Vol.2019-MBL-93 No.5
Vol.2019-ITS-79 No.5
2019/11/20

Anomaly Event Detection using Bus Management Information

-Case Study of Anomaly Operation Status Detection and Its Application

WEI SUN[1] KEI HIROI[1,2]
TAKURO YONEZAWA[1,2] NOBUO KAWAGUCHI[1,2,3]

Abstract: At present, the anomaly detection based on GPS data of buses has become a hot spot in the field of intelligent transportation. The goal of anomaly detection aims to automatically detect the abnormal situation of the city, which is also a key step to develop the intelligence of traffic system. This research aims to propose the anomaly detection at three aspects: the definition, the method, and the practical use of the detection results. In this paper, we focus on the abnormality of traffic situations, which leads to anomaly operation status (bus bunching), caused by mutation of the number of the passenger or the traffic accident. We propose a method for detecting bus bunching. We define a situation that satisfy the three conditions as bus bunching. The analysis result shows that the data of April 2017 of the city Okazaki occurred 49 times and occurred at Kouseicho stop and Daijuji stop at 6p.m frequently. After analyzing the data of the bus bunching, we plan to develop a module to predict the bus bunching.

Keywords: Anomaly detection, Bus location data

1. Introduction

With the rapid development of information and communication technologies and the wide application of location-aware devices such as in-vehicle GPS, the trajectory of buses not only can display a large amount of information about the daily activities of the population, but also can show road traffic conditions to improve the quality of life.

Abnormal vehicle behavior is an important cause of congestion on road and traffic accidents. In this paper, the goal of anomaly detection aims to automatically detect the abnormal situation of the city, which is also a key step to develop the intelligence of traffic system.

As one of the ways to effectively alleviate traffic congestion, urban conventional public transportation has been strongly promoted because of its low price and efficient use of resources. It is a national policy to give priority to the development of urban public transportation. How to further improve the level of public transportation services is urgent the practical significance and the good theoretical research value.

2. Related work

2.1 Analysis using bus location data

Bus location systems have recently emerged that can easily obtain various data, such as arrival time and departure times or travel locations. Few researchers use the bus location data to analyze the factors, predict the delay and the number of passengers.

The research by Ono et al. [1] is based on a comparison of actual data and dia. The average time duration and standard deviation are calculated based on factors such as the presence or absence of typhoons, temperature, month, day of the week, and their trends are analyzed.

Maekawa et al. [2] conducted data analysis for the purpose of predicting arrival time as shown in Figure 2.2. The effect of the delay time of the bus stop that passed immediately before and the required time the relationship between the number of passengers and the delay time is analyzed.

Imai et al. [3] proposed a bus arrival time prediction method using multiple regression analysis and Kalman filter for the operation results data of multiple operators, and verified the effects of data amount, outliers and dynamic accuracy improvement.

Arai et al. [4] proposed an arrival time estimation method and a visualization model. The arrival time estimation model dynamically updates the accuracy via an estimation method using a combination of a multiple-regression model and a Kalman filter. The visualization model analyzes relationships between delays and various factors. The goal of this study is aim to realize a society where people can use buses more comfortably.

2.2 Anomaly detection

Nowadays, there are a lot of algorithms for data mining and anomaly detection. Due to the characteristics of mobile trajectory big data, there is still no mature trajectory anomaly detection algorithm yet.

Outlier detection is an important part of the trajectory anomaly detection. Outliers are data objects that differ in behavior or pattern from the data in the data set. These outlier data objects may contain important information, or information that is of particular interest. Excavation of outliers have a wide range of applications, such as credit card fraud detection, network intrusion discrimination. Three broad categories of anomaly detection techniques exist : Unsupervised anomaly detection, Supervised anomaly detection, and Semi-supervised anomaly detection. Existing outlier detection methods include methods distribution based, distance based, density based, bias based.

