Study of the Intelligent Analysis and Prediction about Subway Passenger Flow during Large-scale Events

Jianchu Kang[1,a], Songsong Pang [2, b] and Yan Zhang[3,c]

[1]School of Computer Science & Engineering Beihang University 100191,Beijing, China

[2]School of Computer Science & Engineering Beihang University 100191,Beijing, China

[3]School of Computer Science & Engineering Beihang University 100191,Beijing, China

[a] kang@nlsde.buaa.edu.cn, [b]skyzhangy@163.com, [c]2329porter@163.com

Keywords: Large-scale events; Subway traffic flow data; Intelligent analysis and forecast.

Abstract. Through analysis of historical data of subway transportation, use the model extracted from the law of normal and large-scale activities' passenger traffic flow and combine the pattern library established by large-scale events' attribute information, this paper describes a method that provides a way to forecast the subway traffic passenger flow when a large-scale activity will happen. Before the occurrence of large-scale activities,We analysis and forecast the in and out passenger flow for the stations ,which around the place where the large-scale activity will happen ,and the entire road network which could describe the effect of this activity to the whole network .This system able to provide passengers with a travel reference to ensure the travel speed, security and comfort, and provide an important basis for the traffic management department to realize the effective real-time scheduling .

Introduction

With the rapid development of urban public transportation, urban passenger flow quantity increases sharply. System efficiency needs to be improved urgently. In such cases, urban public transportation researches focus on how to improve the public transportation's capacity , security and comfort. Urban subway transportation has large volume, high speed, good security, punctuality, protection of the environment, saving energy and land-use characteristics. The world is widely recognized: the fundamental way to solve the traffic problems of the city is to give priority to the development of subway transportation as the backbone of urban public transport system. With the constant improvement of equipments used by the traffic flow data acquisition , intelligent traffic flow analysis and forecasting, as well as intelligent transportation organizations become possible, but those existing studies are mostly focus on daily passenger flow singly or just do some experiments based on analog data . It is hard to meet the microscopic traffic control and real-time induced demand [1], therefore, in-depth study of the traffic flow forecasting theory and method becomes more and more important to ease road congestion, reduce pollution,and intelligent control and real-time induction. In this paper , we describe the way to realize the intelligent analysis and prediction system about subway passenger flow during large-scale events . Real experimental results illustrate this approach.

The Research Process of the Intelligent Analysis and Prediction System

The research on the subway traffic intelligent analysis and prediction system during large-scale events mainly include: data preprocessing, extraction of normal passenger traffic laws, extraction of exception passenger law of large events.We will specifically describe those four aspects below.

Data Preprocessings

Subway traffic flow data refers to all kinds of passenger flow information and data which is calculated and analyzed by subway traffic OD (Orign-Destination, the starting point - the end point) data.

Subway traffic passengers' OD data which is created by the credit card's transaction data when passengers in or out of the station, records passengers' travel processes from the starting point to the destination as well as the time when they in or out station accurately.

According to whether contains the OD travel route ,subway traffic OD data can be divided into two parts of the original OD data and the cleared OD data. The original OD data is used to statistics passenger entry and exit, the cleared OD data is used to analyze the flow of passenger flow distribution. Due to the forecast of staff, personal reasons of passengers and others, there will be some unusual OD data, needed to be preprocessed from the raw OD data, in order to provide good training samples for subsequent model.

Because some invalid data is out late ,data preprocessing is mainly to the OD data, so the passenger flow rule model is established to provide accurate data to support the follow study.

Abnormal data includes extreme data and statistical data in two parts[2]. Extreme data is due to the sinvestigation for passenger flow or the system error operation or other reasons for staff, resulting in abnormal data; statistical data is partly due to the waiting of passenger or pathfinding reasons .

Aiming at the extreme data, we could set the travel time interval, [Tmin-Tmax]where Tmin is the shortest travel time between O and D points; Tmax according to the statistical law of passenger subway time-consuming, can be provided for 4 hours.

