2019 IEEE Pacific Visualization Symposium (PacificVis)

Visual Analytics of Taxi Trajectory Data via Topical Sub-trajectories

Sichen Jin* Yubo Tao† Yuyu Yan‡ Jin Xu§ Hai Lin¶

State Key Lab of CAD&CG, Zhejiang University

Abstract

GPS-based taxi trajectories contain valuable knowledge about movement behaviors for transportation and urban planning. Topic modeling is an effective tool to extract semantic information from taxi trajectories. However, previous methods generally ignore the direction of trajectories. In this paper, we employ the bigram topic model instead of traditional topic models to analyze textualized trajectories to take into account the direction information of trajectories. We further propose a modified Apriori algorithm to extract frequent sub-trajectories and use them to represent each topic as topical sub-trajectories. Finally, we design a visual analytics system with several linked views to facilitate users to interactively explore topics, sub-trajectories, and trips. We demonstrate the effectiveness of our system via case studies with Chengdu taxi trajectory data.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics

1 Introduction

The advanced location-acquisition techniques nowadays have generated a variety of trajectories of humans and vehicles. For example, GPS devices on taxis can record the movement information in real-time, such as the time, geographical location (longitude and latitude), and occupied/vacant status. Thus, large amounts of trajectories are generated which contain valuable movement patterns and can be used to improve urban traffic [18]. However, how to extract and visualize these patterns more effectively for better understanding is still a challenge.

Topic modeling is an effective method to extract semantic topics/patterns from texts and has been applied to taxi trajectories. Chu et al. [7] employed latent Dirichlet allocation (LDA) to extract latent topics from taxi trajectories by mapping the taxi location to the road name, and designed topic maps and routes to visualize the probability-based topical information. Similarly, Tang et al. [13] divided the urban area into grids, encoded the grid id and the time information into a word, and used LDA to extract topics. These works focus on dividing the urban area into several regions represented by a set of road names. But when we focus on the movement understanding of trajectories, the direction of a trajectory, e.g., the order of roads, is also an important property. However, LDA is inferred under the "bag-of-words" assumption, in which the word order is ignored.

In order to utilize the direction information, this paper adopts the bigram topic model to extract latent topics from trajectories by transforming each trajectory into a sequence of road names (Sec. 4.2). Besides, we present the semantic information of topics by frequent sub-trajectories extracted via a modified Apriori algorithm (Sec. 4.3), since sub-trajectories (road names sequences) are more meaningful and understandable than single road name or a region with many roads in previous methods. For example, 'Qingjiang East Road→Chengwen Elevated Road' contains more meaning than the set of 'Qingjiang East Road' and 'Chengwen Elevated Road". Finally, we develop a visual analytics system to present topics, sub-trajectories, and trips for interactive exploration (Sec. 5). The effectiveness of our system is evaluated via case studies of Chengdu taxi trajectory data.

*e-mail: 3150104297@zju.edu.cn
†e-mail: taoyubo@cad.zju.edu.cn (corresponding author)
‡e-mail: yanyuyu001@gmail.com
§e-mail: jinxu.zju@gmail.com
¶e-mail: lin@cad.zju.edu.cn

2 Related Work

Our work is related to trajectory pattern mining [19]. One standard solution is to group similar trajectories into clusters to find common trends of trajectories. Ester et al. [8] modified DBSCAN to group trajectories by the density of their segments in the space. Kim et al. [9] presented a latent topic-based clustering algorithm to discover patterns in the trajectories of geo-tagged text messages. For sequential pattern mining of trajectories, Cao et al. [5] transformed the original sequence into a list of sequence segments, and extracted patterns based on a substring tree structure and the improved Apriori algorithm. In this paper, we combine sequential pattern mining with topic modeling and propose topical sub-trajectories to discover patterns from taxi trajectories.

Visualization has been increasingly used to support the understanding and analysis of trajectory data [4] [6]. To reduce visual clutter of numerous trajectories, Andrienko and Adrienko [1] proposed spatial generalization and aggregation of trajectories for creating legible flow maps. Liu et al. [10] developed several visualization techniques to analyze route diversity patterns from taxi trajectories. Pu et al. [12] presented an interactive visual analytics system, T-Watcher, for monitoring and analyzing complex traffic situations in big cities based on taxi trajectories.