2.3 Bunching detection

According to the different control reference points, the research methods of the bus bunching problem in foreign countries are divided into three categories [8]: based on timetable, based on the headway and based on real-time information. Rufi [9] uses the bus to arrive at the station time, departure time, stop time, the number of passengers getting on and off, and basic data of the line, building

1 Graduate School of Engineering, Nagoya University
2 Institute of Innovation for Future Society, Nagoya University Japan
3 Location Information Service Research Agency

a microscopic traffic simulation platform and establish a line bus simulation model. Chen et al. [10] proposed a control strategy based on multiple control points, considering the passenger getting on and off the car when using the control method. Bartholdi et al. [11] abandoned the traditional timetable and even pre-set the concept of the target head-to-head distance, to control the headway convergence speed and the front-end time limit to strengthen the balance respectively. By adjusting the parameters to increase or decrease an extra bus, the system will adjust the speed of the target vehicle and the distance between the front and the rear.

3. Proposed method

In this paper, there are two types of data anomaly confirmed from the data, one is the abnormality of the data itself, including data missing and data errors. To delete redundancy and use the mean of the previous bus stop and the last bus stop's data to fill in missing data of missing bus stop data, including the latitude and longitude or the time of bus arrival time. The other is the abnormality of traffic situations. It is mainly for the detection of bus bunching, mutation of delay and the passenger number or the traffic accidents. In the detection of bus bunching, there are three conditions for the formation of bunching bus. The actual arrival time difference of two buses passing through the same stop is less than 2 minutes and their schedule arrival time difference is more than 2 minutes. The two buses pass through the same five stops on a road. The number of passengers getting on the one bus is 5 people more than next bus at the same stop.

3.1 Step1: Bus bunching detection

1) Comparing each two diagrams of 523 diagram a day. At the same bus stop, search for each two buses at every bus stop with a difference of actual time ($rtime\_diff$)<120s and a difference of schedule time($stime\_diff$)> $120s$.

   i.$rtime_{diff} = realtime_{d+1} - realtime_d$

   ii.$stime_{diff} = schetime_{d+1} - schetime_d$

   iii.$rtime_{diff} < 120\text{s}\ \&\ stime_{diff} > 120s$

   iv.find $d_{(sys,stop,realtime)}$

2) We would like to define the two buses that passes through five consecutive same bus stops can happen the bun bunching, because two buses often can meet at the same bus stop.

3) Based on the two buses have passed through five consecutive same bus stops, at the same bus stop, that the passengers of the previous bus getting on the bus five more than next bus.

3.2 Step2: Bus bunching prediction

1) In DTW, the distance between each point in two time series is compared, and then the path with the shortest distance between the series is found. This is the DTW distance. Therefore, DTW distance can be defined even if the two sequences have different periodicities / lengths.

   When there are two time series data for which you want to calculate similarity:

$$S = s_1, s_2, \ldots . s_m$$
$$T = t_1, t_2, \ldots . t_n$$

You can think of a grid in which each time series data is arranged horizontally and vertically. Point s_1 of this grid corresponds to the alignment between Point s_2. The Warping path($W = w_1, w_2, \ldots . w_k$)aligns each element of the time series data S and T so that the "distance" is small.

That means W is a series of grid points. Here, if the distance between two elements is defined as $\delta(i,j) = |s_i - t_j|$, DTW is defined as

$$DTW(S,T) = \min_W \sum_{k=1}^{P} \delta(w_k)$$

Therefore, the computational complexity is $O(mn)$.

In this paper, the time series we used are two bus time series where the two bus happened the bus bunching. The length of two time series is same ($m =$ n), so that we propose to calculate the DTW directly.

2) The trend of two time series bus data is very similar, so that we surmise the bus bunching is occurring when the trend of the previous and next buses is similar. We purpose to create time series bus data of each two diagrams and calculate Dynamic Time Warping (DTW) of each diagram.

3) The method of calculate DTW.

   i. From the first bus stop, add bus stop one by one to make some time series.

   ii. Comparing the DTW values of the average delay for these DWT values of two diagrams on Tuesdays in April

   iii. From the results, the delay time and DWT value may change depending on the day and time band, so the average value excluding the anomaly is required.

4. Result

4.1 Data set

The actual data used are bus operation data collected by Meitetsu Bus Co., Ltd., through a bus arrival information system. Their bus arrival information system records data for each bus, including the unique bus ID, the actual/scheduled arrival/departure times at each bus stop, and the numbers of passengers who get on/off at each bus stop. This study uses data from Okazaki City, Aichi Prefecture, Japan, which is a major city with a large number of buses and passengers. Meitetsu Bus Co. has 710 buses, 1539 bus stops and 523 routes in Okazaki City. The data ranges correspond to January 1, 2017, and from January through October 2018. In particular, we use the data collected on April, 2017, the day when the number of passengers was the largest between April, 2017.

