http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (19): 4858-4863, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4858.4863

Method for High-way States Analysis Based on Clustering Algorithm

Lin He, WeiChuan Hong and Yan Chen
Transportation Management College, Dalian Maritime University, Liao Ning,
Da Lian, 116026, China

Abstract: Since the traffic flows are complicated and unstable, there is no standard to classify the traffic flows around the management of traffic, which causes the obstacle to the managers. The purpose of this study is to use flow, velocity, occupancy as input parameters and build up a traffic state classification model based on clustering algorithm. Furthermore, based on the traffic flow theory, this study presents a new method to identify the initial center in clustering in order to avoid the traditional flaws and improve the efficiency in clustering algorithm. Finally, the study utilizes samples to validate the differences and improvement of modified K-means model and modified FCM model. The results prove that modified FCM model is more suitable for the need in traffic management. This model is able to give the exact definition of traffic states, which may discriminate congestion state of high-way and support management of traffic.

Key words: Traffic flow classifying, clustering algorithm, initial centers optimization

INTRODUCTION

Accurate identification of thetraffic statecaneffectively guide thetraffic management and traffic dispersion. In order to achieve this aim, most cities around the world have built their traffic surveillance and control system already, which is used to acquire traffic flowparameters of traffic state. In presents, the representative traffic control systemofincluding TRANSYT system in United States, SCOOT system in Britain and SCATS system in Australia (Xu and Fu, 2009). However, it is still impossible to give a unifiedstandard to identify the traffic state of high-way and quantity of state evaluation. Different people hold different opinion (Hall *et al*., 1992) believed there are three state in traffic flow, including non-congestion, congestion and dispersion state. In 1995, Francois and Willis (1995) proposed that the traffic state should be quantified. In the meantime, he gave severalquantitative indexes to define traffic state. Through the Germanexperimentof expressway traffic flowdata, Kernerand Rehborn (1996) divided traffic state in to three stages, including a free flow, synchronized flow and blocked flow, while Wu (2002) divided them into four stages according to the cars' followingbehaviorand relative position. The free states include free flowing traffic and restrained flowing traffic, while the congestion state include restrained congestion traffic and persistent congestion traffic. There are also different methods to acquire parameters. Qu *et al*. (2013) utilized parameters obtained by fixed monitors to identify traffic state and built up a network traffic stateidentificationmodel which focused on study the relation between velocity and state. Guo *et al*. (2012) built up the threeearly warning mechanisms by velocity parameters caught by the video detection technology. (Luo, 2012) even figured out that there is some relation between noise and identifying traffic state. He believed the traffic state can be identified by noise and this method possess high accuracy. Some scholars tried to use clustering method in data mining to identify traffic state. Xu *et al*. (2012) in Dongnan University used K-means to identify the traffic state and successfully divided them into five categories. However, (Chen *et al*., 2010) figured out it is not accurate to use FCM to identify traffic state without manual intervention. In sum, using theclustering method of traffic state classification becomes the focus of research. Nonetheless, lots of factors, such as the application ofclusteringmethod, traffic flowparameters selection, selection ofinitial clustercenters, will affectthe accuracy of classification.

This essay uses clustering algorithm to identify traffic state. It uses several traffic parameters to build up a traffic identifying model based on clustering algorithm. Traffic flow, velocity and occupancy are normalized and utilized in order to avoid the inaccuracy of identification based on one traffic parameter. In the meantime, the method of getting initial centers is optimized to improve the efficiency of this model. In the end, through the real circumstance, the essay compares the different between K-means and FCM.

Corresponding Author: Lin He, Transportation Management College, Dalian Maritime University, Liao Ning, Da Lian, 116026, China

MODEL BUILDING

Problem description: Traffic parameters, such as velocity, traffic flow, occupancy, are the quantified description of traffic state. The essence of identifying traffic state is the process of classifying traffic objects. While clustering analysis, one of the most method in data mining, is able to classify the data objects through analyze their characteristics, this essay use this analysis to classify traffic state. Through the clustering analysis, one can obtain the exact definition of different traffic state. The clustering centers can be regarded as the quantified definition of each traffic state. One of the most significant advantages of this method is that it can adapt to complex traffic situation.

