JOURNAL OF TRANSPORTATION
SYSTEMS ENGINEERING AND INFORMATION TECHNOLOGY
Volume 13, Issue 5, October 2013
Online English edition of the Chinese language journal

Available online at www.sciencedirect.com
ScienceDirect

Cite this article as: J Transpn Sys Eng & IT, 2013, 13(5), 30−36.

RESEARCH PAPER

Monitoring the Evolution of Traffic on Main City Roads

GONG Jinli[*1], PENG Xianwu[2]

1 School of Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China
2 Sany Heavy Industry Co., Ltd, Changsha 410100, China

Abstract: In order to monitor the evolution of traffic on main city roads, a fixed-point data monitoring system is devised. Firstly, after qualitative traffic data (flow, speed, and occupancy) are acquired, they are then translated into the traffic qualitative state (congested or uncongested) by Fuzzy C-means Clustering algorithm. Secondly, the Congestion Evolution Index is determined using Rescaled Range Analysis of data mining. Finally, by taking a sequence pattern similarity measurement and applying condensed hierarchical clustering methods, the routine pattern is distinguished. Consequently, real-time outlier detection is realized by a distance-based outlier detection algorithm. This algorithm was successfully applied based on 11 days of fixed-point data on the eastern segment of the Shanghai North-South expressway, it is concluded that the outliers distributed in 12:10-13:20, 13:40-14:30 and 17:10-17:15 on September 30.

Key Words: traffic engineering; traffic state evolution, time series, outlier detection, fixed-point data

1 Introduction

When traffic congestion becomes a common difficulty for major cities, monitoring routine congestion evolution on-line and methods to avoid congestion are becoming more important. City main roads (expressway and arterial road) undertake a large number of traffic flows, which play a key role in total road net. The traffic state evolution monitoring on main roads can be conducive to make early warnings for more serious road net congestion in populous areas.

There were many studies conducted about congestion monitoring, both in China and other countries [1-4]. The research organized by institutes and scholars in America, China, Japan and Europe appeared in recent years. With many complex study objects (such as city road, intersection, super highway, road net etc.), application subjects (such as traveler, traffic manager, traffic planning and design person etc.), and data (such as floating car data, dual-loop detection data, video capture data etc.), the application scope of existing methods have greater pertinence in describing and monitoring congestion evolution. On the other hand, accompanied with the development of Information Traffic System (ITS), much of the collected data was detected by dual-loops located on city main road. Therefore, it is meaningful to aquire the congestion information from the data and create early warnings of outlier states which can lead to the road net congestion in populous areas. In this paper, taking the overall development trend of congestion on city main roads is the study object. The time series data mining is applied to the analysis of fixed-point data; (flow, velocity and occupy) both from historical and real-time database. In the end, the accrued time and severity of outlier are detected within the system, which will be useful for the traffic manager and traveler alike.

2 The model of city main road based on fixed-point data

According to the location of dual-loops, the road can be divied into various sections as the basic unit of the system. After the data of flow, velocity and occupancy are obtained, the Fuzzy C-means Clustering is applied to identify traffic state; congested or uncongested and is displayed in Fig. 1 [5]. If the roads meet the upstream and downstream both are congested, they are defined as ‘Failure sections’. The roads which do not meet the condition mentioned above are defined as ‘Non-failure sections’ and are shown in Fig. 2.

Received date: Mar 28, 2013; Revised date: Jun 10, 2010; Accepted date: Jun 20, 2013
*Corresponding author. E-mail: gjl2001cd@126.com

DOI: 10.1016/S1570-6672(13)60122-1

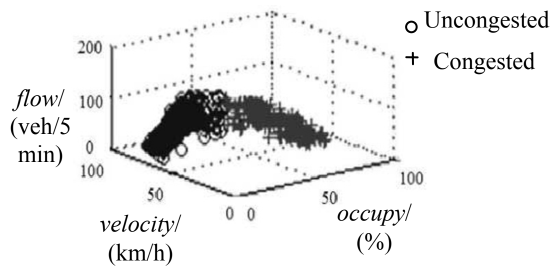

Fig.1 Fuzzy C-Means for traffic qualitative state

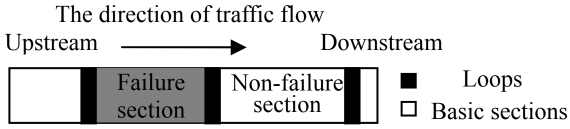

Fig.2 The model of city road

3 The indicators set of traffic state evolution

In 2001, Chen[6] took four indicators; Vehicle Distance Traveled (VDT), Vehicle Hours Traveled (VHT), Total Delay (TD), Productivity Loss (PL) as the traffic operational performance indicators. In order to access the severity of congestion on city main road, the following variables are defined on the basis of indicators mentioned above.

