

Link Speed Estimation and Incident Detection Using Clustering and Neuro-fuzzy Methods

Seung-Heon Lee, M. Viswanathan, and Young-Kyu Yang

College of Software, Kyungwon University, Bokjeong-Dong, Sujung-Gu,
Seongnam-Si, Gyeonggi-do, South Korea 461-701
shleejj@gmail.com, {murl, ykyang}@kyungwon.ac.kr

Abstract. The primary issues in the development of advanced traveler information systems (ATIS) within the intelligent transportation systems (ITS) framework are the optimal estimation of freeway travel time and incident detection with reasonable accuracy. Typically ATIS aims to provide route guidance based on the traveler's requirements using the information gathered from various sources such as loop detectors and probe vehicles. Until recent times traffic information was collected from mostly stationary devices and analyzed. In this research paper we consider data acquired from primarily GPS-based sources. The aim of the research is a comprehensive analysis of collected information from GPS sources using the fuzzy c-means algorithm (FCM) which provides the estimation of link speed. The modified FCM is used to extract patterns from the traffic data collected from a busy network of downtown streets. The link speed estimation is performed using smoothing techniques. Finally we apply the neuro-fuzzy algorithm to the task of incident detection from the traffic patterns.

1 Introduction

In recent years the growth in automobiles sales and the related economic and environmental factors have increased the pressure on infrastructure development. However due to the limits on infrastructure growth efficient and intelligent utilization of the existing transportation infrastructure has become most significant. Intelligent transportation systems are therefore needed to provide a high level of automation in advanced traffic management systems (ATMS) and advanced traveler information systems (ATIS). The accurate estimation of the traffic information is of utmost importance for the drivers as well as the traffic information service provider.

Until recent years, traffic data was collected by the fixed detectors such as loop detector, CCTV and beacon, etc. Many researchers have endeavored to develop reliable travel time forecasting and incident detection models using various methods including historical profile approaches, time series models, neural networks, nonparametric regression models, traffic simulation models, and dynamic traffic assignment (DTA) models [2]. The methods of incident algorithms can be grouped as algorithms based on pattern recognition and statistical methods, algorithms using traffic models and theory, and other hybrid state of the art incident detection methods based on neural networks and fuzzy methods. Due to the recent trend in the wide-spread use of GPS

receivers in car navigation systems, traffic information can be easily collected by the probe cars equipped with GPS receivers. Many traffic information providers are offering link speeds estimated solely by GPS data due to the low cost and data availability in real-time. This thus requires developing novel algorithms for extracting traffic patterns, different from the algorithms used for the conventional fixed detectors.

The aim of research is to propose and test a new algorithm for calculating the travel speed and incident detection for the collected information from the probe cars equipped with GPS receivers. For this aim, we automatically cluster the GPS data into five levels of speed layers which are very slow, slow, middle, high and very high speed using Fuzzy C-Means (FCM) and apply these layers of speed patterns to predict the link travel speed and incident detection.

2 Existing Approaches in Link Speed and Incident Detection

Typically many conventional methods use statistical and simple neural network algorithms to calculate the link travel speed. The data set collected by the fixed detection devices are speed, road occupancy rate, and the amount of traffic flow and available on for the fixed locations where the devices are installed. But, acquired data from GPS probe cars are only speeds at specified time and location. And therefore traffic information from GPS devices requires novel methods and algorithms for analysis.