The basic definition of the model is as follows:

- **Input:** Trafficflowsat asection of road (included velocity, traffic flow, occupancy)
- **Output:** Quantified definition of each traffic state, each traffic membership to different state
- **Clustering model:** K-means, FCM

Model assumption: In order to simplify complex question, one can make several assumption as follows.

Assuming one way traffic can be divided into three categories: Free, Congestion, High-flow. The characteristics of Free are low flow, high velocity and low occupancy. The characteristics of Congestion are low flow, low velocity, high occupancy. The characteristics of High-flow are high-flow

Data listed follows are regarded as unreasonable data, which are eliminated for the inaccuracy:

- Average occupancy is higher than 100%
- Average velocity is higher than 150 km h^{-1}
- When the flow is lower than 0 cars h^{-1}, the velocity is higher than 0 km h^{-1} or the occupancy is higher than 0%
- When the occupancy equals 0%, the flow is higher than 0 cars h^{-1}
- Occupancy, velocity and flow were all 0 (There is no cars pass by)

Model Improvement

Optimizationof the initial center: The above mentioned selected initial points directly influencethequality and efficiency ofclustering method. If the method randomly selects the initial centers, the result would easily fall into partial optimal and increasethe number of iterations, which may make the clustering less effective. Therefore, we should optimize selecting initial centers of clustering methods. Because most of the optimizations of selecting initial centers are based on the algorithm, which make the clustering much more difficult. Besides, some of these methods are not effective when the data come from traffic flows. In sum, the optimization based on algorithm has two main disadvantages: (1) Complex algorithm, which causes bad efficiency; (2) The quality of method is not ideal, because it don't base on the traffic characteristics.

According to assumption, this essay optimizes the initial centers as follows:

- **Input**

High-flow value: Q_H; High-velocity value: V_H; High-occupancy value: P_H

Low-flow value: Q_L; Low-velocity value: V_L; Low-occupancy value: P_L

- **Output:**

Free: $\{Q_{X1}, V_{X1}, P_{X1}\}$

Congestion: $\{Q_{X2}, V_{X2}, P_{X2}\}$

High-flow: $\{Q_{X3}, V_{X3}, P_{X3}\}$

Specific processes as follows:

- **One-way trafficstate is divided into 3 states, set 3objects sets:** Free X1, Congestion X2, High-flow X3. Iterating all the velocity, flow, occupancy of the traffic objects, if the object satisfies the condition below, put it into the objects set.

Free X1: $Q_{free}<Q_L$, $V_{free}>V_H$, $P_{free}<P_L$

Congestion X2: $Q_{congestion}<Q_L$: $V_{congestion}<V_L$: $P_{congestion}>P_H$

High-flow X3: $Q_{high}>Q_H$

- **Calculate average value of each objects set:** Free:

$$Q_{X1}=(Q_L+Q_1+Q_2+\ldots+Qn)/n+1$$

$$V_{X1}=(V_H+V_1+V_2+\ldots+Vn)/n+1$$

$$P_{X1}=(P_H+P_1+P_2+\ldots+Pn)/n+1$$

Congestion:

$$Q_{X2} = (Q_L+Q_1+Q_2+\ldots+Qn)/n+1$$

$$V_{X2} = (V_L+V_1+V_2+\ldots+Vn)/n+1$$

$$P_{X2} = (P_H+P_1+P_2+\ldots+Pn)/n+1;$$

High-flow:

$$Q_{X3} = (Q_L+Q_1+Q_2+\ldots+Qn)/n+1$$

$$V_{X3} = ((V_H+V_L)/2+V_1+V_2+\ldots+Vn)/n+1$$

$$P_{X3} = ((P_H+P_L)/2+P_1+P_2+\ldots+Pn)/n+1$$

- **Initial centers are as follows:**

$$\text{Free: } \{Q_{X1}, V_{X1}, P_{X1}\}$$

$$\text{Congestion: } \{Q_{X2}, V_{X2}, P_{X2}\}$$

$$\text{High-flow: } \{Q_{X3}, V_{X3}, P_{X3}\}$$

When calculating the average value, it is found that there are several problems if just regard all the qualified objects' average value as the initial centers:

