is conducted in sequences among traffic networks.

The problem of discovering conduction rules from historical object movements is very challenging. Typically, the patterns are not explicitly specified and have to be discovered from the data. The patterns can be thought of as (possibly noncontiguous) sequences of events that reappear in history data series. In addition, given that we expect to extract conduction rules from adjacent locations at every instant of each period, the conduction rule mining is a spatial–temporal process, which increases the complexity of the mining tasks.