Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 31 (2014) 48 – 57
Procedia Computer Science

2nd International Conference on Information Technology and Quantitative Management, ITQM 2014

VAPA: Vehicle activity patterns analysis based on Automatic Number Plate Recognition System Data

Yuyan Sun[a,b,c], Hongsong Zhu[a,b,]*, Xinyun Zhou[a,b], Limin Sun[a,b]

[a] Beijing Key Laboratory of IOT information security technology, Institute of Information Engineering, CAS, Beijing，100093, China
[b] *State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing，100093, China*
[c] *University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract

With the explosive growth and wide spread of traffic cameras deployed on the road networks, the amount of Automatic Number-Plate Recognition (ANPR) data captured daily by traffic cameras is very substantial. In this paper, we apply data-mining techniques to discovering vehicle activity patterns from ANPR data. We propose some quantitative indicators of vehicle trace features, and vehicle activity classification method based on the feature of vehicle trace. Evaluations base on collected ANPR records show the capability and efficiency of the proposed approach.

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.

Keywords: Datamining, ANPR, Activity Pattern, Classification;

1. Introduction

The advanced positioning, monitoring, communication and other information technologies nowadays are widely used in intelligent transportation system. Intelligent traffic data analysis has become an active research area in recent years[1], for example the vehicles activity pattern classification.

By the vehicle activity pattern classification analysis, traffic management department could get the percentage composition of various activity pattern types of vehicles. Furthermore, we could get the related regional distribution

* Corresponding author. Tel.: +8613611261703; fax: +8601082546564..
E-mail address: zhuhongsong@iie.ac.cn

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.
doi:10.1016/j.procs.2014.05.244

of, to more effectively deploy resource to ensure the stable situation of the traffic. And the vehicle behavior analysis result would provide important reference information on traffic signal timing conducting, for example, public buses route designing, logistics traffic limiting program, commuter car traffic scheduling and so on. Abnormal activity pattern detection and notification could help processing emergencies events and reduce the impact on traffic flow.

Intelligent transportation system aggregated various types of traffic data information, including vehicle real-time speed, road flow, accidents, construction lane, as well as parking spaces occupation and so on. The traffic data information can be acquired by magnetic, wave, video and mobile communications technology. For example, the positioning device with GPS module or mobile phone installed in vehicle[2,3] could record vehicle location, instantaneous speed, travel time, travel speed, long-term running trajectory and other traffic information. And the induction coil buried in the intersection[4] and the video camera deployed at fixed location could get the intersection traffic flow.

The likelihood of the interacting multiple model tracking[5] is used to classify the vehicle's turning behavior. The hidden Markov modeling is used to model the typical trajectories and classification highway driving type[6]. The Bayesian networks are used for classifying the vehicle's behavior and predicting the vehicle's lane change[7]. Quaternion-based rotationally invariant longest common subsequence (QRLCS) metric is used to match observed trajectories to preordered trajectories, and characterize behavior at roundabouts[8]. Variation Gaussian mixture modeling is used to classify and predict the long-term trajectories of vehicles[9]. Discrete Fourier transformation method was proposed for seasonal traffic variation pattern classification through[10].

The amount of Automatic Number-Plate Recognition (ANPR) data captured daily by traffic cameras is very substantial[11]. ANPR data recorded the spatial and temporal attribute information of all vehicles on the road. Some selected characteristic values such as frequency, date and location sequence, could indicate a wealth of knowledge, for example the vehicle active behavior. We could observe and analyses the vehicle behavior based on the ANPR data and region GIS data. The vehicles could be classified when the characteristic value sequences match the temporal and spatial similarity threshold.

The remainder of the paper is structured as follows. We describe the spatial and temporal feature quantify of vehicle trajectory in Section 2 and Section 3. In Section 4 our vehicle activity pattern classification method is established. Section 5 starts with the ANPR data filter and redundant and then gives the performance evaluations and some interesting analysis result. Finally we make some concluding remarks.

2. Spatial Feature Quantify

In this section we discuss the spatial feature extraction from vehicle's ANPR records. And the spatial feature quantify algorithm detail was shown in Fig. 1.(b).All deployed gateways have been divided into two parts, Border gateways and Interior gateways as shown in Fig. 1.(a).