Latent topic extraction has been widely used for visual analytics of taxi trajectories. Chu et al. [7] transformed the taxi trajectory into a document consisting of the taxi traversed road names and applied LDA to extract semantic topics to show the probability-based maps and routes. Al-Dohuki et al. [3] further mapped GPS points into road/POI names to support flexible semantic queries over a text search engine. Wu et al. [17] utilized non-negative matrix factorization (NMF) to extract latent activity patterns from three types of heterogeneous mobility data (taxi trajectories, metro passenger RFID card data, and telco data). Tang et al. [13] divided the urban area into grids, encoded the grid id and the time information into a word, and applied LDA to extract latent topics of taxi trajectories. Different from previous works, we adopt the bigram topic model rather than traditional topic models, such as LDA, to include the direction information of trajectories in the topics. Besides, while previous works usually represent topics by top scored 'word' (road name), we represent topics by top scored 'phrases' (sub-trajectories) to facilitate understanding of the semantic information in each topic.

3 Overview

Base on the discussion with transportation engineers and urban planning professionals, we found that the direction of each trajectory is important in decision-making such as figuring out the specific direction blocked. So our system aims to analyze the movement patterns in the taxi trajectories with the direction attribute included. Unlike previous works which use a group of road names

2165-8773/19/$31.00 ©2019 IEEE
DOI 10.1109/PacificVis.2019.00027

to represent a topic, we employ sub-trajectories (road names sequences) to represent a topic. Thus, the direction information is included in the order of road name sequences. In addition, to help users better manage the current urban traffic and make future urban planning, we visualize the extracted topics and sub-trajectories for interactive and visual exploration. According to the requirements, we describe the following tasks for our system:

T1: Trajectory topic construction
A suitable model is required to extract topics/patterns from taxi trajectories. The model should be based on both the road names and directions of trajectories.

T2: Topical sub-trajectory extraction
To better present the meaning of topics, we need to extract frequent sub-trajectories (defined in Section 4.3) from taxi trajectories. These frequent sub-trajectories are further associated with topics as topical sub-trajectories.

T3: Trajectory pattern understanding
To facilitate users understand the latent semantic information, we visualize the extracted topics and their associated sub-trajectories from the spatial, temporal, and other dimensions with linked views.

Based on the above tasks, we propose our framework in Fig. 1. We first clean the data and textualize the trajectories into documents with sequences of road names. We then employ a bigram topic model to generate latent topics/patterns from sub-trajectories (T1) and propose a modified Apriori algorithm to extract frequent sub-trajectories from textualized documents (T2). Finally, we visualize topics and topical sub-trajectories via the map, stacked bar chart, circular heatmap, etc. (T3).

4 Pattern Extraction based on Sub-trajectories

This section describes how to preprocess taxi trajectories, construct patterns via topic modeling, and extract frequent sub-trajectories.

4.1 Trajectory Preprocessing

The taxi trajectory data used in this paper contains 1.4 billion trajectories of 14 thousand taxis in 6:00 - 24:00 from August 03, 2014 to August 30 in Chengdu. Each record contains the following fields: taxi ID, latitude and longitude, time, and occupied/vacant status indicating whether the taxi has a customer.

We first segment the trajectory of a taxi into several trips according to the occupied/vacant status, and each trip corresponds to one taxi service from pickup to dropoff. Sometimes, GPS positions may gravely deviate from the actual position. By calculating the speed of every GPS position according to its previous position, we remove the errant data from our dataset of which the speed is over 300 km/h. We finally use the road matching algorithm [7] to find the closest road of each GPS position. Then, a sequence of GPS positions is transformed into a readable sequence of road names.

4.2 Trajectory Topic Construction

Topic modeling is generally based on the "bag-of-words" assumption and ignores the order of words. However, the trajectories have the direction information. For instance, most cars go into town for work in the morning and out of town at night. These two different patterns may be modeled into one topic if the direction is ignored [7]. In this paper, we distinguish these two kinds of trajectories via the direction information. In natural language processing (NLP), n-gram language models predict the current word according to previous $n-1$ words. When $n=2$, the n-gram language model can be simplified as a bigram language model. Wallach [16] combined the bigram language model with topic modeling and proposed the bigram topic model to consider the order of words.

We apply the bigram topic model to taxi trajectories. Unlike texts, the location of a trajectory is spatially continuous. Also, the topic of each location of trajectories is predicated according to the previous and current location. Thus, the topic has the directionality. Compared with traditional topic models, the bigram topic model takes the direction into consideration, i.e., the order of road names.

