Trajectory Outlier Detection for Traffic Events: A Survey

Kiran Bhowmick and Meera Narvekar

Abstract With the advent of Global Positioning System (GPS) and extensive use of smartphones, trajectory data for moving objects is available easily and at cheaper price. Moreover, the use of GPS devices in vehicles is now possible to keep a track of moving vehicles on the road. It is also possible to identify anomalous behavior of vehicle with this trajectory data. In the field of trajectory mining, outlier detection of trajectories has become one of the important topics that can be used to detect anomalies in the trajectories. In this paper, certain existing issues and challenges of trajectory data are identified and a future research direction is discussed. This paper proposes a potential use of outlier detection to identify irregular events that cause traffic congestion.

Keywords Trajectory data · Map matching · Trajectory outlier detection
GPS data · Similarity measures

1 Introduction

Trajectory data is the data about moving objects like vehicles, animals, and people. With the advent of technologies of mobile computing and location-aware services, there is a massive generation of this trajectory data. Trajectory data analysis helps in identifying several real-world phenomena like, for example, identifying the buying patterns of humans in shopping malls, understanding the migratory patterns of animals and birds, and detecting the travel path of hurricanes and tornadoes. Vehicle trajectory analysis is used in various traffic dynamics like traffic jam prediction, roadmap construction, and real-time traffic flow; in route navigation like suggest optimum route, suggest places to visit, and identify frequently visited

K. Bhowmick (✉) · M. Narvekar
Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai, Maharashtra, India
e-mail: kiran.bhowmick@djsce.ac.in

M. Narvekar
e-mail: meera.narvekar@djsce.ac.in

S. Bhalla et al. (eds.), *Intelligent Computing and Information and Communication*,
Advances in Intelligent Systems and Computing 673,
https://doi.org/10.1007/978-981-10-7245-1_5

locations; in itinerary navigation like discover popular routes, predict next location; and in other applications like predicting human flow, suggesting alternate transport means, etc. [1–4].

In recent years, there has been an increase in the research of outlier analysis of trajectory data. This outlier analysis is used to identify anomaly in the regular data. Outlier detection can also be used to identify events that occur in regular traffic. This paper discusses different outlier detection techniques, the research literature available, and suggests the use of trajectory outlier detection for detecting irregular events such as traffic congestion due to road accidents or roadblocks or any other reason.

2 Related Concepts

2.1 Basic Concepts

This section describes the concepts related to trajectory mining and outlier detection basics.

2.1.1 Trajectory data

Definition 1: A trajectory T is an ordered list of spatiotemporal samples *p1, p2, p3, …, pn*. Each $pi = (x_i, y_i, t_i)$, where x_i, y_i are the spatial coordinates of the sampled point and t_i is the timestamp at which the position of the point is sampled, with $t1 < t2 < \ldots < tn$ [5].

In general, a trajectory depicts the time and location of a moving object where it can be tracked. A single trajectory can include many trajectory points and can have any length for a single moving object. Basically, the length depicts the places that the object has traveled from point A to point B before it comes to a halt. A sub-trajectory depicts a partition or a segment of a full trajectory.

2.1.2 Trajectory outlier

Definition 6: An outlier in a trajectory data is an item that is significantly different from other items in terms of some form of similarity [1].

Outliers are objects that are inconsistent with remaining data items. Outliers are also sometimes considered to be noise. But the noise for someone can be someone else's data [6]. For example, in the following, figures t4 and t5 are outliers in the path of the regions R1 and R2.

However, t4 can be considered as an alternate path, while t5 can be a long detour (Fig. 1).

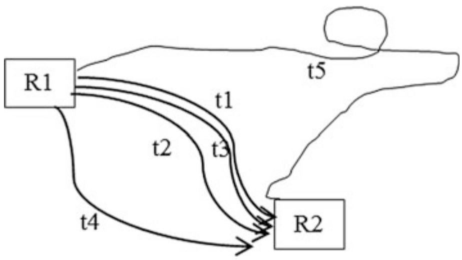

Fig. 1 Outliers between regions

2.2 Similarity Measures

The distance between two trajectories is usually measured by some kind of aggregation of distances between trajectory points. The various similarity measures are defined in [1, 7]. These include closest pair distance (CPD), sum of pairs distance (SPD), Euclidean distance (ED), dynamic time warping (DTW and PDTW), edit distance measures (ERP and EDR), and LCSS. The CPD, SPD, and ED require equal length of the trajectories and calculate the minimum distance between two trajectories. Edit distances, DTW and LCSS, assume variable length trajectories.