Some examples of traffic information modeling include the ADVANCE project [9] which estimated the link travel time for expressway and arterial roads using traffic volume and occupancy ratio from detector data. Sisiopiku et al.[10] present a study of correlation between travel time and traffic volume and occupancy rate for estimation of link travel time. They use the NETSIM model for link travel time estimation and analyzed the relation between traffic volume and travel time as well as occupancy and travel time using Fuzzy and estimated link travel time. Nelson et al. [11] used traffic volume and occupancy rate as input data and estimated travel time using fuzzy-neural system studied integration of neural network with fuzzy rule with empirical results that suggested that the integration improved the performance. Some well-known pattern recognition algorithms include the California algorithm [5][7][8] that detects incidents using occupancy rate at specified times, APID algorithm [6][7][8] that compares occupancy between two detector areas, and wave analysis, PATREG, and the Monica algorithm. Statistical algorithms include SND and Bayesian algorithms [7] that use probability of incident occurrence through blocked crossroads of downstream traffic. Algorithms using traffic model and theory are Dynamic Algorithms that assumed sudden change in the freeway system to follow predictable patterns due to the integrated relation between speed-density and traffic-density in detection analysis. Additionally based on Catastrophe Theory, the (modified) McMaster Algorithm uses traffic theory as explanation of "Jump" of speed in the operation of freeway that has little change of occupancy. The final group of hybrid algorithms includes Fuzzy Set algorithms, Neural Network Algorithms, Probabilistic Neural Network and Automatic Vehicle Identification Systems [5][7]. The algorithms as mentioned above suffer from issues such as excessive training time for data analysis and the intervention of professionals for generating useful patterns. This research primarily attempts to adapt Fuzzy C-Means algorithm for solving these issues.

3 Design of Link Speed and Incident Detection Estimation System

The primary motivation is to generate patterns for calculating the link speed using an adapted FCM algorithm. The data collection interval is of 5 minute duration and the probe cars are GPS-equipped taxis. The data thus acquired includes patterns such as stopping and running, slow and fast running and searching-running. Due to the various traveling patterns, calculation using time mean speed has many gaps in comparison with actual speed. Also, it is difficult to detect incidents since the taxi probe cars have frequent lower speeds. Therefore, this research proposes a new speed calculation method for optimal estimation of the travel speed and robust incident detection. Figure 1 shows structure of system for link travel speed estimation and incident detection.

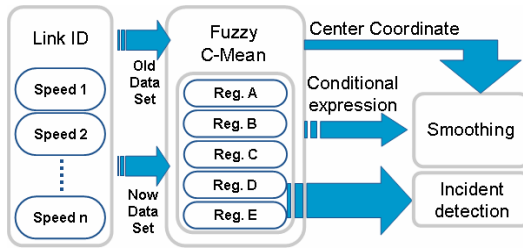


Fig. 1. Structure of system for calculating link speed and incident detection

Since urban traffic patterns have high variation extensive calculations are required and it is challenging to integrate all links. Therefore, in this research, a method using FCM is applied to analyze traffic patterns of each link and reduce calculation volume when generating patterns through training. For calculation of vehicle speed V using GPS we divide length of link by time using pass time of previous node and current node.

$$V = \text{Length of Link} / (T_{nL} - T_{pL}) \tag{1}$$

Using calculated link speed, we classify data into same time of interval of 5 minutes. Classified data is calculated three pattern (high, middle, low speed) using FCM. FCM recognizes spherical clouds of points in p-dimensional space. Having a finite set of objects $X=\{x_1, \dots, x_n\}$ and the number of cluster centers c to be calculated, the assignment of the n objects to the c clusters is represented by the proximity matrix $U=[U_{ik}]$. With $k=1, \dots, n$ and $i=1, \dots, c$, $u_{ik} \in [0,1]$ expressing the fuzzy proximity or affiliation of object x_k to cluster center v_i . Matrix U is used to calculate region of speed and v_i is used to calculate link speed. After generating the cluster we obtain the values of the cluster center and min-max value of cluster. Min-max values are used to decide boundary of speed layer from measured real-time GPS data for calculation of link speed and center values are used in conversion from low-speed to high speed.