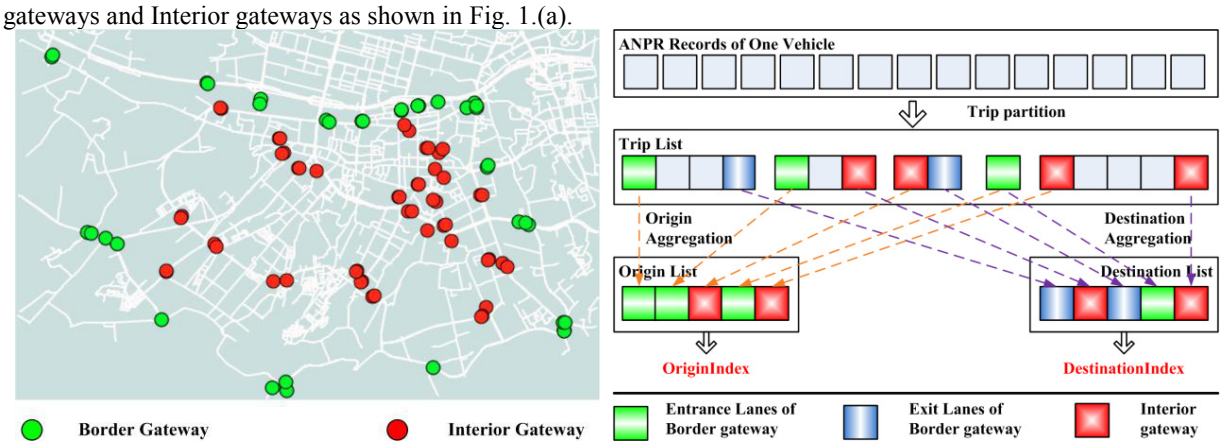

Fig.1. (a) Gateway geography division; (b) an example of spatial feature quantify algorithm

The division is base on the gateways deployment information and the GIS map data. Each Border gateway's lanes were divided into "exit lane" and "entrance lane". When the vehicle passes the Border gateway's exit lane, it will leave this region. And the vehicle will enter this region when it running on the Border gateway's entrance lane.

On processing the vehicle's ANPR records, the spatial feature quantify algorithm first splits the records list into multiple trip lists, which records a short-term trip trace, by the time interval between ANPR records. The gateway of one trip's first record is the origin of the trip, the gateway of last record is the destination of the trip.

Then the origins of all trips would be collected in a list and same to the destination. We design trip origin index (TOI) to indicate source feature of vehicle trace and trip destination index (TDI) to indicate destination feature of the vehicle trace. The higher TOI is, the more possibility the vehicle is foreign. The calculation method of TOI and TDI was show in Algorithm 1.

Algorithm 1: The spatial feature quantify algorithm

```
Input:  tripList: an empty list records the vehicle's trips;
        recordList: a list records the ANPR records refers to the vehicle with plateNo
        threshold: the minimal time duration between two trips
        entranceDict: a hash table records entrance lane of Border Gateway;
        exitDict: a hash table records exit lane of Border Gateway;
Output: spatialFeature: an object contains (TOI: trip origin Index, TDI: trip destination Index)
new trip object and do initialization with recordList[1];
TOI =0, TDI=0
for i=2;i≤ recordList.count; i++ do:
|   if recordList [i]. time - recordList [i-1]. time > threshold then
|   |   push trip to tripList
|   |   new trip object and do initialization with recordList [i];
|   else
|   |   push recordList [i]to trip
|   end
end
for i =1; i≤ tripList.count; i++ do:
|   trip= tripList[i]
|   if entranceDict .has_key((trip [1]. gateway, trip [1]. lane)) then
|   |   TOI = TOI + 1
|   end
|   if exitDict .has_key((trip [-1]. gateway, trip [-1]. lane)) then
|   |   TDI = TDI +1
|   end
end
new spatialFeature object
spatialFeature.TOI=TOI/ tripList.count
spatialFeature.TDI=TDI/ tripList.count
return spatialFeature
```