2.3 Map Matching

Map matching is a preprocessing technique in the trajectory data mining used to convert the sequence of GPS points to a sequence of road segments. There are various categories of map matching algorithms depending upon the type of information considered and the range of sampling [8]. The category with the type of information includes geometrical, topological, weight-based, probabilistic, and advanced, whereas the range of sampling has two types local/incremental and global methods.

2.4 Outlier Detection Techniques

This section introduces different outlier techniques.

Distance-based outlier detection techniques These techniques consider the distance between trajectory and its neighbors. If the trajectory does not have neighbors above a particular threshold, it is an outlying trajectory [9, 10].

Density-based outlier detection techniques This means a trajectory is an outlier if its density is relatively much lower than that of its neighbors [9, 11].

The distance-based outlier techniques detect global outliers due to the way the distance is measured between the points. It also requires uniform distribution of data. With different density distributions of data, the distance-based techniques cannot be used. The density-based outlier techniques detect local outliers and can do analysis in different density distributions but require a large number of k-nearest neighborhood queries and are computationally very intensive.

Historical similarity-based outlier detection techniques This technique suggests the use of historical similarity to identify trends between data points. At each time step, a road segment is compared with its historical recorded temporal neighborhood values to identify similarity values. A drastic change in the values results in the outlier being detected.

Motifs-based outlier detection techniques A motion classifier to detect trajectory outliers using an objects motion feature is called as motifs. These motifs are a sequence of motion features with values related to time and location. The classifier works on this high-dimensional feature space to distinguish an anomalous trajectory with a normal one.

3 Related Work

Although it is an upcoming field, a lot of work have already been done in the outlier detection and overall trajectory mining. With respect to the different similarity measures, H. Wang et al. in their paper [7] have done a comparative study about various similarity measures under different circumstances.

Map matching helps overcome the problem of uncertainty, but the existing map matching algorithms also require high sampling rate of the data which is not always the case. Yin Lou et al. in [12] have used spatiotemporal matching and analysis to find the candidate graph which helps to map low sampling digital data to a roadmap. K. Zheng et al. [13] proposed the use of similarity between query data points and historical data to infer path for low sampling data. The inference is done using traversal graph and nearest-neighbor-based heuristics approach. In [14], the authors propose a technique to consider the complex nature of the urban roads and address the problem of identifying a data point on elevated roads. The problem of identifying the candidate segment selection for the first GPS point is addressed in [15], where the authors employ a heuristic A* algorithm to find the shortest path between a previous point and the candidate segment. Paolo Cintiaa and Mirco Nannia in [16, 17] argue that the shortest path should not be the measure to find the best matching point on the road network when the data samples are low and that a time-aware Dijkstra's shortest path is suggested to match the points. G. Hu et al. in [18] suggested the use of an information fusion of data to accurately perform map matching as well as consider the complexity of road networks. And lastly, the authors in [19] suggest using LCSS and clustering algorithm totally avoiding the use of any map matching algorithm. They conclude with their experimental results that none of the measures are better than others. In all, each similarity measure is affected by the decrease in the sampling rate. None of the related techniques consider time-aware heuristics methods together with the complexity of road networks like multilayer and parallel roads, complex interchanges, and elevated roads.

Outlier detection has also been researched in different scenarios. Next, paragraph review of literature on the research is been discussed.