4.2 Result of bus bunching detection

We propose to use the data of April 2017, that the bus bunching has occurred 25 times. The bus bunching often happened at route Okudonojin ya ~ Kosei-cho ~ Higashi Okazaki ~ JR, and route Daimon-eki ~ Kosei-cho ~ Higashi Okazaki ~ JR, which include the four bus stops: Tonobashi, Kosei-cho, Daikiji, Okazaki Genkikan mae at 6pm. There are two types of bus bunching showed at Fig 4.1 and Fig 4.2. The Fig 4.1 showed the two buses in the same route at 7am, and the FIG 4.2 showed the two buses in the different route at 2pm. The horizontal axis is bus stop sequence and the vertical axis is the time which include actual time and scheduled time. As we can found that when the difference of actual time of two buses became very small, that the bus bunching was happening. The red marks in the Fig 4.1 and Fig 4.2 are bus stops where the bus bunching happened.

The Fig 4.3 and Fig 4.4 showed the actual time of difference is floating in a small range. It means that the pervious bus passed a bus stop, then the next bus also passed the same bus stop in a short time. If the two buses passed five stations continuously, we can make sure the bus bunching happened.

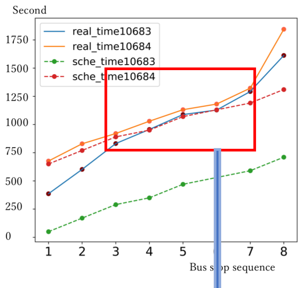

Fig4.1 the actual time and scheduled time of same route

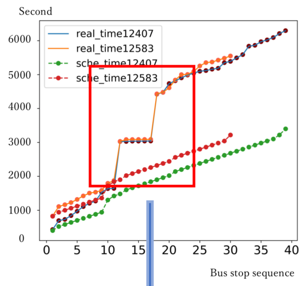

Fig4.2 the actual time and scheduled time of different route

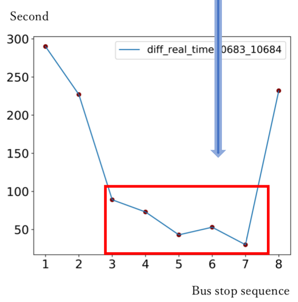

Fig4.3 the difference of two actual time of the same route

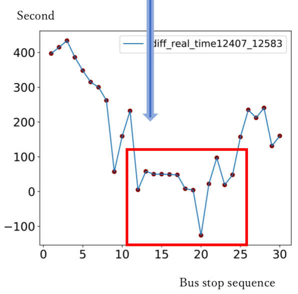

Fig4.4 the difference of two actual time of the different route

As a result, the bus bunching often happened at Okazaki near the subway station, because the most of the buses in Okazaki pass the subway station, and the traffic volume on this road in Dalian is also big. Those reasons lead to the Traffic jam behavior at bus delay caused the bus bunching that an anomaly situation.

4.3 Result of bus bunching prediction

The Fig 4.5 shows that the bus bunching happened between the bus 10148 and bus 10313. The red point and the orange point are the road where two buses have happened the bus bunching which began at the bus stop sequence from 0.

Fig 4.5 shows the situation with the condition of the number of the previous bus getting on the bus is more than next bus. The Fig 4.6, Fig 4.7 show the situation without the condition of the number of the previous bus getting on the bus is more than next bus.

Considering the influence of various factors, and this bus bunching occurred on Tuesday, April 4. In order to taking into the impact of the week and time period, we propose to calculate the average of delay of Tuesdays April without April 4. As the Fig 4.6 shows relationship of the mean delay and the delay of two buses. It can make sure that the previous bus 10313 do not delayed for so long as usual. When the bus bunching was happening, the bus 10313 has a long delay in that day. So that it may be something happened at the road, caused the bus delay, reason to the bus bunching.

