ESS2019
2019/9/6

Traffic Congestion Detection from Taxi Trajectory Data

Thi Thi Shein[1] and Makoto Imamura[2]

Due to the advance of location acquisition technologies, the large amount of trajectory data can be collected in the form of spatial-temporal data. Extraction meaningful knowledge from the trajectory data is a popular topic in the data mining area. Thus, this paper aims to detect traffic congestion in the form of gathering pattern. Thus, the gathering pattern defined as the congregation of vehicles, which stay together with high density in some area for a long duration. To retrieve the gathering pattern efficiently, we propose an incremental gathering pattern discovery framework over the taxi trajectory data. Effectiveness and efficiency are conducted on real taxi data and synthetic data.

1. Introduction

As the rapid development in location tracking devices, such as Global Positioning System (GPS), Radio Frequency Identification Devices (RFID), and elementary attached on the wildlife, spatial-temporal trajectories of moving objects has become increasingly available. Extraction knowledge from the trajectory is a popular research trend in the area of data mining. In this paper, we developed the system efficiently to detect traffic jam, namely gathering pattern.

The gathering pattern was introduced by Zheng et al.[1]. The gathering pattern is detected by considering the following properties:1) **Size**: A gathering must contain a large number of objects, 2) **Density**: Those objects must be in the form of dense group, 3) **Duration**: A group of these objects must be last for a certain time period, 4) **Stability**: A group of these objects must be relatively stable. 5) **Member**: A group of these objects must contain some dedicated members that last for a certain time period.

2. Our Approach

In our approach, there have three main phases: 1) clustering 2) crowd detection 3) gathering retrieving. The first phase, clustering, is to detect the group of objects at each timestamp. In this paper, we used the micro-group based clustering that we proposed in [2]. Crowd detection is a group of these objects that last of certain time periods. From these crowds, we retrieve the gathering in the form of traffic jam. Fig.1 shows the framework of our approach.

Fig.2 illustrates an example of gathering pattern discovery from the moving object trajectories. If we set a candidate size threshold (δ_s) =3, a candidate duration threshold(δ_t) =3, then the system detects the cluster sequence $\langle c_1^1, c_2^1, c_3^2\rangle$, $\langle c_1^2, c_2^2, c_3^2\rangle$ as crowd patterns. Since the cluster, c_3^1, is too far away from the clusters c_2^1 and c_2^2, it is not the member of the crowd patterns. If we add participator size threshold (ρ_s)=3, participator lifetime threshold (ρ_t)=2, then the system results only $\langle c_1^2, c_2^2, c_3^2\rangle$ as a gathering pattern because it contains four participators$\{o_5, o_6, o_7, o_8\}$ along the time.

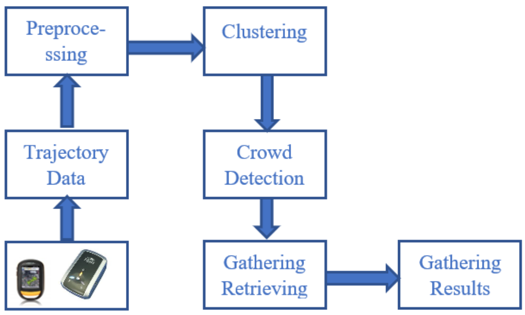

Fig. 1. Framework for gathering pattern discovery

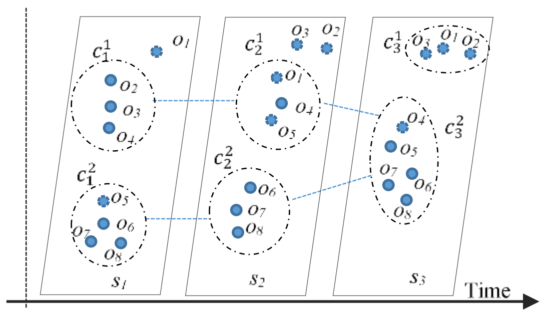

Fig. 2. Illustration of gathering pattern discovery

[1, 2] Department of Embedded Technology, Tokai University.