Lee et al. first proposed a trajectory outlier detection algorithm TRAOD [19] to detect outlying sub-trajectories using a partition and detect framework. It uses density as well as distance-based measures and does not suffer from local density problem and has minimal overhead for density measures. However, it has a complexity of $O(n^2)$ and considers only spatial data. Piciarelli et al. in [20] used a single-class SVM classifier on fixed-dimensional vectors of spatial data to identify outliers. But the algorithm uses a training data with outliers. Outliers not seen in training data will not be detected. In [21], Li et al. proposed a temporal outlier detection TOD algorithm that considers time dimension and historical data to detect outliers. The authors suggest the use of exponential function to detect outliers. This exponential function, however, cannot be used for spatial dimension. Yong Ge et al. in [22] proposed TOP-EYE to detect an evolving trajectory using a decay function. If a trajectory is going to be an outlier, then it will be detected at an early stage by this method. Wei Liu et al. in [23] suggested the use of frequent pattern tree to identify the causal interactions between the detected outliers. They have considered temporal information about current and historical time frames trajectory to compare using Euclidean distance for detecting outliers. Daqing Zhang et al. proposed their algorithm iBAT [24] for detecting outliers based on the isolation techniques. An iTree based on the data-induced random partition is generated for all the trajectories. If a trajectory is an outlier, then it will have comparatively shorter paths than those of a normal trajectory. Alvares et al. in [25] detected an outlier between surveillance camera and inbound trajectories by analyzing the behavior of trajectories. If a trajectory seems to avoid the object in certain patterns, then it is detected as outlier. But it fails to identify if the avoidance was forced or intentional. Chawla et al. in [26] used PCA analysis to identify anomalies and then L1 optimization to infer possible routes whose flow traffic caused the outlier. Zhu et al. in [27, 28] used time-dependent transfer graphs to identify outliers online.

From the above literature review of outlier, it is observed that, while considering trajectory data, its inherent characteristics affect the outlier detection method. Preprocessing of the data can reduce most of the challenges. But the fact that the trajectory data is not equally distributed on all roads, it is important to consider historical data. The choice of hybrid outlier detection technique and suitable similarity measure can improve the accuracy majorly.

4 Issues and Challenges Identified

Trajectory data can be efficiently used in traffic management. Considering the characteristics of trajectory data and the literature review, the following issues and challenges related to trajectory data, map matching, and outlier detection are identified.

4.1 Trajectory Length and Sampling

Every trajectory have different lengths, i.e., the number of data points is different for each, i.e., T1 = (p1, p2, …, pn) and T2 = (q1, q2, …, qm) and n and m can be different. It is important to identify the region of interest to consider the start and end of each trajectory.

4.2 Low and Uneven Sampling Rates

The GPS data of the moving vehicles usually have a low sampling rate to avoid the overhead of communication cost, data storage, and battery life of devices. So the time interval between two consecutive GPS points can be very large (average sampling rate 2 min). This leads to an uncertainty of the path taken by a trajectory between these time gaps. Also, the data sampling rates are different across all trajectories, i.e., T1 may be sampled for every 1 m and T2 is sampled every 2 m.

4.3 Trajectory Directions, Regions, and Road Networks

Trajectories moving in different directions and different regions should be considered different. Trajectories moving in different directions but close proximity and those moving in the same direction but different regions should be considered as different. Also for vehicles, the trajectories are bound by the underlying road networks, so the data points should match the road networks map.

4.4 Similarity Measures

There are different types of similarity measures as mentioned in Sect. 2.2. However, with different lengths, low samplings, uncertainties, and noises, it is difficult to identify a suitable similarity measure to compare trajectories together. All the measures described in above section are sensitive to decrease in sampling rate. It is a challenge to process data with low sampling rate.

4.5 Map Matching Challenges

Most of the existing map matching algorithms assume high sampling of data as they usually perform local or incremental approach. Also, they consider only spatial data for mapping to road networks. However, there is a lot of information available that can improve the map matching considerably like time, speed, and direction. The complexity of urban road networks is another issue. The topology of road networks in urban areas has many challenges like parallel roads, multilayer roads, complex interchanges, and elevated roads. Considering all these to perform map matching increases the complexity, but will be realistic and accurate.

4.6 Outlier Detection Challenges

Different types of outlier detection techniques exist, viz., distance-based and density-based. The existing trajectory outlier detection algorithms have different techniques and each tries to solve the outlier detection with a different objective. The issue is to devise an algorithm that can overcome the problem of uncertainty, low sampling, and uneven length data with mixed distribution of trajectories and considering spatial as well as temporal similarity constraints efficiently identifies outliers even with large amount of data. A data structure that can help to deal with the variety of information and the length of the data is needed. The algorithm should be able to identify sparse as well as dense trajectories for online data. Also, there needs to be a way to propagate the effects of the outlier to linked traffic nodes.