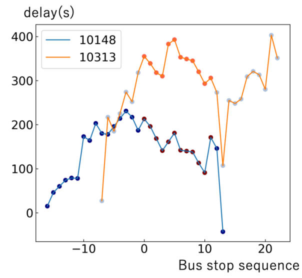

Fig 4.5 delay On April 4

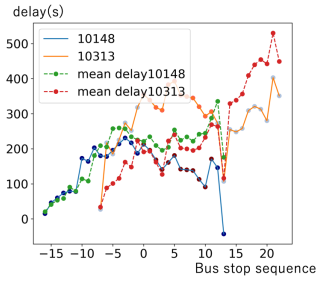

Fig 4.6 delay on April 4 and average delay of April

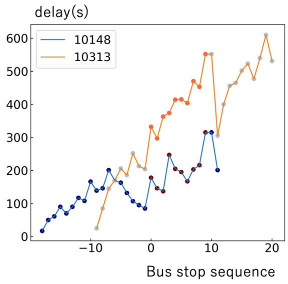

Fig 4.7 delay On April 10

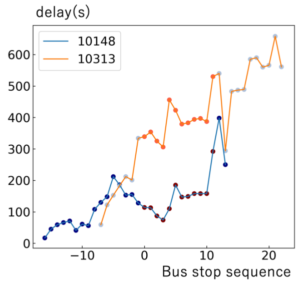

Fig 4.8 delay on April 11

The same situation happened at the same buses 10313 and 10148 on April 10 and April 11. The bus also had a long delay on the road. The April 10 and April 11 do not meet the conditions of the previous bus getting on the bus is more than next bus, so that the delay of April 10 and April 11 do not delay in the number of passengers.

We propose to use bus delay to calculate the DTW values of two buses time series. Considering the influence of passenger number of bus bunching happened, we separate to calculate the DWT values of bus bunching without the condition of the number of the previous bus getting on the bus is more than next bus or not.

i. The DTW values of bus bunching without the condition of the number of the previous bus getting on the bus is more than next bus.

The Table4.1 showed that the DTW values of bus 10313 and 10148. Because of the scheduled time of bus 10313 is early than 10148, so we calculate the DTW values of 10626 to make sure the values change when the bus bunching happening. When using bus stop 1~ bus stop 30 to make a time series, the DTW value become bigger and bigger.

ii. The DTW values of bus bunching with the condition of the number of the previous bus getting on the bus is more than next bus.

To show some typical data in Table 4.1:

Table 4.1: DTW values of 10313 on April

Source data	Compare data	Bus stop 1~2	Bus stop 1~3	Bus stop 1~4	…	Bus stop 1~30
April 4 of 10313	April 11 of 10313	100.24	105.38	113.44	…	605.41
	April 18 of 10313	133.06	177.18	227.69	…	488.22
	April 25 of 10313	157.71	172.26	223.24	…	431.62
	April mean of 10313	96.39	115.08	141.16	…	322.25
April 11 of 10313	April 18 of 10313	47.20	96.78	135.13	…	444.41
	April 25 of 10313	77.80	92.52	136.23	…	484.90
	April mean of 10313	26.77	40.49	48.81	…	382.91
April 18 of 10313	April 25 of 10313	30.61	34.08	31.67	…	177.53
	April mean of 10313	36.77	65.43	89.62	…	278.16
April 25 of 10313	April mean of 10313	64.03	64.31	88.12	…	318.93

As Fig4.9 shows that blue points in the DTW of April 4 and April 11 is very similar, much smaller than others. The black points of DTW of April 18 and April 25 is also very similar, and smaller than others. The DWT of April 4 and April 18 is much bigger than DWT of April 4 and April 11. That means the April 4 and April 11 without considering the number of passengers, have a similar delay condition. Both of them meet the condition of delay on the road. As the Fig 4.9 showed, when the DTW value has a range of 100~170, is easily to happen the bus bunching.

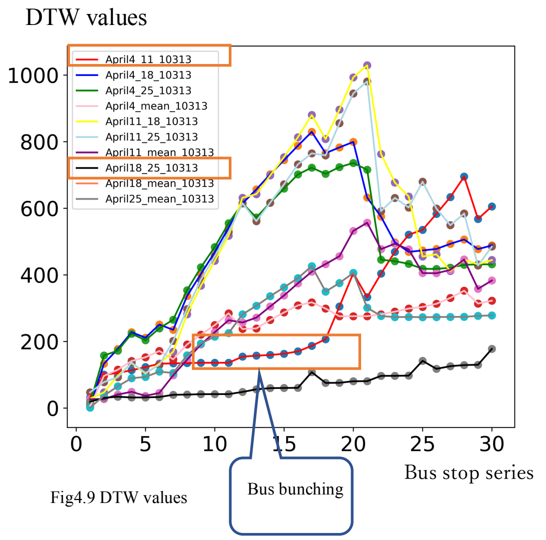

Fig4.9 DTW values

The Fig 4.9 showed the DTW values calculated by delay data. Considering factor of the number of passengers, we propose to adjust the weight of factor of the number of passengers and the weight of factor delay or the other factors to improve prediction accuracy.