3. Temporal feature quantify

The vehicles present some certain dates in a long-term time. And the vehicle would active at certain hours such as the evening hour or rush hour. We acquire the daily attribute of the temporal ANPR records queue by Discrete Fourier Transforms[12].As shown in fuction.1, $f(n)$ is presence value of the vehicle at the date point n. $F(m)$ is the spectral coefficient for the mth wave component. N is number of date time points, which is the days of the collected data. For the input function f (n) is entirely real then the negative frequencies are completely redundant, giving no

additional information. The magnitude squared of the Fourier coefficients, $|F(m)|^2$, called the power. The $|F(0)|^2$ is called the presence frequency ratio (PFR), and the mean of $|F(m)|^2$ is called presence discrete coefficient(PDC)

$$F(m)=\sum_{n=0}^{N-1} f(n)\exp\left(\frac{-i2\pi mn}{N}\right) \tag{1}$$

Furthermore, the rush-hour active index (RAI) and evening-hour active index (EAI) was designed to indicate hourly feature of vehicle trace. We first count the numbers of the commuter hour appearances and evening hour appearances base on trip list of the vehicle. And then the RAI and EAI of the vehicle were calculated as shown in fuction.2 and fuction.3, base on the statistical result.

$$RAI=\sqrt{1-\left(1-\frac{1}{N}\sum \text{commuterhourCount}\right)^2} \tag{2}$$

$$EAI=\sqrt{1-\left(1-\frac{1}{N}\sum eveninghourCount\right)^2} \tag{3}$$

The *commuterhourCount* is number of appearances at commuter hour, and the *eveninghourCount* is number of appearances at evening hour. *N* is number of trip in the vehicle's trip list. If the vehicle always actives at evening, the EAI would be close to 1. And the RAI would be close to 1 if the vehicle mostly running at rush hour.

The temporal feature extraction mechanism detail was shown in Algorithm 2.

. Algorithm 2: The temporal feature quantify algorithm

Input: *plateNo*: the vehicle's number-plate;
tripList: a list records the ANPR records refers to the vehicle with *plateNo*
commuterhourDict: a hash table records the hours of commuter hour time, such as 8:00am and 6:00pm;
eveninghourDict: a hash table records the hours of evening hour time, such as 2:00am and 23:00pm;
Output: *temporalFeature*: an object contains *(PFR:* presence frequency ratio*, PDC:* presence discrete coefficient, *RAI*: rush-hour active index, *EAI*: evening-hour active index*)*

```
new dateList object and do initialization with 0;
hotCount=0; quietCount=0;
for i=1;i≤ tripList.count; i++ do:
|    originRecord= tripList [i] [1]
|    tripDate= originRecord.time.date
|    tripHour= originRecord.time.hour
|    dateList[tripDate]=1
|    if commuterhourDict.has_key(tripHour) then
|    |    commuterhourCount = commuterhourCount+1
|    else if eveninghourDict.has_key(tripHour) then
|    |    eveninghourCount = eveninghourCount +1
|    end
end
F=FFT(dateList), Computes the forward Discrete Fourier Transforms and returns the coefficients F
new temporalFeature object
temporalFeature.PFR = Square (F[0])
temporalFeature.PDC= Mean ( Square (F) )
```

```
temporalFeature.RAI = Sqrt(1- Square (1-hotCount / tripList.count))
temporalFeature.EAI = Sqrt(1- Square (1-quietCount / tripList.count))
return temporalFeature
```

4. Vehicle Activity Classification

Varied spatial and temporal characteristics indicate varied vehicle activity patterns. In order to classify the patterns, this paper uses K-means clustering methods[13] to classify the vehicle groups based on spatial and temporal features of ANPR trace data, including trip origin Index(TOI), trip destination Index (TDI), presence frequency ratio (PFR), presence discrete coefficient (PDC), active time index (ATI), mentioned in section 2 and section 3. There are five typical vehicle activity patterns: tourism, commuter, public, logistics and taxi. As the result, the number of cluster k is set to 5.