According to the statistical abnormal data, the path travel time X approximately obey the normal distribution $N(\mu, \sigma^2)$, we could use the characteristics of normal distribution: $P\left(\left|\frac{x-\mu}{\sigma}\right| \leq 3.9\right) = 1$, and delete a statistical anomaly data by using successive iteration method.Zhichun - Xizhimen data on March 1, for example, before and after the treatment ,the frequency distribution of travel time is shown as follows .
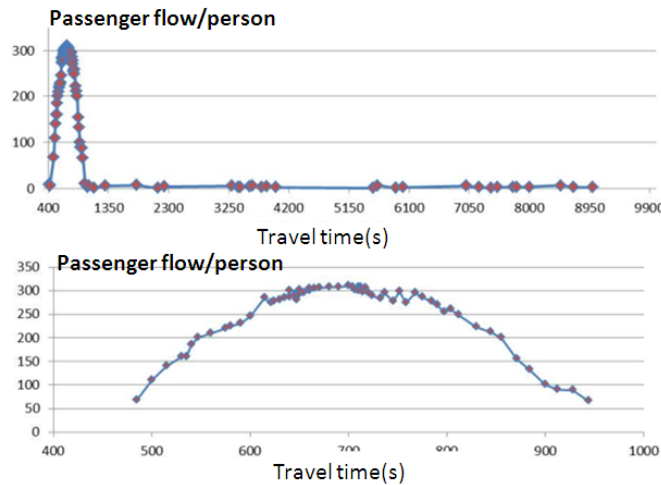

Fig. 1   Zhichun - Xizhimen Travel time frequency distribution graph on March 1

Establish the Extraction Model of Normal Passenger Subway Transportation Law

According to subway traffic characteristics of passenger flow, the intelligent analysis of the applicability of the SMO model is applied to short-term forecasts of subway traffic passenger flow, which gets the the law of normal passenger subway traffic in the same time.

This task provides a basis for the separation of the effect of the weather ,large-scaled events and other factors on passenger flow ,and it also provides the statistical basis for the filling of missing data[3].The basic idea of SMO algorithm is described as follow:The algorithm of SMO is the same like other improved SVM algorithms.The former could resolve the task to the minist scale,every time it just need to resolve the problem of two data samples, by analysis compared to other algorithms,like the SVM, which need to resolve the problem of two data samples at least,which means to improve the corresponding Lagrange factors,Because of the existing of equation restrain,it could not just improve a single variable. If given a group of samples of $(x_1, y_1), (x_2, y_2) ..., (x_n, y_n)$,the model of recurrent SVM could be expressed by as follow.

$$\min \frac{1}{2}\omega^T \omega + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1}$$

$$\text{s.t.} \begin{cases} y_i - \omega^T x_i - b \leq \varepsilon + \xi_i \\ \omega^T x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \tag{2}$$

ω is the normal vector of classification hyperplane,C is the penalize factor,which is used to control the level of the penalization,b is the constant factor, ε is insensitive loss function;ξ_i, ξ_i^* are slack variables. Adopting the antithetical parallelism,we could transform the orginal problem of original planning to quadratic programming,and establish the equation of Lagrange.

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i x_j) \alpha_i \alpha_j \tag{3}$$

$$\text{s.t.} \begin{cases} 0 \leq \alpha_i \leq C^+ \ if \ y_i = +1 \\ 0 \leq \alpha_i \leq C^- \ if \ y_i = -1 \\ \sum_{i=1}^n y_i \alpha_i = 0 \end{cases} \tag{4}$$

α_i is the factor of Lagrange,K(x$_i$,x$_j$) is the core function.Separately,$C^+$and C^- are the plus-minus penalize factors.

The process of SMO contains two steps as follow[4]. Firstly,because the linear rule in the former equation makes at least two Lagrange factors changed,we select two improved Lagrange factors.Secondly,there is only the improval problem to the two factors by using intelligent analysis method.Aiming to resolve the improval problem of the two factors,we need to calculate the restraint of the two factors firstly,and resolve the minimize problem with restraint.

Through analysising features of subway traffic flow,we adopt subway traffic flow data per five minute between five o'clock and twenty-four o'clock and gain the input and output from the model of SMO as follow: $x_t, t = 1, 2 \dots, n$ is the data calculated five per minute by the flow out of the station, $\hat{x}_t$,forecasted by the former data ,is the flow data of entry station at the time t . While the nearest K unit time sequence data as input ,the forecasted value of $\hat{x}_t$ as output.This is mean that the input sequence is thevector$i_t = (x_{t-1}, x_{t-2}, \dots, x_{t-k})$,the output sequence is $o_t = (\hat{x}_t)$,According to the p units of samples , $\{(i_1, o_1), (i_2, o_2), \dots, (i_p, o_p)\}$, and the principle described in the former about supporting vector machine regression, we could establish the forecasting model as the follow.

$$o_t = \sum_{m=1}^n (\alpha_i^* - \alpha_i) K(i_m, i_t) + b \tag{5}$$

So,α_i^*, α_i are the lagrange factors imported by calculation. $K(i_m, i_t)$ is the core function.Through the train model,we could obtain the value of the factor α_i^*, α_i, b, to move forward a single step, establish the extraction model.The process of applying the former principle in our research is described by the follow Fig. 2.