5. Conclusion and future work

The bus operation situation is easily influenced by various factors, such as traffic congestion or weather conditions. The number of passengers is also a factor that affects the delay which can cause the bus bunching: the more passengers need to get on and off, the more time the bus spend at the bus stop, and the more arrival time at the next stop will be delayed as the number of passengers increases .In this study, for the purpose of detecting anomaly events, considering the past bus location data and the number of passengers, examination of data analysis, detection and prediction methods for one of the anomaly situations -bus bunching. In future work, based on the results of this analysis, we devised predictions of bus bunching considering the number of passengers. Continuing to detect other anomaly events, including traffic accident, celebration event.

Reference

[1] In Japanese (“Gps を活用したバスの定時運行評価に関する事例分析. 土木計画学研究・講演集, 2003.)
[2] In Japanese(前川裕一，林夏美，牧野友哉，白石陽. バス到着時刻におけるバス運行所要時間データと乗降客数データの活用. マル チメディア通信と分散処理ワークショップ, 2013.)
[3] Hitomi IMAI, Kei HIROI, Nobuo KAWAGUCHI. Arrival Time Estimation and Visualization based on Bus Traffic Data” ITS AP Forum 2018.
[4] Takehiro ARAI, Kei HIROI, Nobuo KAWAGUCHI. Simulation for Passengers Convenience using Actual Bus Traffic Data
[5] Mei Chen, Xiaobo Liu, Jingxin Xia: Dynamic Prediction Method with Schedule Recovery Impact for Bus Arrival Time, Transportation Research Record Journal of the Transportation Research Board, pp.208 - 217 (2005)
[6] Location Information Service Research Agency, http://lisra.jp/en.CitedJanuary2018
[7] Zhang Jian, Li Mengtian, Ran Bin, Li Wenquan. Causes and forecast modeling of conventional bus bunching. Dongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Southeast University (Natural Science Edition)
[8] Zolfaghari S, Azizi N, Jaber M Y. A model for holding strategy in public transit systems with real-time information [J]. International Journal of Transport Management, 2004 , 2(2): 99-110. DOI:10.1016/j.ijtm. 2005. 02. 001.
[9] Rufi F M. Optimization analysis of the number and lo- cation of holding control stops to prevent bus bunching. [D]. Stockholm, Sweden: Kungliga Tekniska Hogskolan, 2011.
[10] Chen Q, Adida E, Lin J. Implementation of an iterative headway-based bus holding strategy with real-time information [J]. Public Transport ゾ 2012 ゾ 4(3) : 165-186. DOI: 10. 1007 / s12469-012-0057-1.
[11] Bartholdi J J, Eisenstein D D. A self-coordinating bus route to resist bus bunching [J]. Transportation Re- search Part B ゾ 2012 ゾ 46 (4) : 481-491. DOI: 10. 1016 /j. trb. 2011. 11. 001.
[12] Chen Zhang and Jing Teng. Bus dwell tme estimation and prediction: A study case in shanghai-china. Procedia - Social and Behavioral Sciences, Vol. 96, pp. 1329 - 1340, 2013.
[13] Claudio Piciarelli, Christian Micheloni, Gian Luca Foresti. Trajectory-Based Anomalous Event Detection
[14] Takashi Nagatani. Chaos control and schedule of shuttle buses. Physica A: Statistical Mechanics and its Applications, Vol. 371, No. 2, pp. 683 - 691, 2006.
[15] Yingfeng Cai, Hai Wang, Xiaobo Chen, Haobin Jiang. Trajectory-based anomalous behaviour detection for intelligent traffic surveillance

Acknowledgement

We wish to thank Meitetsu Com and Meitetsu Bus Company Limited for insightful suggestions and provision of bus traffic data. This research and development work was supported by the JST OPERA.