4.3 Topical Sub-trajectory Extraction

Topic modeling generates several distributions over words for latent topics, and the top ranking words are usually used to represent topics. However, single words are hard to convey the meaning of topics, such as "data", "mining". But we can easily infer the meaning from phrases ("data mining"). Thus, many works extract the top-ranked phrases (key phrases) to describe the latent topics [11] [14]. Similarly, we use a sub-trajectory (a sub-sequence of road names), instead of single road names, to present the topic with its direction information. If sub-trajectories appear frequently, we call them **frequent sub-trajectories**, similar to key phrases in NLP.

There are many phrase extraction methods in NLP, primarily focusing on extracting noun phrases via part-of-speech tags. However, trajectories do not have this kind of syntactic structure. Thus, we propose a modified Apriori algorithm [2] to extract frequent sub-trajectories. The frequent sub-trajectory set L_1 is initialized with the road names of trajectories. For each iteration, our algorithm generates the candidate frequent sub-trajectory set C_k via adding adjacent road names to the front or the end of sub-trajectories in the last frequent sub-trajectory set L_{k-1}, where k is the length of sub-trajectories. We further filter sub-trajectories in C_k, if the support of a sub-trajectory (the number of trajectories containing the sub-trajectory dividing by the total number of trajectories in the data) is less than the minimum support threshold it will be discarded. After removing these less frequent sub-trajectories, we obtain the frequent sub-trajectory set L_k. Finally, we extract frequent sub-trajectories $\bigcup L_i, i \in [2,k]$ as representative sub-trajectories.

As the number of sub-trajectories would be too large, we further aggregate sub-trajectories to reduce redundancy and encourage diversity during visual exploration. For a frequent sub-trajectory, its sub-trajectories are also in the frequent sub-trajectory set, and they may be redundant. The Apriori algorithm can extract strong rules with high confidence, which indicates how often the rule has been found to be true. Thus, we can use the confidence to identify redundant frequent sub-trajectories. If A is the sub-trajectory of B, and the confidence $A \rightarrow B$ (the support of B divided by the support of A) is more than the minimum confidence threshold, we remove the frequent sub-trajectory A from the frequent sub-trajectory set.

According to the bigram topic model, we can calculate the conditional probability distribution of a road name w over each topic z, $p(w|z)$. A frequent sub-trajectory s consists of several road names. We multiply the probability distribution of road names in the frequent sub-trajectory and calculate the frequent sub-trajectory probability distribution over each topic $p(s|z) \in R^m$, where m is the number of topics. Then we use frequent sub-trajectories to represent one topic. Each frequent sub-trajectory is associated with the topic with the highest probability if the probability is also higher than the minimum topic threshold. These topic-associated frequent sub-trajectories are called **topical sub-trajectories** in this paper.

5 Visual Design

Based on extracted topics and topical sub-trajectories, we design a visual analytics system to explore themes of taxi trajectories interactively. Our linked views consist of three parts, as shown in Fig. 1. The control views help users find topics and topical sub-trajectories of interest. The map view presents the spatial distribution of topical sub-trajectories or trips on the map. The detail views show the detail information of topics and their sub-trajectories. The video in supplemental material shows the interface of our system.

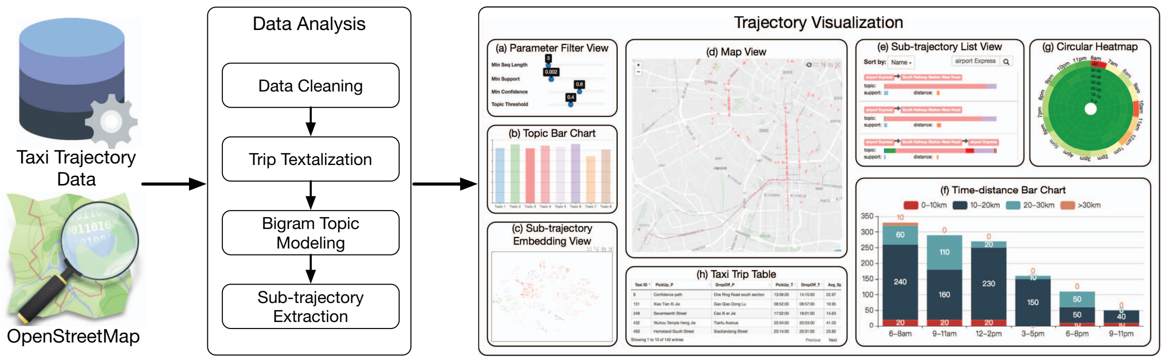

Figure 1: The visual analytics pipeline for taxi trajectories, including data analysis module and trajectory visualization module.