Using matrix U , we make conditional expression for pattern of speed. Each cluster has maximum value, minimum value and center value for the speed. These values are used for calculating the link speed using the real time collection speed. V_T is time mean speed and can be calculated by (2). V_T is difficult to use for the link speed

production which uses GPS because volume of collected data using GPS is small until recently. Also, this equation has tendency to ignore high speed data and calculates to speed of middle inside putting first. Therefore, we propose a new speed production method for GPS data.

$$V_T = \frac{1}{N} \sum_{i=1}^N V_i \tag{2}$$

$$V_T = \frac{1}{N} \left(\sum_{j=0}^n V_{i1} + \sum_{k=0}^m (\alpha V_{i2} + (1 - \alpha)v_i) \right) \tag{3}$$

As v_i is center value of cluster which is included maximum speed, V_{i1} is all of speed which is included cluster. And V_{i2} is mean low speed that is little higher than center value of cluster. α are threshold values of speed for making near in high-speed layer. We calculate near maximum speed according to allow many weights to high speed more than low speed. In the above case, maximum speed include in one of pattern. But, FCM is able to get speed which is not include in speed domain U because it is not reflect on region which is not included speed data. Therefore, we calculate new standard speed non cluster v_i using (4). It is calculated by providing weight on center of upper and lower cluster of collected maximum speed. w is weight for connection between upper cluster and lower cluster and has a region from 0 to 1. Weight is calculated using (5).

$$v_i^{non\ cluster} = w \bullet v_i^{upper\ cluster} + (1 - w) \bullet v_i^{lower\ cluster} \tag{4}$$

$$w = N_{lower} / (N_{upper} + N_{lower}) \tag{5}$$

N which is used to calculate weight is number of objects including upper and lower cluster. A cluster with many objects has higher weight in relation to others. Once the weight is computed, the base speed for non-pattern area and link speed can be computed using (3) and (4). The estimated link speed is usually larger than the mean speed because the highest speed data of the link is used for the calculating the link speed. Finally, if estimated speed is higher than the speed limit, it is not usable due to the traffic regulations. Therefore, we will offer users the regulation speed in those cases.

The other aim of this paper is to adapt the ANFIS algorithm to detect incidents in a scenario where the available input data consists of speed measurements based on moving probe vehicles. For incident detection, input values are center values that were generated from our FCM algorithm, collected speed and previous mean speed. Work of first measure that whether or not inclusion about distribution of collected speed in low speed layer. At this time, in case of including all of collected speed in region D-E, our system regard as incident and decide to incident using collected speed, previous mean speed and cluster value of now and previous through ANFIS. The fuzzy membership function that used in ANFIS is constructed generating 36 rules for incident detection. After calculating the speed and pattern, we used this data as inputs and detect incidents through ANFIS. Figure 2 presents some of the fuzzy rules generated.

```

If (AVGSPD is SS) and (PDSPD is SS) and (PPATTERN is VLS) then (RESULT is INCIDENT)
If (AVGSPD is SS) and (PDSPD is MS) and (PPATTERN is VLS) then (RESULT is INCIDENT)
If (AVGSPD is SS) and (PDSPD is LS) and (PPATTERN is VLS) then (RESULT is INCIDENT)
.....
If (AVGSPD is SS) and (PDSPD is LS) and (PPATTERN is VLS) and (NPATTERN is LS) then (RESULT is INCIDENT)
If (AVGSPD is SS) and (PDSPD is MS) and (PPATTERN is VLS) and (NPATTERN is LS) then (RESULT is INCIDENT)
If (AVGSPD is SS) and (PDSPD is SS) and (PPATTERN is VLS) and (NPATTERN is LS) then (RESULT is INCIDENT)
    
```

Fig. 2. Fuzzy rules for incident detection

First, the four input values which include speeds of link and center values of cluster current and previous in each observation are given to the system. For each value the membership degree is calculated by the fuzzy membership function and these values are passed into the neural network according to the fuzzy rules. The neural network multiplies weights to value of all rules and then the resulting values after addition are sent to the output layer. These values undergo de-fuzzification by the fuzzy membership function and the final output value is acquired. The structure of the adapted ANFIS algorithm is shown in figure 3. Once the error rate of the model reaches 0.01 or the learning has gone through 600 epochs, the process is terminated.