- Average values are not ideal when compares to Professional's ideal initial centers, which causes a little deviation
- Divisor might be zero during the clustering algorithm, especially for FCM. The situation is as follows:

$$\mu_{ik}^{(b)} = \{\sum_{j=1}^{c}[(\frac{d_{ik}^{(b)}}{d_{jk}^{(b)}})^{\frac{2}{m-1}}]\}^{-1} \quad (1)$$

If there is only one chosen object in the objects set, when we calculate the initial centers by the Eq. 1, d_{ik} would be zero, which may cause the infinite result.

Therefore, in order to avoid this circumstance, we should add the limited condition's object into the objects set before calculating.

During the calculation, high or low flow, velocity and occupancy should be set rigorously for cutting down the iterations and improve the velocity of calculating. This method supports that there is no object in each objects set. In sum, this method no only satisfies the requests of professionals' definition in traffic states but also combines the real traffic objects, which cut down the subjectivelyinfluence of professional.

Correctionand normalizationof theoriginal sampledata: Because the magnitude of flow is much higher than the magnitude of velocity and occupancy, during the Euclidean distance calculating, the result of flow is higher than result of velocity and occupancy, which causes it domains the classification of initial centers. Therefore, the result of initial centers has flow with big difference and velocity, occupancy with small difference. In order to avoid this circumstance, we need to preprocess the traffic objects and make sure that their flow, velocity and occupancy are at the same magnitude.

The preprocessing is as follows:

- Find the highest and lowest flow, velocity, occupancy in the objects as HighQ, HighV, HighP, LowQ, LowV, LowP
  Re-calculate the parameters in each object as follows:

$$X_q = \frac{X_q\text{-lowQ}}{\text{highQ-lowQ}}*100\% \quad (2)$$

$$X_v = \frac{X_v\text{-lowV}}{\text{highV-lowV}}*100\% \quad (3)$$

$$X_p = \frac{X_p\text{-lowP}}{\text{highP-lowP}}*100\% \quad (4)$$

Through this method, we can ensure that all the parameters are at the same magnitude, which can avoid the inaccuracy of clustering because of different magnitude. Of course, after clustering, we should use the opposite method to restore the parameters.

Model procedure: According to the analysis upon, we can build up a clustering model of traffic states as Fig. 1.

- **Input:** Time (years, months, days), traffic flow direction (up, down)
- **Output:** Clustering membership matrix, iteration, clustering centers

EXPERIMENTAL RESULTS

Calculation procedure: Take the traffic flows from January 1st to December 31th, 2010 as an example. We use the clustering method to define the traffic states and compare the differences between K-means and FCM. Sample data is showed as Fig. 2.

- Input the high flow 800, low flow 200, high velocity 90, low velocity, 40, high occupancy 50, low occupancy 5, which can be used to calculate the initial centers