5.1 Control Views

The **parameter view** allows users to change the parameters of topical sub-trajectory extraction, as shown in parameter filter view (Fig. 1(a)). It contains four parameters: the minimum length of sub-trajectories, the minimum support threshold, the minimum confidence threshold, and the minimum topic threshold. The length of sub-trajectory is the number of road names. For example, the length of the sub-trajectory 'Qingjiang Road→Chengwen Road' is two.

The **topic bar chart** displays the number of trips in each topic, which contains topical sub-trajectories in the topic, as shown in topic bar chart (Fig. 1(b)). Each bar corresponds to one topic with one unique color. The user can interactively select one topic for a detailed analysis of its topical sub-trajectories.

The **sub-trajectory embedding view** shows all topical sub-trajectories via t-SNE [15] to reveal the similarity between them, as shown in sub-trajectory embedding view (Fig. 1(c)). The distance between sub-trajectories is calculated based on the topic distribution of sub-trajectories. Each circle corresponds to one sub-trajectory. Its size encodes its support value, and its color is the same as the color of its associated topic in topic bar chart (Fig. 1(b)). When hovering on a circle, it will be enlarged and the map view will highlight this sub-trajectory. When a circle is selected by clicking, the map view will show the sub-trajectory and taxi trips that contain the sub-trajectory on the map, and other views will also display the detail information of the sub-trajectory. This view also supports the brush operation to select several sub-trajectories interactively.

5.2 Map View

The **map view** displays topical sub-trajectories and taxi trips on the map, as shown in map view (Fig. 1(d)). These trajectories are encoded by lines with arrows to reveal the direction information of trajectories. Topical sub-trajectories are visualized with the same color of their associated topics, and taxi trips are colored in gray. The width of each sub-trajectory is encoded as its support value.

The map view supports many interactions for detailed analysis. When hovering on a sub-trajectory, its corresponding circle will be highlighted in the sub-trajector embedding view. When the user selects a sub-trajectory, the map view will show the taxi trips that contain the selected sub-trajectory or its sub-trajectories due to the aggregation. The sub-trajectory list view will scroll to the selected sub-trajectory, and other views will also display the detail information of the sub-trajectory. The brush operation allows users to select multiple trajectories by region and discover their distribution in the sub-trajectory embedding view.

5.3 Detail Views

The detail information of a sub-trajectory includes its sequence of road names, support value, and distance. For the trips containing a sub-trajectory, we provide their pick-up times, speeds, and distances to analyze traffic conditions, such as traffic congestion. Thus, we design the following views for detailed visual exploration.

The **sub-trajectory list view** shows the topical sub-trajectories with the meta information, as shown in sub-trajectory list view (Fig. 1(e)). It uses the road name sequence to describe each sub-trajectory briefly, and assists users to quickly find sub-trajectories according to the road names. Since the road is divided into several segments and textualized into different words to improve the granularity of analysis, we merge the words with the same road name to simplify the road name sequence in the list view. Furthermore, the topic distribution of the sub-trajectory is displayed by a stacked bar chart below the road name sequence, and two additional bar charts show its support value and distance of the sub-trajectory. The list view support sorting by the road name, support value, distance, and length, and searching for topical sub-trajectories by the road name.

For the selected sub-trajectory, the **time-distance bar chart** illustrates the distribution between the pick-up time and the distance of trips that contain the sub-trajectory, as shown in time-distance bar chart (Fig. 1(f)). The x-axis represents the pick-up time, and the y-axis represents the number of taxi trips in different distance ranges via a stacked bar chart. We define four distance ranges: 0 - 10 km, 10 - 20 km, 20 - 30 km, $>$ 30 km. When a topic is selected in the topic bar chart but no specific sub-trajectory is selected, the time-distance bar chart will show the distribution of all sub-trajectories in the topic.