The all vehicles appeared in collected ANPR data are divided into six groups as follow:

- Tourism vehicles: the vehicles from foreign area travel in this region for one or few days and would not revisit in short-term. The origins and destinations of foreign commuter vehicles are mainly Border gateway, and the most appearance time is daytimes. So the PFR and PDC values of the tourism vehicle trace would be both low.
- Commuter vehicles: The commute vehicles mainly used for commute between home and work places in working days. The origins of commuter vehicles are mainly Interior gateway, and the most appearance time is rush hours. So their PFR, PDC, ATI values would be high and the TOI value would be low.
- Public vehicle: these vehicles have a fixed route and schedule, so the active time and gateway is regular, such as, bus, cleaning vehicle and so on. The origins and destinations of public vehicle route mostly contain Border gateway. So their PFR, PDC, ATI, TOI, TDI values would be high.
- Logistics vehicles: the vehicles transport supplies and resources to this region from foreign area. And most of logistics vehicle appear during the evening for efficiency and safety. So their ATI values would be very low and the TOI and TID values would be high.
- Taxi vehicles: these vehicles keep active at any time and mostly present every day with random route. For all taxi vehicles have two drivers alternating in turn, the taxi active all day and night. So their PFR, PDC values would be high and ATI value would be low.

5. Experiment Evaluation

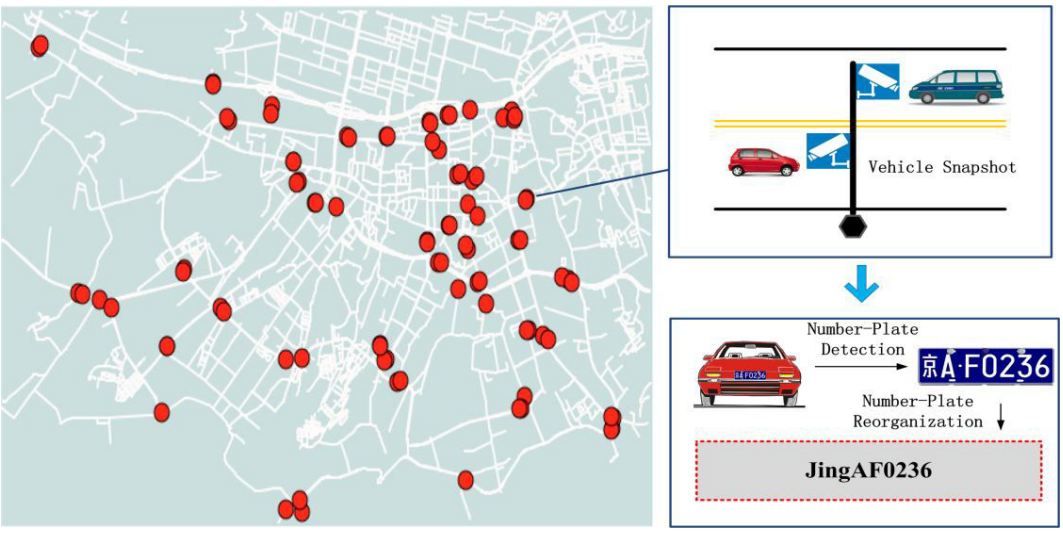

Fig.2. Automatic Number-Plate Recognition System Deployment;

We collected nearly 114 million ANPR records between August 2013 and October 2013 form a region in China. There are nearly 314 Automatic Number-Plate Recognition camera devices deployed at 112 gateways in the region as show in Fig. 2. Each camera devices capture one road direction and usually two ANPR cameras deployed at one gateway for different direction. Some gateway has four ANPR cameras for the road is wider than normal.

5.1. Date Filter and Redundant

Table 1 gives one example of collected ANPR data. ANPR cameras capture large amounts of pictures when detecting the vehicle's movement. There are many unrecognized plate numbers which are marked as "Unknown" because the correctly detection rate and the reorganization rate of plate number on the captured image could not reach 100%. And about 3% to 5% reorganization result was wrong, that is say the plate number would be recognized as another number. According to our experimental records and pictures, one vehicle would be captured 2 to 10 more times under busy traffic condition. The total amount of duplicated records could be up to more than 10 percent per day, which would interfere with data analysis or data mining.