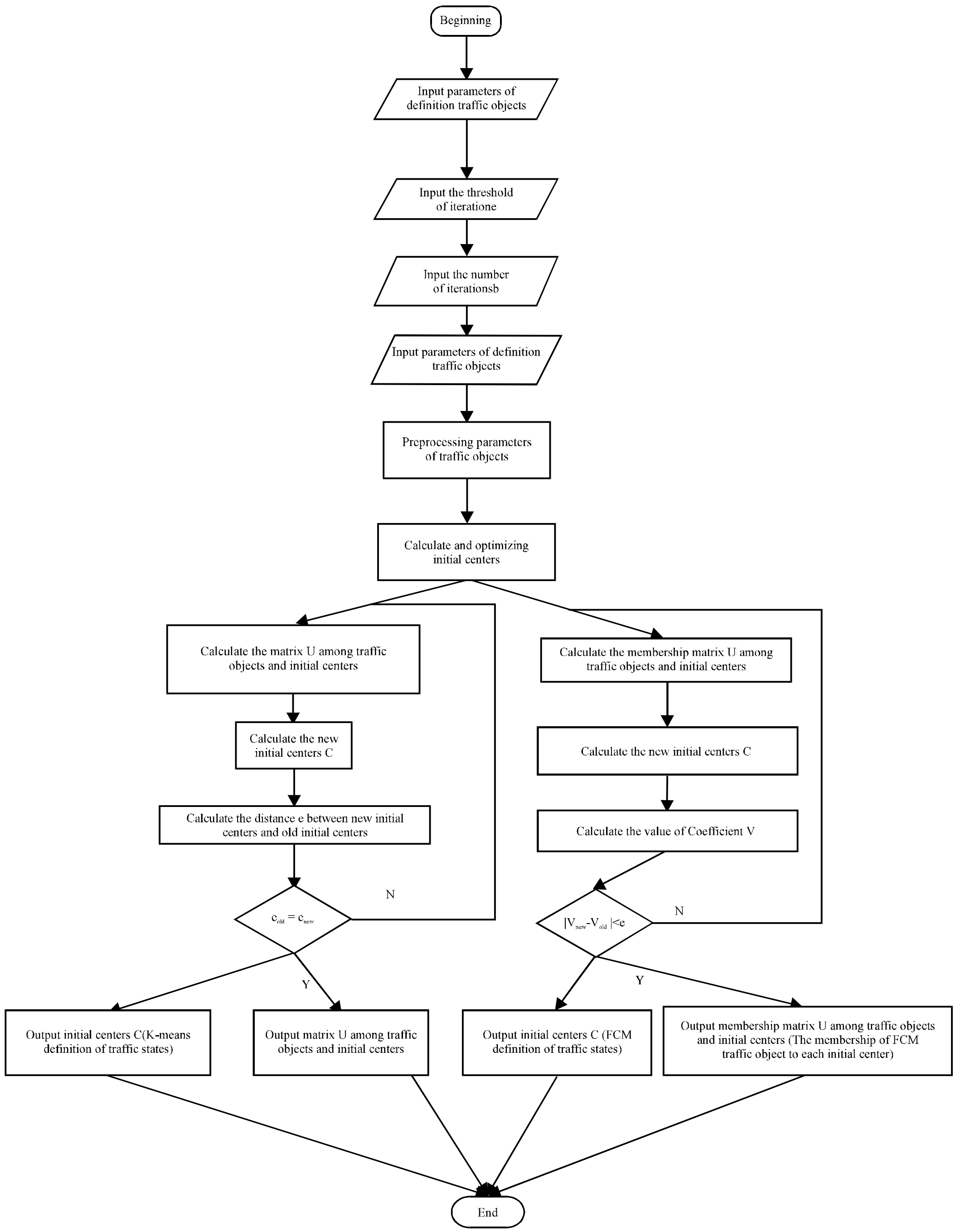


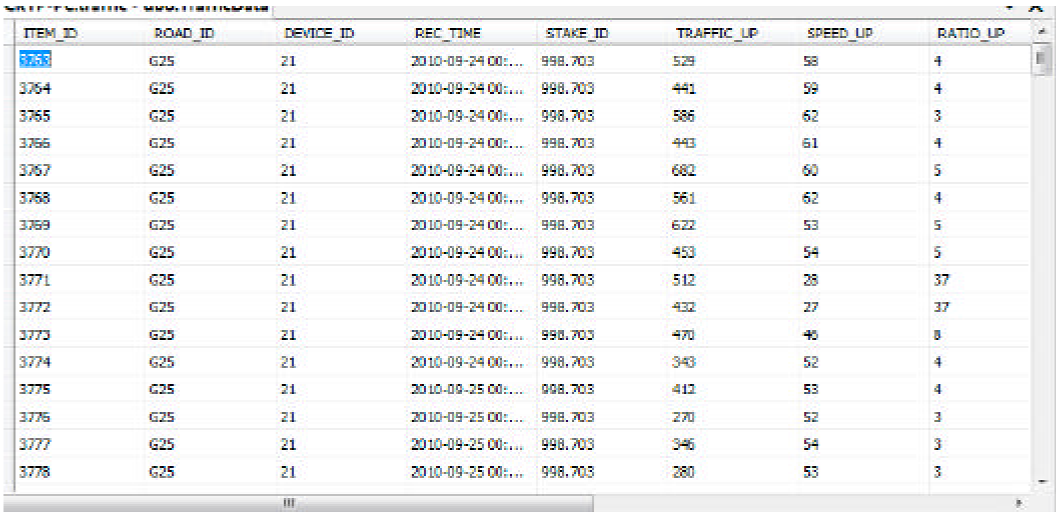
Fig. 1: Classification of traffic state based on clustering model

- Input clustering threshold 0.3
- Processing K-means, the results are as Table 1 and Table 2
- Processing FCM, the results are as Table 3 and 4

Result analysis: Under the sametraffic conditions, through the comparison of K-means and FCM we can find:

Clustering centers: There is no obvious difference between the K-means and FCM's clustering centers. They are about (268, 68, 2.7), (430, 30, 38) (525, 72.5, 4), which means there is little different between K-means and FCM in definition of traffic states.

States matrix: After comparing the matrix between K-means and FCM, we can find that in FCM the highest



ITEM_ID	ROAD_ID	DEVICE_ID	REC_TIME	STAKE_ID	TRAFFIC_UP	SPEED_UP	RATIO_UP
3763	G25	21	2010-09-24 00:...	998.703	529	58	4
3764	G25	21	2010-09-24 00:...	998.703	441	59	4
3765	G25	21	2010-09-24 00:...	998.703	586	62	3
3766	G25	21	2010-09-24 00:...	998.703	443	61	4
3767	G25	21	2010-09-24 00:...	998.703	682	60	5
3768	G25	21	2010-09-24 00:...	998.703	561	62	4
3769	G25	21	2010-09-24 00:...	998.703	622	53	5
3770	G25	21	2010-09-24 00:...	998.703	453	54	5