The **circular heatmap** illustrates the distribution between the pick-up time and the speed of a sub-trajectory or all sub-trajectories in a topic, as shown in circular heatmap (Fig. 1(g)). The whole day is divided into 18 hours, from 6 o'clock to 24 o'clock. Each hour is represented by seven arcs. The speed increases from inside to outside in seven ranges, respectively 0 - 9 km/h, 19 km/h, 10 - 20 to 29 km/h, 30 - 39 km/h, 40 - 49 km/h, 50 - z59 km/h, and $>$ 60 km/h. The color from green to red reveals the number of taxi trips.

The **taxi trip table** (Fig. 1(h)) contains the detail information of taxi trips that contain the selected sub-trajectory, including taxi id, pick up position, dropoff position, pick up time, dropoff time, average speed, max speed, min speed, and distance.

6 Case Studies

Our case studies analyze the trajectories on August 22, 2014, and there are 419,410 trips after data cleaning. There are 8 patterns/topics according to the bigram topic model. The larger the topic number is, the better patterns will be explored. But a large number of topics is hard to analyze.

6.1 Topic Analysis

We first explore topics and their topical sub-trajectories from the spatial, temporal, and other dimensions. As shown in Fig. 1(b), the trip number of each topic is roughly the same, and the number of Topic 7 is less compared with other topics.

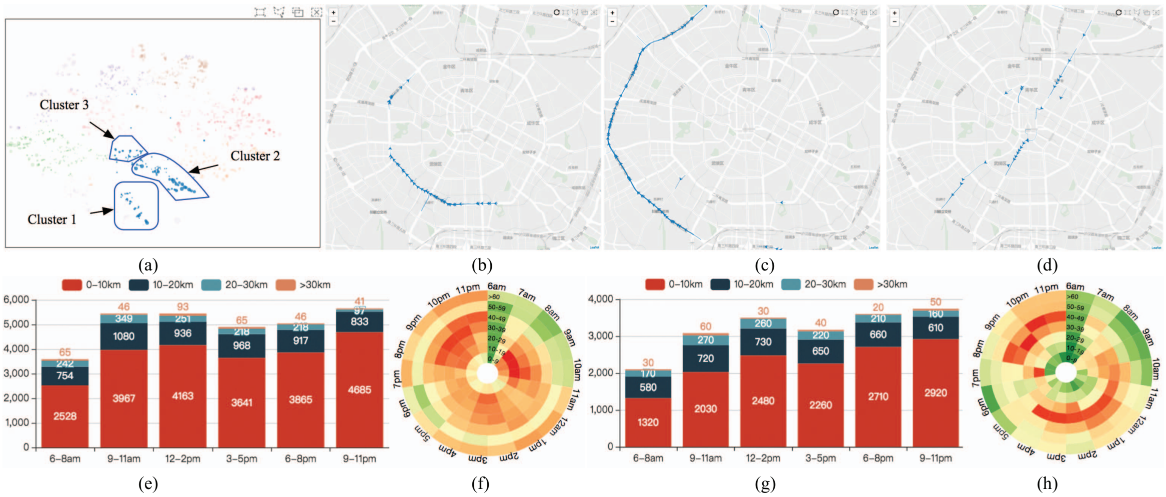

Figure 2: Topic 1 has three sub-trajectory clusters (a): near 2nd Ring Road south section and west section (b), from South 3rd Ring Road 5th section to North 3rd Ring Road 1st section (c), and from the northeast region to the southwest region (d). The detail information of the first cluster is shown in (e) and (f), and the detail information of the third cluster is shown in (g) and (h).

Topic 1: As shown in Fig. 2 (a), topical sub-trajectories of Topic 1 can be roughly divided into three clusters. After selecting these clusters via the brush operation, these spatial distributions are displayed in Fig. 2 (b-d), respectively. The first cluster is near 2nd Ring Road south section and west section, and the second cluster is from South 3rd Ring Road 5th section to North 3rd Ring Road 1st section. Their directions are both clockwise, and the distributions of the pick-up time, distance, and speed are similar in the first two clusters. As shown in Fig. 2(e), the numbers of trips remain roughly unchanged after 9 am, and most distances of trips are within 10km. The circular heatmap in Fig. 2(f) also suggests that the number of trips increases rapidly from 6 am to 8 am, and the peak is the night from 8 pm to 11 pm. The average speed of these sub-trajectories is between 10 km/h and 49 km/h. The third cluster corresponds to the sub-trajectories from the northeast region to the southwest region, and the number of trips is much less than the first two clusters, as shown in Fig. 2(g)(h). The distribution between the pick-up time and distance is similar to the first two clusters, but the speed distribution is different, only 10 - 29 km/h from 5 pm to 7 pm.