3. Experiment

In this section, we conduct the effectiveness and efficiency comparison on real (T-drive)[3] and synthetic [4] datasets. The improvement of our approach: gathering discovery with micro-group based clustering algorithm (MG) was compared with the state of the art pattern discovery methods: 1) gathering pattern discovery with buddy based clustering algorithm (BG) [1], 2) convoy discovery (CD) [5], swarm pattern discovery (SW) [6], and 4) platoon pattern discovery (PT) [7]. Fig.3 describes the datasets and parameter settings for experiment evaluation.

Dataset	Obj#	Points#	Duration	Frequency
T-Drive (D1)	1164	312,936	36 hours	432 (5min/ snapshot)
Synthetic (D2)	1586	231,201	24 hours	1440 (1 min/snapshot)
Candidate size threshold (δ_s):10-30; default:15				
Candidate duration threshold (δ_t): 10-30; default:15				
Candidate distance threshold (δ_d): 300 meters				
Participator size threshold(ρ_s) : $\delta_s/2$				
Participator lifetime threshold(ρ_t) : $\delta_t/2$				
Micro-group distance threshold (ε): 100 meters for D1, D2;				
Micro-group size threshold (γ): 3				

Fig. 3. Parameter Settings and Datasets

3.1. Effectiveness

We evaluated the number of patterns (as effectiveness) on real taxi data: T-drive (D_1).In this evaluation, we divided a day into three periods, peak time (7 am to 10 am and 4 pm to 8 pm), work time (10 am to 4 pm), and casual time (8 pm to 5 am).

Fig.4 shows the effectiveness evaluation based on the setting of the default parameter.

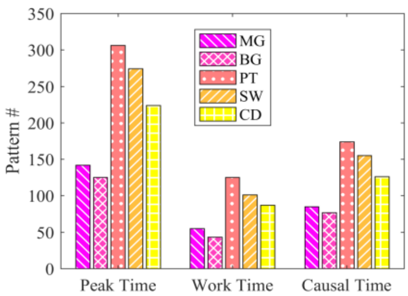

Fig. 4. Effectiveness on T-drive (D1)

3.2. Efficiency

Fig.5 shows the comparison for the running time of all pattern discovery methods based on the default parameter settings.

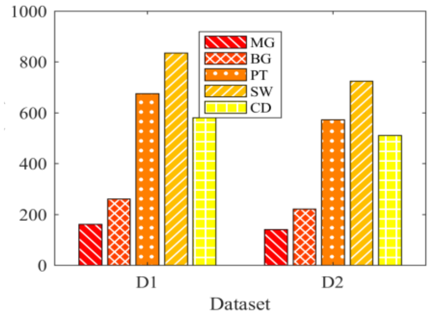

Fig. 5. Efficiency on all datasets

References

[1] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, “Online discovery of gathering patterns over trajectories,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 8, pp. 1974–1988, 2014.

[2] S. Puntheeranurak, T. T. Shein, and M. Imamura, “Efficient Discovery of Traveling Companion from Evolving Trajectory Data Stream,” 2018 IEEE 42nd Annu. Comput. Softw. Appl. Conf., pp. 448–453, 2018.

[3] J. Yuan, Y. Zheng, C. Zhang, and W. Xie, “T-Drive: Driving Directions Based on Taxi Trajectories,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2, no. 3, Elsevier B.V., 2010, p. 593.

[4] M. F. Mokbel et al., “MNTG: An extensible web-based traffic generator,” Lect. Notes Comput. Sci. including Subser. Lect. Notes Artif.

[5] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen, “Discovery of Convoys in Trajectory Databases,” 2010.

[6] Z. Li, B. Ding, J. Han, and R. Kays, “Swarm: Mining relaxed temporal moving object clusters,” VLDB Int. Conf. Very Large Data Bases, vol. 3, no. 1–2, pp. 723–734, 2010

[7] Y. Li, J. Bailey, and L. Kulik, “Efficient mining of platoon patterns in trajectory databases,” Data Knowl. Eng., vol. 100, no. Yuxuan Li, pp. 167–187, 2015.