Table 1. An example of ANPR records

Record Time	Gateway	Lane	Plate Number
2013-10-05 13:03:01	6013	2	JingVB3654
2013-10-05 13:03:01	7045	1	JingAT3421
2013-10-05 13:03:02	6013	2	JingVB3654
2013-10-05 13:03:02	7021	2	Unknown

The redundant records elimination and compression was shown in Algorithm 1. The algorithm would first filter out records of unrecognized and some sensitive plate number, such as political or military vehicles, based on the vehicle registry organization. Then the algorithm would discard those other records with same gateway and plate number within a certain threshold window. The default threshold is 60 seconds, which means if a vehicle plate number is reported many time at one gateway in 60 seconds, the algorithm would keep the first records and discard others.

As shown in Fig. 3, the filter and redundant algorithm can filter out about 15.1% of incorrect recognized ANPR records and reduce about 5.79% of redundant data on average.

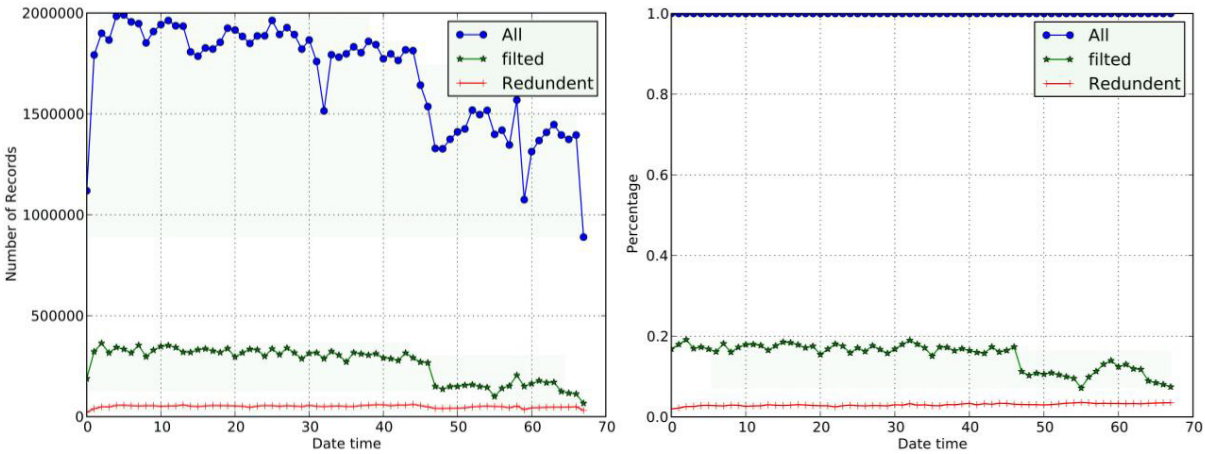

Fig.3. (a) Numbers of filtered and redundant records; (b) Percentage of filtered and redundant records

5.2. Performance of spatial feature quantify

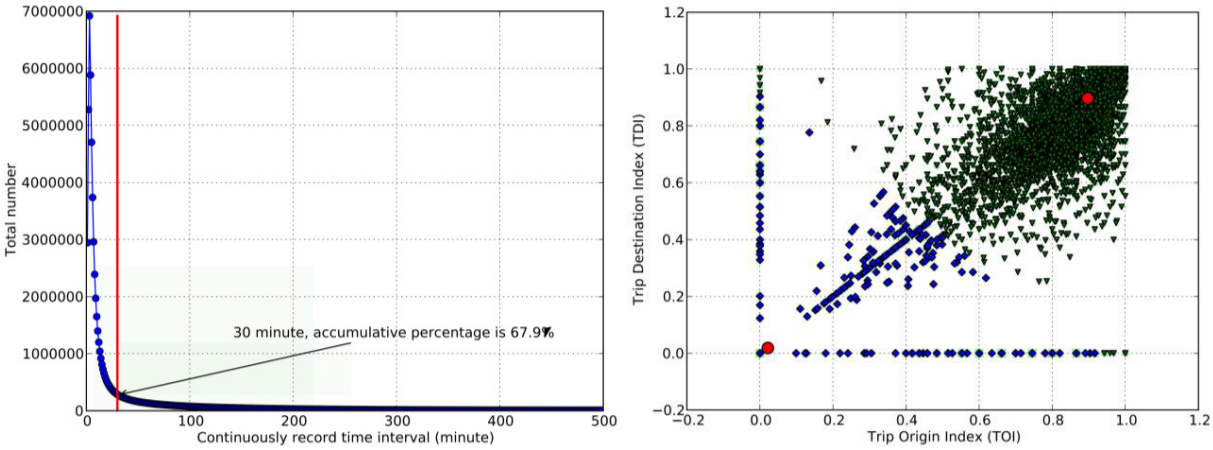

Fig.4. (a) Accumulative percentage of continuously record time interval; (b) TOI and TDI extraction result and classification