(1) Qualitative indicator of Failure

$$N_s(k) = \sum_{i=1}^{N} Num(\{B(k)\}) \quad (1)$$

where *Num*() is a counting function for set; N is the number of road sections; k is a number of detected time intervals, and one interval is usually equal to 5 or 15 minutes; $B(k)$ is a set of failure roads at No. k interval. According to the definition, N_s is less than or equal to N.

(2) Quantitative indicators of Failure

$$TVH(k) = \sum_{i\in L} \rho_i(k)\Delta x_i T_s \quad (2)$$

$$TVM(k) = \sum_{i\in L} \rho_i(k) V_i(k)\Delta x_i T_s \quad (3)$$

$$TD(k) = \sum_{i\in B(k)} (1 - V_i(k)/v_i)\rho_i(k)\Delta x_i T_s \quad (4)$$

$$TPL(k) = \sum_{i\in B(k)} (1 - q_i(k)/Q_{M,i})\Delta x_i T_s \quad (5)$$

where $TVH(k)$ is total Vehicle Hours Traveled at No. k interval, and its unit is veh(h), which denotes the total hours in that the road occupied by vehicles within statistical time. $TVM(k)$ is total Vehicle Distance Traveled, and its unit is vehicles-km, which denotes the total vehicle traveled distances within statistical time; $TD(k)$ is total Delay, and its unit is vehicles-hour, which denotes the total vehicle travelled delay caused by 'Failure sections'; $TPL(k)$ is total Productivity Loss, and its unit is km(h), which denotes the total loss of traffic capacity caused by 'Failure sections'; L is the set of all road sections; $\rho_i(k)$ is density of section i at No. k interval, and its unit is vehicles per km; $V_i(k)$ is space mean speed of section i at No. k interval, and its unit is km per hour; v_i is free flow speed of section i at No. k interval, and its unit is km per hour; $q_i(k)$ is flow out of section i at No. k interval, and its unit is vehicles per hour; $Q_{M,i}$ is traffic capacity of section i, and its unit is vehicles per hour; Δx_i is the length of section i, and its unit is km; T_s is sampling period (length of time interval) and its unit is h. The data from detection loop only contains flow q and time mean speed v_t, the space mean speed V and density ρ should be computed in following equations [7]:

$$V = v_t/(1 + c.v.^2) \quad (6)$$

$$\rho = q(1 + c.v.^2)/v_t \quad (7)$$

In the equations, $c.v.$ is the coefficient of variation, and its value range from 0.08 to 0.17.

The indicators mentioned above constitute the set of traffic congestion's assessment, which can be used as the basic indicators to determine "Congestion Evolution Index".

4 Design of traffic congestion evolution monitoring system

4.1 The determination of Congestion Evolution Index module

After the qualitative and quantitative indicators are defined, their consistency of trend changes should be verified in order to determine comprehensive index. The comprehensive index can then describe traffic state evolution from both a qualitative and quantitative standpoint. In time series data mining, R/S analysis is a method to validate whether or not the time series have a certain trend, by which the Hurst index can be calculated. The more similar Hurst indexes are the more consistent the two time series are. The calculation procedure is shown in Eq. (8) through Eq. (11) below [8]. Given a time series X which contains N values, $X(1)$, $X(2)$,⋯, $X(N)$, and its mean value $\overline{X}$, time series Y can be calculated in the following equation:

$$Y(k) = X(k) - \overline{X}, \quad k = 1,2,\cdots,N \quad (8)$$

Based on Y, time series Z and its range R can be calculated in the following equations:

$$Z_k = Y(1) + Y(2) + \cdots + Y(k), \quad k = 1, 2, \cdots, N \quad (9)$$

$$R = \max(Z_1, Z_2, \cdots, Z_N) - \min(Z_1, Z_2, \cdots, Z_N) \quad (10)$$

Then Hurst index can be calculated by Eq. (11).

$$(R/S)_n = cH^n \quad (11)$$

where n is the length of subsequence after time series X is divided into equal parts. Here, the value of n lies in the scope of [1, N] and it can be divided exactly by N; S is the standard deviation of subsequence; H is Hurst index whose value lies in the scope of [0, 1]. And the closer to 1 the value is, the more persistent the sequence is; c is a constant. For example, if N=100, the number n may be equal to 2, 4, 5, 10, 20, 25 or 50. $(R/S)_n$ will be computed firstly by the Eq.(8)-Eq.(10), and then H will be calculated by establishing regression equation of $(R/S)_n$ on n.