Topic 3 and Topic 8: Fig. 3 shows topical sub-trajectories of Topic 3 and Topic 8 in the opposite direction on many same roads. The direction in Topic 3 is from north to south, and the sub-trajectories are mainly on the West 2nd Ring Road, South Renmin Road, and Kehua Road. Besides, it contains the sub-trajectories from the downtown to Chengdu Shuangliu International Airport. In contrast, the direction of Topic 8 is from south to north. They are mainly on the East 2nd Ring Road, South Renmin Road, Kehua Road and from Chengdu Shuangliu International Airport to the downtown. The distributions of the pick-up time, speed and distance are similar for these two topics. As shown in the time-distance bar chart and circular heatmap, the number of trips is minimal before 7 am and stays basically unchanged after 7 am, and the peak is after 8 pm. The average speed of these trips is 10 - 29 km/h, and the speed is higher in the night from 7 pm to 11 pm. Similarly, we can also find two topics in the opposite direction between east and west, Topic 2 and Topic 4, in the supplemental material. Thus, our method can effectively distinguish trajectories with different directions, even on the same roads.

6.2 Sub-trajectory Analysis

We analyze topical sub-trajectories in several topics.

Airport to downtown: We first analyze the sub-trajectories containing ‘airport express’ in Topic 8. As shown in the map view in Fig. 3(b), we can find that there is only one sub-trajectory containing ‘airport express’ in Topic 8. We directly select the sub-trajectory in the map view. As shown in Fig. 4, the taxis pass through airport express and then travel through 2nd Ring Road, 3rd Ring Road, South Renmin Road and Kehua Road to different regions of Chengdu. From the time-distance bar chart, the pick-up time distribution has two peaks, 12 am - 2 pm and after 9 pm, and the average distance is between 20 km and 30 km. The average speed is mainly greater than 60 km/h due to the expressway.

Figure 3: Topical sub-trajectories of Topic 3 (a) and Topic 8 (b).

Downtown to airport: The sub-trajectories from the downtown to Chengdu Shuangliu International Airport in Topic 3 are shown in Fig. 3(a). By searching ‘airport express’ in the sub-trajectory list view, as shown in Fig. 1(e). We get result related to ‘airport express’ and select them for detailed analysis in Fig. 1. In the map view, the taxis from the central city pass through South Renmin Road and Kehua Road to airport express, and the taxis from other parts of Chengdu pass through 2nd Ring Road and 3rd Ring Road to

Figure 4: The topical sub-trajectories from the airport to downtown.

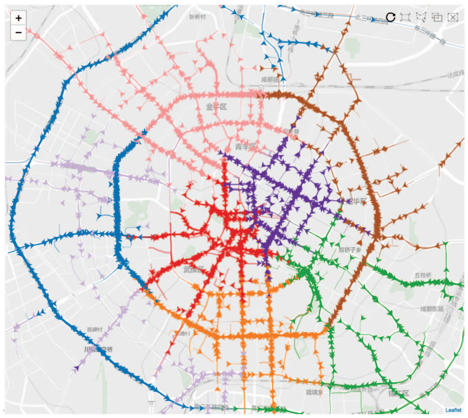

Figure 5: The result of LDA on the same data.

airport express. In the time-distance bar chart, the average distance of trips is in 10 - 30km, which is similar to the sub-trajectories from the airport to downtown. However, the pick-up time distribution is different from Fig. 4 which is mostly in the morning, especially 6 - 7 am and 10 - 12 am, as shown in the circular heatmap.

6.3 Discussion

Comparison: Previous works [7] [13] [3] employ LDA to extract trajectory topics and adopt a group of road names to represent each topic. Since LDA is based on the "bag-of-words" assumption, it does not consider the order of words. Thus, they cannot distinguish different directions of trajectories on the same roads. In this paper, we use the bigram topic model to include the direction information of trajectories in the modeling process. Fig. 5 shows the result of LDA on the same data. It is clear that LDA well separates the sub-trajectories in the geographic space, and this may omit these long trajectories, such as from northeast to southwest in Fig. 4. Moreover, sub-trajectories with different directions are grouped in the same topic. Instead of single road names, we propose topical sub-trajectories to represent the meaning of topics. This would be more readable and understandable for users during visual exploration.