The longest path length between each deployed gateways is about 11.9km. Then the driving time of the path is about 30 minute in real driving test experiment at rush hour. As show in Fig 4.(a), the intervals of continuously records of vehicles are mainly under 1 hour. About 67.9% of interval values were less than 30 minute. According to the records interval distribution shown in Figure4.a, we set the trip time interval threshold to be 20 minute in Algorithm.2.

The TOI and TDI extraction result is show in Fig 4.(b). The TOI and TDI feature set were divided into two clustering by K-means clustering methods with k=2.

5.3. Performance of temporal feature quantify

The FFT transfer result of data-time list is show in Fig 5.(a). The PFR and PDC extraction result is show in Fig 5.(b). And the PFR and PDC feature set were also divided into three clustering by K-means clustering methods with k=3.

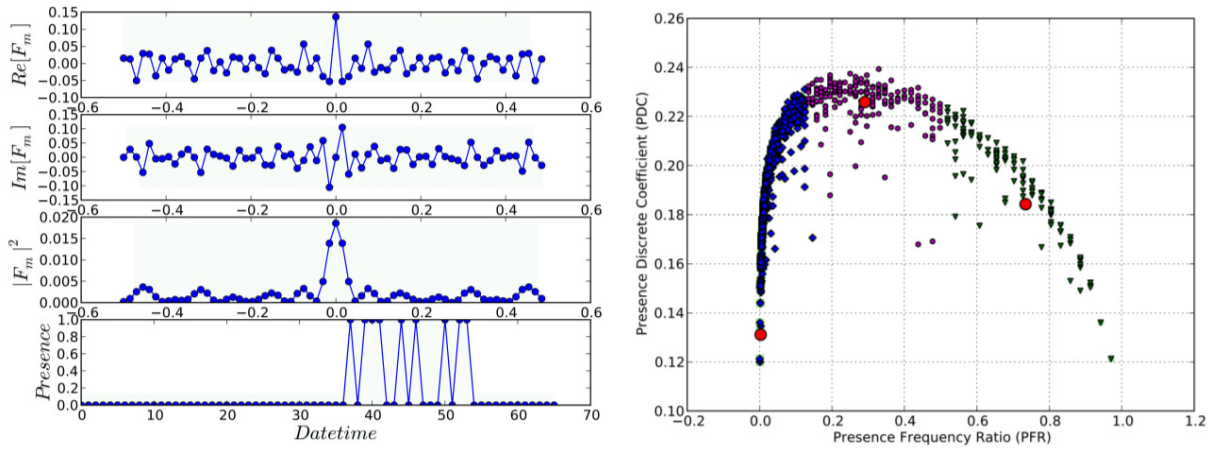

Fig. 5. (a) An example of FFT calculation; (b) PFR and PDC extraction result and classification