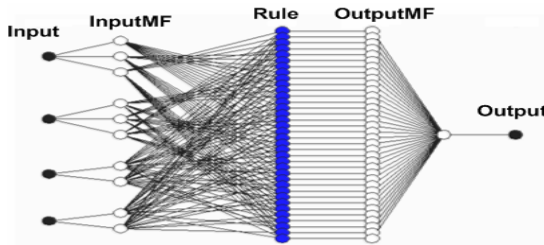


Fig. 3. Anfis model structure

4 Empirical Framework and Analysis

In the creation of speed pattern, we employ taxi-based probe cars for data collection. For link speed analysis and data generation the GPS-based probe cars acquire data for a 1 hour period between 9-10 am. The data thus collected is studied for patterns using FCM. Table 1 shows center value of each cluster after analysis using FCM at 5 minute interval until during the 1 hour period. We generate 5 clusters and get data about center and min-max values of cluster for all days excepting weekend. Table 2 presents results of comparing our algorithm with time mean speed and actual speed. Our algorithm is often closer to the actual speed than the time mean speed. The results from the following table suggest that our algorithm has less variance in approximating the actual speed in comparison to time mean speed. Through performance testing, the average error rate of speed between our proposed algorithm and time mean speed are 4.36km and 9.20km respectively.

Table 1. Center of calculated speed pattern using Fuzzy C-Mean

Speed	Very Low	Low	Middle	High	Very High
09:00-09:05	4	8	10	15	23
09:05-09:10	4	9	13	20	28
09:10-09:15	4	8	11	17	29
09:15-09:20	6	10	15	22	29
09:20-09:25	6	10	16	24	38
09:25-09:30	4	8	12	18	32
09:30-09:35	5	10	15	30	60
09:35-09:40	5	10	13	20	39
09:40-09:45	5	11	19	29	51
09:45-09:50	5	10	17	29	49
09:50-09:55	6	12	20	30	48
09:55-10:00	6	12	23	38	62

Table 2. Comparison of calculated speed using mean, test probe car and our algorithm

5 interval	Link 1			Link 2		
	mean	actual	new	mean	actual	new
09:00-09:05	47	43	54	15	31	20
09:05-09:10	36	41	43	24	31	31
09:10-09:15	48	45	48	19	29	25
09:15-09:20	24	43	40	20	26	27
09:20-09:25	33	37	41	22	24	29
09:25-09:30	34	38	44	15	27	25
09:30-09:35	42	42	48	10	25	12
09:35-09:40	25	48	42	26	23	27
09:40-09:45	37	50	49	12	24	18
09:45-09:50	36	51	43	14	29	26
09:50-09:55	43	47	45	20	33	33
09:55-10:00	42	45	42	23	37	36

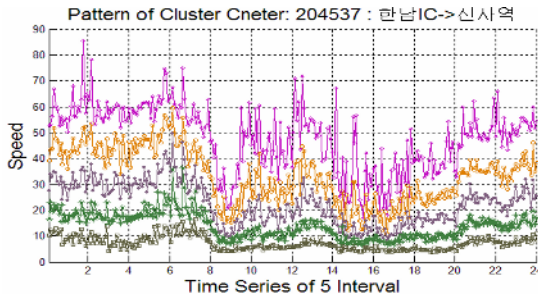


Fig. 4. Speed variation over 1 month between 0:00 am to 12:00 pm

During the office-going hours and closing times from 2:00 PM, there is sudden lowering of speed and at low-peak hours rising speeds are observed as in figure 4. Furthermore, speed distribution varies across different links due to the state of road and environmental factors. Figure 4 shows variation of speed about the center value of each cluster.

Performance of the proposed model about incident detection is evaluated by using the detection rate and false alarm rate of incident detection. Formula for the incident detection rate (DR) and false alarm rate (FAR) are as follows:

$$DR = t / (t + t'), \text{ FAR} = f / (t \times s) \tag{6}$$

t is number of real incident, t' is number of incident that were undetected, f is number of false alarms on incident detection and s is number of links.