Parameters analysis: For sub-trajectories, we set the minimum length of sub-trajectories as 3, the minimum support as 0.002, the minimum confidence as 0.6, and the minimum topic threshold as 0.4. After extraction and aggregation, we generate 1,633 topical sub-trajectories. We find the minimum support value can ignore most less important sub-trajectories and preserve meaningful sub-trajectories. Also, a small minimum confidence can aggregate similar sub-trajectories. For the minimum topic threshold, a larger value will filter out too many sub-trajectories and a smaller value will introduce lots of sub-trajectories that are not associated with topics.

7 Conclusion and Future Work

We have presented a visual analytics system for taxi trajectory analysis. The taxi GPS location is transformed into a road name by the map match algorithm, and the bigram topic model is applied to extract hidden topics in the trajectory. To better convey these topics, we propose a modified Apriori algorithm to construct frequent sub-trajectories. Our case studies demonstrate our system can effectively extract meaningful topics with direction information and their topical sub-trajectories, and facilitate users visually explore movement behaviors. In the future, we intend to analyze taxi trajectories in a long time to extract weekly or monthly movement patterns. In addition, we plan to fuse different data sources, such as taxi data, bike sharing data, and social media data, to fully understand movement behaviors in the city.

Acknowledgments

This work was supported by the National Key Research & Development Program of China (2017YFB0202203), National Natural Science Foundation of China (61472354 and 61672452), and NSFC-Guangdong Joint Fund (U1611263).

References

[1] N. Adrienko and G. Adrienko. Spatial generalization and aggregation of massive movement data. *IEEE TVCG*, 17(2):205–219, 2011.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB'94*, pp. 487–499, 1994.

[3] S. Al-Dohuki, Y. Wu, F. Kamw, J. Yang, X. Li, Y. Zhao, X. Ye, W. Chen, C. Ma, and F. Wang. SemanticTraj: A new approach to interacting with massive taxi trajectories. *IEEE TVCG*, 23(1):11–20, 2017.

[4] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[5] H. Cao, N. Mamoulis, and D. W. Cheung. Mining frequent spatio-temporal sequential patterns. In *Proceedings of ICDM'05*, pp. 82–89, 2005.

[6] W. Chen, F. Guo, and F. Y. Wang. A survey of traffic data visualization. *IEEE TITS*, 16(6):2970–2984, 2015.

[7] D. Chu, D. A. Sheets, Y. Zhao, Y. Wu, J. Yang, M. Zheng, and G. Chen. Visualizing hidden themes of taxi movement with semantic transformation. In *Proceedings of IEEE PacificVis 2014*, pp. 137–144, 2014.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of KDD'96*, pp. 226–231, 1996.

[9] Y. Kim, J. Han, and C. Yuan. Toptrac: Topical trajectory pattern mining. In *Proceedings of KDD'15*, pp. 587–596, 2015.

[10] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni. Visual analysis of route diversity. In *Proceedings of IEEE VAST*, pp. 171–180, 2011.

[11] Z. Liu, W. Huang, Y. Zheng, and M. Sun. Automatic keyphrase extraction via topic decomposition. In *Proceedings of EMNLP'10*, pp. 366–376, 2010.

[12] J. Pu, S. Liu, Y. Ding, H. Qu, and L. Ni. T-Watcher: A new visual analytic system for effective traffic surveillance. In *Proceedings of IEEE ICMDM*, vol. 1, pp. 127–136, 2013.

[13] Y. Tang, F. Sheng, H. Zhang, C. Shi, X. Qin, and J. Fan. Visual analysis of traffic data based on topic modeling. *Journal of Visualization*, 21(4):661–680, 2018.

[14] N. Teneva and W. Cheng. Salience rank: Efficient keyphrase extraction with topic modeling. In *Proceedings of ACL*, pp. 530–535, 2017.

[15] L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *JMLR*, 9:2579–2605, 2008.

[16] H. M. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of ICML'06*, pp. 977–984, 2006.

[17] W. Wu, Y. Zheng, N. Cao, H. Zeng, B. Ni, H. Qu, and L. M. Ni. Mobiseg: Interactive region segmentation using heterogeneous mobility data. In *Proceedings of IEEE PacificVis 2017*, pp. 91–100, April 2017.

[18] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM TIST*, 5(3):38:1–38:55, 2014.

[19] Y. Zheng and X. Zhou. *Computing with Spatial Trajectories*. Springer Publishing Company, Incorporated, 1st ed., 2011.