Taking advantage of the passenger flow data ranging from January 1 to March 15 about Wukesong Station,We extracted the inbound passenger traffic law of Wukesong station 5min timeshare week shown in the follow Fig. 3
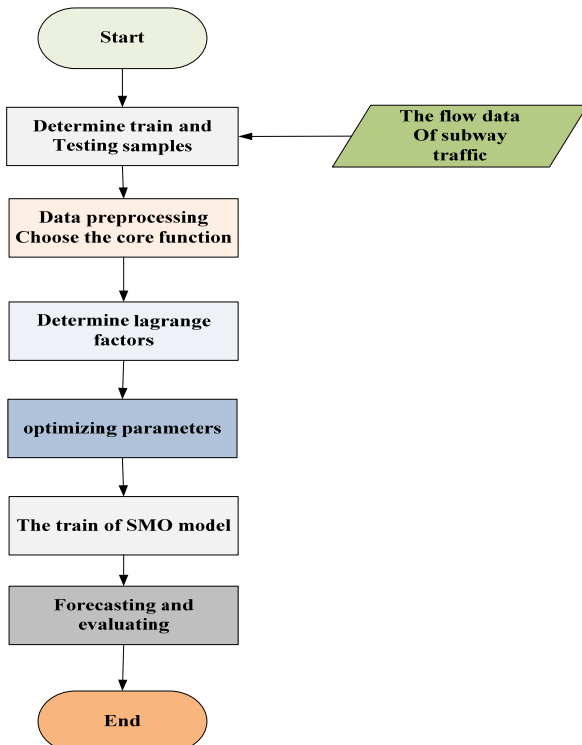
**Fig. 2 The passenger flow forecasting
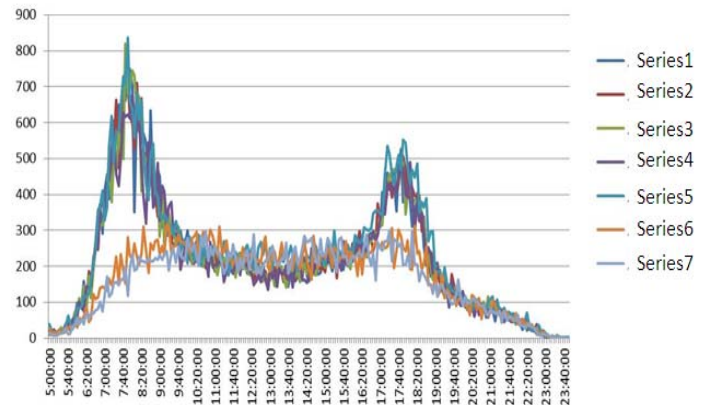of PSO-SMO flow-process**

**Fig. 3 The law of inbound flow in
Wukesong station every 5 Minute**

Establish the Extraction Model of Abnormal Passenger Subway Transportation Law

Large-scale activities are devided into two categories: ongoing activitives, and one-time activitives. Persistent large-scale activities for its passenger flow characteristics, We analyze the applicability of Gray - Markov model to predict extraction based on the persistent large-scale activities exception passenger law; For a one-time large-scale events, according to its passenger flow characteristics, EM clustering algorithm classifiy the passenger flow during the large-scaled events,and extract the one-time large-scale activities' abnormal passenger rule. This section is an important prerequisite for the establishment of large-scale activity patterns library.EM algorithm is an iterative algorithm, a method of maximum likelihood estimation of the parameters proposed in 1977 by Dempster, Laind, Rubin, this method can be widely applied to deal with missing data, censored data, with sick of data, such as the so-called incomplete data. The EM model specific process is as follow.