5 Research Problem

The outlier detection is still an upcoming research due to its application in traffic management. A huge number of trajectory data are generated for a single moving object and there are thousands of moving objects in the form of GPS enabled vehicles and human beings carrying smartphones. The large number of sampling points and the uneven length requires a suitable data structure like a modified R-tree

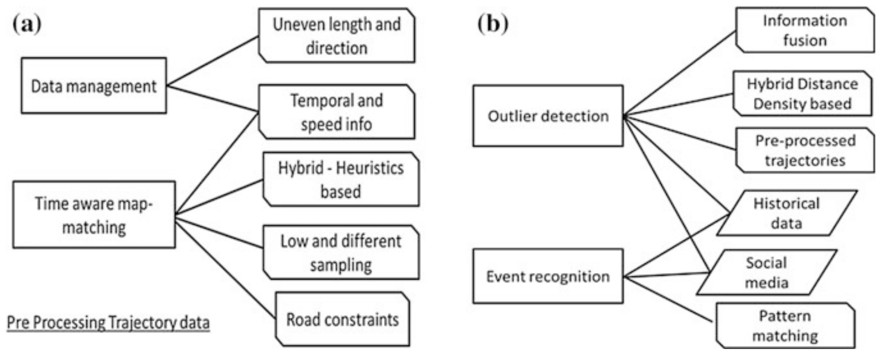

Fig. 2 **a** Preprocessing **b** Outliers between regions

for fast retrieval and indexing of the trajectory data. The R-tree should be able to store temporal and related information (Fig. 2).

The trajectory data has uneven rate of sampling and an inherent uncertainty about the locations between two sampled points. This uncertainty between sampling points can be reduced by using map matching algorithms. But the existing map matching algorithms consider only spatial data and work well with high sampling. Also, it is important to consider the urban road complexity like multi-layer, complex interchanges, and elevated roads. With low sampling, this makes it more challenging. The use of shortest path to find the actual location of the trajectory is not helpful. There has to be a consideration of hybrid technique like combining the global method with weight-based technique or global method with advanced or probabilistic method. A time-aware heuristic method can improve the accuracy of map matching algorithms in low sampling data. Since the distribution of trajectory data is skewed, using only density-based or distance-based method of outlier detection is not the solution. A hybrid approach is used by many existing algorithms but temporal information is not considered. Using time-dependent popular routes graph based on historical data to deal with the changing nature of outliers is beneficial. The detected outliers can be used to identify events by performing pattern matching and comparing it with similar patterns that were seen earlier when an irregular event occurred.

6 Conclusion

This paper is a part of an on-going research on identification of traffic events using vehicle trajectory. This paper mainly focusses on the identifying problems in trajectory data and outlier detection. A detailed literature survey on existing map matching techniques and outlier detection technique is discussed. The paper also discusses the proposed research problem for identifying traffic events using outlier detection techniques.

7 Future Scope

In the future, we plan first to enhance our research by designing a novel outlier detection technique focussing on the issue of low sampling and uncertainty using taxi trajectory data and second to design pattern matching algorithms to identify traffic events.