3771	G25	21	2010-09-24 00:...	998.703	512	28	37
3772	G25	21	2010-09-24 00:...	998.703	432	27	37
3773	G25	21	2010-09-24 00:...	998.703	470	46	8
3774	G25	21	2010-09-24 00:...	998.703	343	52	4
3775	G25	21	2010-09-25 00:...	998.703	412	53	4
3776	G25	21	2010-09-25 00:...	998.703	270	52	3
3777	G25	21	2010-09-25 00:...	998.703	346	54	3
3778	G25	21	2010-09-25 00:...	998.703	280	53	3

Fig. 2: Sample data

Table 1: States matrix of k-means

Traffic ID	State 1	State 2	State 3
...			
7	0	0	1
8	0	0	1
9	0	1	1
10	0	1	0
11	0	0	1
12	1	0	0
13	1	0	0
...	...	...	...
76	0	0	1
...			

Table 2: K-means states definition centers

Traffic state	Up flow	Up velocity	Up occupancy
1	269.25766	67.69459	2.7549715
2	429.94778	28.68667	39.4360430
3	532.81540	72.40531	4.2607765

Table 3: Membership states matrix of FCM

Traffic ID	State 1	State 2	State 3
...			
7	0.1804201000	0.15061142	0.6689685000
8	0.3274119200	0.11935783	0.5532302000
9	0.0204692780	0.95525060	0.0242800750
10	0.0030799909	0.99377090	0.0031491057
11	0.3048190500	0.24710323	0.4480777000
12	0.6155373500	0.11410562	0.2703570000
13	0.4403583400	0.12223884	0.4374029000
...	...	...	...
76	0.2003154800	0.40245354	0.3972309800
...	...	...	...