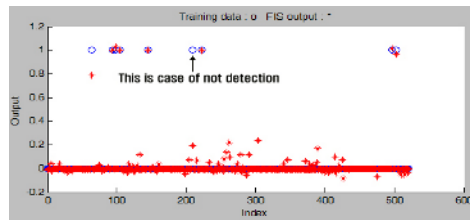


Fig. 5. Result of learning using ANFIS

Figure 5 shows result of learning by using speed and increase and decrease rate. Blue circle represents target value and red star the real output value. In figure5 successful incident detection in seen in 8 cases and misclassification in a single case when using a data set of 1200 input values. Based on the empirical analysis the average detection rate is over 83% and the false alarm rate is under 23%. Our algorithm has plausible performance when applied to the collected GPS data and is robust in the presence of limited data.

5 Conclusion

The speed variation of vehicles in urban streets poses serious challenges in link travel time estimation and incident detection. Additional factors like Road state, environmental factors and the integration of speed data from multiple links result in a high volume of computation. In attempting to develop an optimal link travel time algorithm we adapt the well-known FCM for clustering and link speed estimation from real-time GPS-based probe car data. A robust incident detection model was also developed using the ANFIS system in conjunction with the feedback from our FCM algorithm. The empirical analysis suggests that our approach offers plausible performance in the presence of limited data and high speed variance. Our approach also offers a reduction in the computational costs and robustness in the presence of link failures. While this research aims to enhance accuracy in estimation of link speed and incident detection further work is required in fusing data from multiple sources and integrating speed information from arterial roads and expressways.

Acknowledgement

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program super-vised by the IITA (Institute of Information Technology Assessment).

References

1. Hoppner, F.: "A Contribution to Convergence Theory of Fuzzy c-Mean and Derivatives," IEEE Transactions of Fuzzy System, Vol. 11. NO 5, 2003
2. You, J., Kim, T.J.: "Development and evaluation of a hybrid travel time forecasting model," Transportation Research Part C: Emerging Technologies, Vol. 8. Issues 1-6, 2000, pp. 231-256
3. Bruce R. Hellinga, Liping Fu, "Reducing bias in probe-based arterial link travel time estimates," Transportation Research Part C: Emerging Technologies, Vol. 10. Issue 4, 2002, pp. 257-273
4. Jesús Lázaro, , Jagoba Arias, José L. Martín, Carlos Cuadrado, Armando Astarloa, "Implementation of a modified Fuzzy C-Means clustering algorithm for real-time applications," Microprocessors and Microsystems, Vol. 29. Issues 8-9, 2005, pp. 375-380
5. Dipti Srinivasan, Ruey Long Cheu, Young Peng Poh and Albert Kim Chwee Ng, "Development of an intelligent technique for traffic network incident detection", Engineering Applications of Artificial Intelligence, Vol. 13, Issue 3, 1 June 2000, pp. 311-322.
6. Dipti Srinivasan, Xin Jin and Ruey Long Cheu, "Adaptive neural network models for automatic incident detection on freeways", Neurocomputing, Vol. 64, March 2005, pp. 473-496.
7. Sheu, J.B.: "A fuzzy clustering-based approach to automatic freeway incident detection and characterization", Fuzzy Sets and Systems, Vol. 128, Issue 3, 16 June 2002, pp. 377-388.
8. Jih-Biing Sheu, "A sequential detection approach to real-time freeway incident detection and characterization", European Journal of Operational Research, Vol. 157, Issue 2, 1 September 2004, pp. 471-485.
9. Boyce, D., Roupail, N., Kirson, "A Estimation and measurement of link travel times in ADVANCE project," Proceedings of Vehicle Navigation and Information Systems Conference, IEEE, 1993, pp.62-66.
10. Sisiopiku, V., Roupail, N.: "Exploratory Analysis of the Correlations Between Arterial link Travel Times and Detector Data from Simulation and Field Studies," ADVANCE Working Paper No. 25, Oct. 1993.
11. Nelson, P. and Palacharla, P.: "A neural network model for data fusion in ADVANCE," in Proceedings of Pacific Rim Conference, Seattle, Wash., 1993, pp. 237-43.