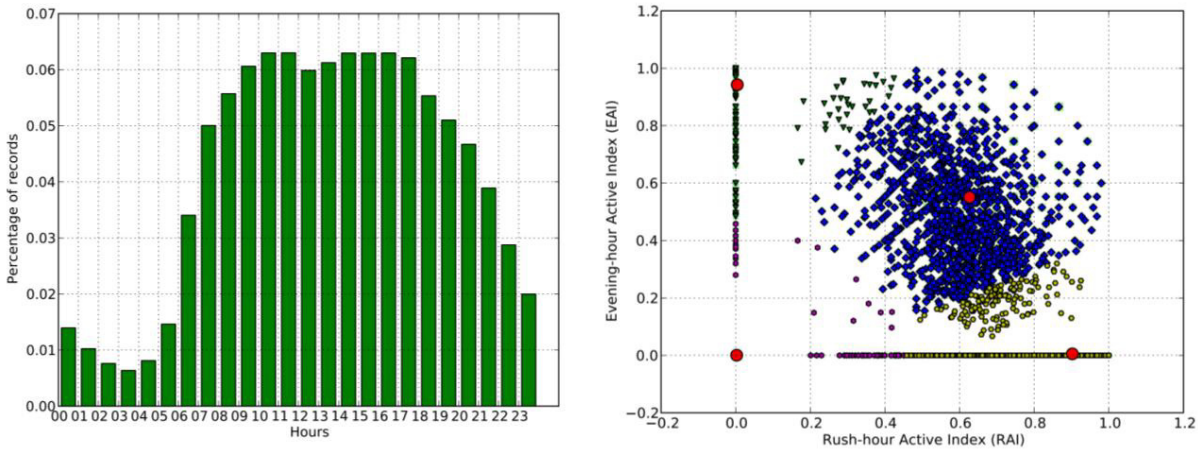

Fig. 6. (a) Percentage of ANPR records numbers at different hour time; (b) RAI and EAI extraction result and classification

According to the ANPR records numbers distribution at different hours shown in Fig 6.(a), it is easy to measure the vehicle activities or traffic at different hour. We choose the 0.am to 6.am to be the keys of *eveninghourDict* , and the 8.am to 19.pm to be the keys of *commuterhourDict* in Algorithm 2.

The RAI and EAI is shown in Fig 6.(b). The RAI and EAI feature set were divided into four clustering by K-means clustering methods with k=4.

5.4. Performance of classification

In the experiment, there were about 3,726,951 vehicles captured by the deployed ANNR system. The active pattern features of each vehicle's, that is (TOI, TDI, PFR, PDC, RAI, EAI), were computed first. Then the set of vehicles features were divided into five clustering by K-means clustering methods with k=5. We consider these clusters are more likely to be regarded as varied active pattern.

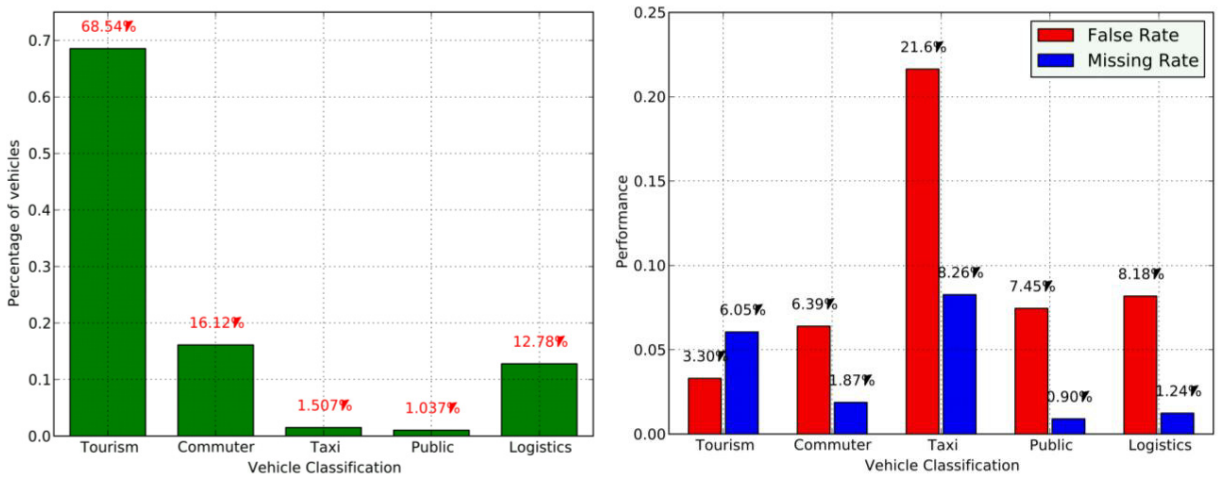

Fig. 7. (a) Number share of active pattern classification; (b) the precision of classification result

The number share of classification is shown in Fig 7.(a), the capture rate of tourism pattern is 68.54%, commuter pattern is 16.12%, taxi pattern is 1.507%, public pattern is 1.037% and logistics pattern is 12.78%.