Table 4 FCM States Definition Centers

Traffic state	Up flow	Up velocity	Up occupancy
1	268.80096	69.54950	2.6615498
2	434.74185	30.38108	37.3452870
3	520.93170	73.01341	4.0739164

smembership object belong to the same state in K-means, which means in clustering, K-means and FCM are similar in traffic objects' classification.

However, the objects in K-means are forced to be divided in each traffic states, while in FCM each object has their own membership to each traffic state. According to the real environment, a splendid clustering method should not only provide that whether or not an object is congest but also provide that the degree of congestion of each object or the membership to each traffic states. For example, in table 1 and table 3, according to traffic flow number 76's parameters, we can find it is prone to congestion but it is not congestion. As a traffic state manager, we should pay more attention to this traffic object. However, K-means divides it to the High-flow state. Compare to K-means, FCM give the membership of this state, which describe this state much more clear than K-means and warn the traffic manager. Therefore, from a practicalpoint of view, FCM can describe the traffic objects more precisely than K-means and help the traffic manager to handle the traffic flows.

CONCLUSION

Through clustering analysis, we can acquire the potential relation between traffic objects and traffic states. This essay successfully uses the real traffic objects to divide the states in high-way and give the clustering centers as the definition of each traffic states. By this definition, traffic manager can precisely describe the traffic objects according to FCM matrix. However, we need to notice that while processing clustering, the clustering algorithm will categorize all the traffic objects according to the initial centers. Therefore, if there is no traffic object in some traffic states, the clustering algorithm will still categorize some objects to these states to ensure the integrity of clustering. So, when we use the clustering to define traffic states, we should ensure that all the traffic states have at least one traffic object.

ACKNOWLEDGMENTS

Supported by "The Ministry of education of Humanities and Social Science Fund for young" (No. 13YJCZH052). Supported by "NSFC" (Natural Science Foundation of China) (No. 70940008). Supported by " the Fundamental Research Funds for the Central Universities" (No. 3132013074).

REFERENCES

Chen, F., Y.H. Jia and Z.H. Niu, 2010. Classification of traffic flow based on fuzzy c-means clustering. Proceedings of 2010 Asia-Pacific Conference on Information Theory, December 11-12, 2010, Chinese Institute of Electronics, Information Theory Branch, XiAn China, pp: 1100-1103.

Hall, F.L., V.F. Hurdle and J.H. Banks, 1992. Synthesis of recent work on the nature of speed flow and flow-occupancy (or density) relationships on freeways. Transp. Res. Rec., 1365: 12-18.

Francois, M.I. and A. Willis, 1995. Developing effective congestion management systems. Federal Highway Administration, Technical Report No.8, p: 22.

Guo, Y.R., B.T. Dong and L. Wu, 2012. Traffic state recognition and dynamic evaluation based on velocity study. J. Highway Transp. Res. Dev., 29: 26-31.

Kerner, B.S. and H. Rehborn, 1996. Experimental properties of complexity in traffic flow. 53 4275: 4278-10.1103/PhysRevE.53.R4275.

Luo, X.L., 2012. Traffic flow state identification based on traffic noise signals. J. Tongji Univ. (Nat. Sci.)., 40: 1821-1824.

Qu, Z.W., Q. Wei, Y.M. Bie, H. Zhu and D.H. Wang, 2013. Method for traffic state identification based on fixed detector. J. Central South Univ. (Sci. Technol.), 44: 403-410.

Xu, L. and H. Fu, 2009. Intelligent Prediction Theory and Methods of Traffic Information. Science Press, Beijing.

Xu, C., P. Liu, W. Wang and Z. Li, 2012. Evaluation of the impact of traffic states on crash risks on freeways. Accid. Anal. Prev., 47: 162-171.

Wu, Z.F., 2002. Ecoclimatic study on vegetation transition zones in Northeast China. Scientia Geographica Sinica, 22 (2): 219-225, (In Chinese).