References

1. Zheng Y. 2015. "Trajectory Data Mining: An Overview", ACM Transactions on Intelligent Systems and Technology (TIST), Volume 6 Issue 3, May 2015 Article No. 29
2. Zaiben Chen et. al., "Discovering Popular Routes from Trajectories", *ICDE Conference 2011*, IEEE 2011, DOI 978-1-4244-8960-2/11
3. Lin, Miao and Wen-Jing Hsu. "Mining GPS Data for Mobility Patterns: A Survey". *Pervasive and Mobile Computing* 12, Elsevier, pp. 1–16, July 2013
4. Li Z, et. al., MoveMine: Mining Moving Object Databases. *SIGMOD '10 ACM,* pp. 1203–1206, June 2011
5. Wagner et. al., "Mob-Warehouse: A Semantic Approach for Mobility Analysis with a Trajectory Data Warehouse", Springer, pp. 127–136, Nov 2013
6. J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. Morgan Kaufmann, 2006
7. H. Wang et.al., "An Effectiveness Study on Trajectory Similarity Measures", Proceedings of the Twenty-Fourth Australasian Database Conference (ADC 2013), Adelaide, Australia, vol. 137, 2013
8. Mahdi Hashemi and Hassan A. Karimi, "A critical review of real time map matching algorithms: Current issues and future directions", Computers, Environment and Urban Systems, Elseveir, vol 48, pp. 153–165, Aug 2014
9. Gupta, Manish et al. "Outlier Detection for Temporal Data: A Survey". *IEEE Transactions on Knowledge and Data Engineering* 26.9, vol. 25, no. 1, 2014: 2250–2267
10. Knorr, Edwin M., Raymond T. Ng, and Vladimir Tucakov. "Distance-Based Outliers: Algorithms And Applications". The VLDB Journal, 8.3–4 (2000): 237–253
11. Liu, Zhipeng, Dechang Pi, and Jinfeng Jiang. "Density-Based Trajectory Outlier Detection Algorithm". Journal of Systems Engineering and Electronics 24.2 (2013): 335–340
12. Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories", Proceedings of the 17th ACM SIGSPATIAL—GIS '09
13. K. Zheng, Y. Zheng, X. Xie, and X. Zhou, "Reducing uncertainty of low-sampling-rate trajectories", 28th IEEE International Conference on Data Engineering. IEEE, 2012, pp:1144–1155
14. Z. He, S. Xi-wei, P. Nie and L. Zhuang, "On-line map-matching framework for floating car data with low sampling rate in urban road networks", IET Intelligent Transport Systems, vol. 7, no. 4, pp. 404–414, 2013
15. M. Quddus and S. Washington, "Shortest path and vehicle trajectory aided map-matching for low frequency GPS data", Transportation Research Part C: Emerging Technologies, Elsevier, vol. 55, pp. 328–339, March 2015
16. Paolo Cintiaa, and Mirco Nannia "An effective Time-Aware Map Matching process for low sampling GPS data", Elsevier, March 2016
17. G. Hu, J. Shao, F. Liu, Y. Wang and H. Shen, "IF-Matching: Towards Accurate Map-Matching with Information Fusion", IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 1, pp. 114–127, Oct 2016

18. J. Kim and H. Mahmassani, “Spatial and Temporal Characterization of Travel Patterns in a Traffic Network Using Vehicle Trajectories”, Transportation Research Procedia, vol. 9, pp. 164–184, July 2015
19. J. G. Lee, J. W. Han and X. L. Li. “Trajectory outlier detection: a partition and detect framework”. 24th International Conference on Data Engineering ICDE, IEEE, pages 140–149, 2008
20. Piciarelli, C., C. Micheloni, and G.L. Foresti. “Trajectory-Based Anomalous Event Detection”. IEEE Trans. Circuits Syst. Video Technol. 18.11 (2008): 1544–1554
21. X. L. Li, Z. H. Li, J. W. Han and J. G. Lee. “Temporal outlier detection in vehicle traffic data”. 25th International Conference on Data Engineering, pages 1319–1322, 2009
22. Yong Ge, et. al., “TOP-EYE: Top-k Evolving Trajectory Outlier Detection”, CIKM’10 ACM, 2010
23. Wei Liu et. al., “Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams”, KDD’11, ACM, August 2011, California, USA
24. Daqing Zhang et. al., “iBAT: Detecting anomalous taxi trajectories from GPS traces”, UbiComp’11, ACM 978-1-4503-0603-0/11/09, September 2011, Beijing, China
25. L. Alvares, A. Loy, C. Renso and V. Bogorny, “An algorithm to identify avoidance behavior in moving object trajectories”, Journal of the Brazilian Computer Society vol. 17, no. 3, pp. 193–203, 2011
26. S. Chawla, Y. Zheng, and J. Hu, Inferring the root cause in road traffic anomalies, 12th IEEE International Conference on Data Mining. IEEE, 141–150, 2012
27. Zhu, Jie et al. “Time-Dependent Popular Routes Based Trajectory Outlier Detection”. Springer International Publishing (2015): pp. 16–30, Switzerland, https://doi.org/10.007/978-3-319-26190-4_2
28. Zhu, Jie et al. “Effective And Efficient Trajectory Outlier Detection Based On Time-Dependent Popular Route”. *Springer Science* (2016), NY, https://doi.org/10.007/s11280-016-000-6