To evaluate the performance of the classification, we randomly sampled 20,000 plate numbers as well as the captured pictures with these numbers. And the volunteers judge the vehicle type by the appearance of vehicle in the captured pictures. Then we get the precision of classification show in Fig 7.(b). The mean percentage error (MPE) is 4.76%

The false rate of taxi vehicle classification result is little high that reaches 21.6%. It might because there were so many illegal taxis without business license. So these vehicles were recognized as normal commuter car by volunteers but actually their activities were similar with taxi.

6. Conclusion

This paper proposes the vehicle trace feature quantify and activity classification issue to dig out presence patterns. By the vehicle activity pattern classification analysis, traffic management department could acquire the regional distribution of various types of vehicles and road running, to more effectively deploy resource to ensure the stable situation of the traffic. And the vehicle behavior analysis result would provide important reference information on traffic signal timing conducting, events organization developing traffic organization during construction or, trucks traffic limiting program, commuter car traffic scheduling and so on.

For future works, automatic abnormal vehicle behavior achieves detection and notification could improve the speed of monitoring and processing emergencies events and reduce the impact on traffic flow.

Acknowledgements

The work has been funded by Program of National High Technology Research and Development Program of China (2012AA050804), National Key Technology R&D Program (2012BAH20B03), S&T Innovation Base construction and development special projects of Beijing Municipal Commission of Science and Technology(Z131101002813085), and "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDA06040100).

References

1. Sivaraman S, Trivedi M M. Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis. *J IEEE Transactions on Intelligent Transportation Systems* , vol.14, no.4, Dec. 2013, p.1773-1795
2. Wiest J, Hoffken M, Kresel U, Dietmayer K. Probabilistic trajectory prediction with Gaussian mixture models. *2012 IEEE Intelligent Vehicles Symposium (IV)*, 3-7 June 2012, p.141,146
3. Yuyan S, Yan L, Xinyun Z, Limin S. City bus arrival time prediction based on similarity of road conditions. J Ruan Jian Xue Bao/Journal of Software, v 23, n SUPL, October 2012, p. 87-99
4. Otto C, Leon F P. Long-term trajectory classification and prediction of commercial vehicles for the application in advanced driver assistance systems. *American Control Conference (ACC)*, 2012, p.2904-2909
5. Barth A, Franke U. Tracking oncoming and turning vehicles at intersections. *2010 13th International IEEE Conference on intelligent Transportation Systems (ITSC)* , 19-22 Sept 2010, p. 861-868
6. Sivaraman S, Morris B. T, Trivedi M M. Learning multi-lane trajectories using vehicle-based vision. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 6-13 Nov. 2011, p. 2070-2076
7. Kasper D, Weidl G , Dang T , Breuel G. Object-oriented Bayesian networks for detection of lane change manoeuvres. *J IEEE Intelligent Transportation Systems Magazine* , vol.4, no.3, Fall 2012, p. 19-31
8. Hermes C, Einhaus J, Hahn M, Wöhler C, Kummert F. Vehicle tracking and motion prediction in complex urban scenarios. *2010 IEEE Intelligent Vehicles Symposium (IV)* , 21-24 June 2010, p. 26-33
9. Wiest J, Hoffken M, Kresel U, Dietmayer K. Probabilistic trajectory prediction with Gaussian mixture models. *2012 IEEE Intelligent Vehicles Symposium (IV)*, 3-7 June 2012, p.141-146
10. Luou S, Chenxi L, Fang Z, Weiming L. Discrete Fourier Transformation for Seasonal-Factor Pattern Classification and Assignment. *J IEEE Transactions on Intelligent Transportation Systems*, vol.14, no.2, June 2013, p.511-516
11. Homayounfar A, Ho A T S, Zhu N, Head G, Palmer P. Multi-vehicle convoy analysis based on ANPR data. *4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011)*, 3-4 Nov. 2011, p. 1-5.

12. W. S. Steven. *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. San Diego, CA: Spectrum, 2002.
13. Macqueen J. Some Methods for classification and Analysis of Multivariate Observations. *J Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. California, USA, 1967, p. 281–297.