Gray model in dealing with volatile data, the prediction accuracy will be reduced[5]; Markov chain model applied to forecast the volatile data, but requires state with no aftereffect[6]. Gray - Markov makes up for the lack of both at the same time, greatly improve the accuracy of prediction, and provides an effective combination of predictive models for stochastic volatility data. The Gray - Markov model specific process is as follow.
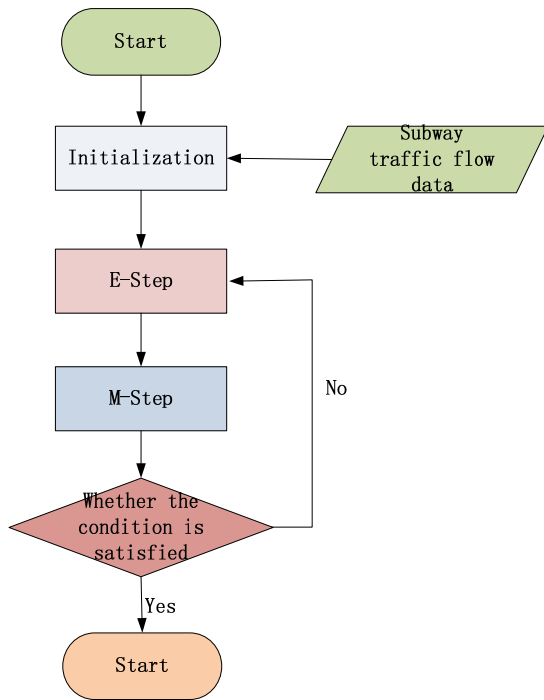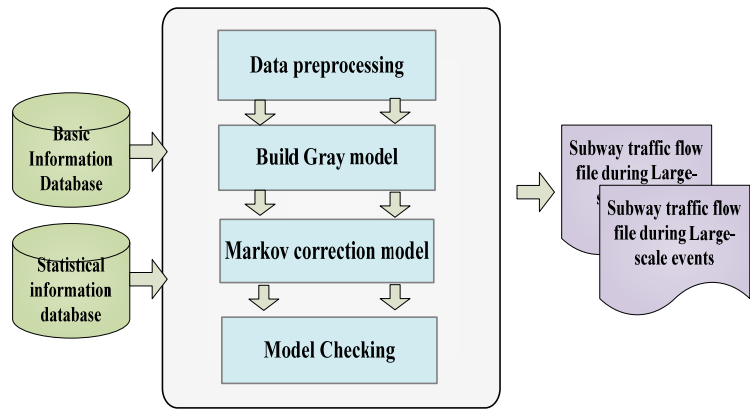
Fig. 4 Gray - Markov flow chart

Fig. 5 EM model process

Summary

In this paper, for the lack of existing domestic and foreign research methods and models of large-scale events, the intelligent analysis and prediction for subway traffic flow modeling is researched. PSO-SMO could intelligent correct the analysis of large-scale events impact on passenger flow, at the same time provide the statistical basis for the filling of missing data. After verification of the instance, we prove the validity of the model. Grey - Markov model provide a theoretical basis for the passenger flow forecast prior to the occurrence of large-scale events. For the intelligent dispatching it provides analysis data input. Therefore, the city public transport in the scheduling research can make the city bus operation efficiency partial improved.

References

[1]Benxuan Zhang, Research on Urban Rail Passenger Flow, Master's thesis, School of Transportation, 2009. (in Chinese)

[2] Renbin He, Study on Macroscopieal Models of Divinable Passenger Flow of Subway, Master's thesis, Chongqing University, 2002. (in Chinese).

[3] Chrobok R, Kaumann O, Wahle J, et al. Different Methods of Traffic Forecast Based on Real Data[J] .European Journal of Operational Research, 2004, 155(3) : 558.

[4] Smith B L, Williams B M, Oswald R K. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting [J] . Transportation Research Part C, 2002, 10(4) : 303.

[5] Xue-zhi Kou,Qi-yi Zhang.The Forecast for the Wear Trend of the Diesel Engine Based on Grey Markov Chain Model [C] Dec.2009. Proceedings of the 2009 Second International Symposium on Computational Intelligence and Design(in Chinese)

[6] Yen-Tseng Hsu, Ming-Chung Liu, Jerome Yeh, Hui-Fen Hung,Forecasting the turning time of stock market based on Markov-Fourier grey model [J], May. 2009 Expert Systems with Applications: An International Journal, Vol 36,No 4(in Chinese)