TOP-EYE: Top-k Evolving Trajectory Outlier Detection

Yong Ge[1], Hui Xiong[1], Zhi-Hua Zhou[2], Hasan Ozdemir[3], Jannite Yu[3], K.C. Lee[3]
[1] MSIS Department, Rutgers University
E-mail: hxiong@rutgers.edu, yongge@pegasus.rutgers.edu
[2] National Key Laboratory for Novel Software Technology, Nanjing University,China
E-mail: zhouzh@lamda.nju.edu.cn
[3] Panasonic System Solutions Development Center (PSDU) of USA,
E-mail: {timucin, jyu, kclee}@research.panasonic.com

ABSTRACT

The increasing availability of large-scale location traces creates unprecedent opportunities to change the paradigm for identifying abnormal moving activities. Indeed, various aspects of abnormality of moving patterns have recently been exploited, such as wrong direction and wandering. However, there is no recognized way of combining different aspects into an unified evolving abnormality score which has the ability to capture the evolving nature of abnormal moving trajectories. To that end, in this paper, we provide an evolving trajectory outlier detection method, named TOP-EYE, which continuously computes the outlying score for each trajectory in an accumulating way. Specifically, in TOP-EYE, we introduce a decay function to mitigate the influence of the past trajectories on the evolving outlying score, which is defined based on the evolving moving direction and density of trajectories. This decay function enables the evolving computation of accumulated outlying scores along the trajectories. An advantage of TOP-EYE is to identify evolving outliers at very early stage with relatively low false alarm rate. Finally, experimental results on real-world location traces show that TOP-EYE can effectively capture evolving abnormal trajectories.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Algorithms, Experimentation

Keywords

Outlier Detection

1. INTRODUCTION

Advances in sensors, wireless communication, and information infrastructures such as GPS, WiFi, Video Surveillance, and RFID have enabled us to collect large amounts of fine-grained location traces (trajectory data). Such a large number of trajectories provide us unprecedented opportunity to automatically discover useful knowledge, such as identifying suspicious activities, understanding the behaviors of moving objects as well as the patterns of transportation networks, which in turn delivers intelligence for real-time decision making in various fields, such as surveillance, transportation management, and border security.

Recent years have witnessed an increasing interest in trajectory outlier detection [2, 7, 1, 3], which aims to detect suspicious moving objects automatically. However, while various aspects of abnormality of moving objects have been exploited, there is no recognized way of combining different aspects into an unified evolving abnormality score which has the ability to simultaneously capture the evolving nature of many abnormal moving trajectories. For example, in Figure 1, among a large number of trajectories, both Trajectory A and Trajectory B show certain degree of abnormality in terms of their evolving moving directions. An intriguing question from this case is how to dynamically and systematically identify top-k outlying trajectories with respect to different features among a huge number of trajectories . The answer to this question is very important, since outlying trajectories usually carry interesting information of potential problems which requires real-time attention and should be detected and dealt with at the early stage.

To this end, in this paper, we propose a top-k evolving trajectory outlier detection method, named TOP-EYE, to capture and maintain top-k evolving outlying trajectories in a real-time fashion. The objective of TOP-EYE is to be more tolerant to noisy disturbance and detect trajectory outliers at the early stage. To achieve this objective, the key challenge is how to develop an evolving outlying score which can capture the evolving nature of outlying trajectories. In TOP-EYE, we take both current and past outlierness of each trajectory into consideration when designing the outlying score function. Specifically, the past outlier score of each trajectory is mitigated by a decay function and the mitigated score is accumulated with the instant outlier score to produce the current outlier score for this trajectory.

In this paper, we mainly consider two types of outlying trajectory: outliers in terms of direction and outliers in terms of density. First, unlike traditional vector data, trajectories are complex data for which traditional distance measures and outlying score functions cannot be used di-

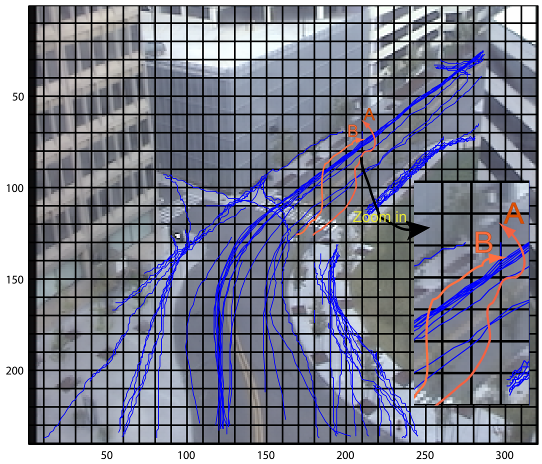


Figure 1: Top-K Evolving Outliers: Motivation

rectly. To deal with this challenge, a data transformation process is applied. Specifically, we discretize the continuous space into small grids and use a probabilistic model to turn the direction information of trajectories in a grid into a vector with eight values to indicate the probabilities of moving towards eight directions within this grid.

Therefore, we can generate the direction trend within a fixed area by summarizing the moving directions from large amounts of trajectories for a period of time. Then, once some objects move across this area along the completely different directions from the summarized directions, we can detect them as outliers in a real-time fashion by measuring the similarity between the directions of the observed objects and the summarized directions. Also, with this discretized space, we can compute the density of trajectories in each grid conveniently. The trajectory density within each grid is estimated as the number of trajectories across this grid. We can obtain the trajectory density distribution with sufficient historical trajectory data. The outlying score of a new trajectory can then be measured based on the density of trajectories in the grids where this trajectory actually passes.

One benefit of this data transformation is to save the computational cost for finding trajectory outliers. For trajectory data, some people defined the distance measures between trajectories at the point level and detect outliers based on this distance measure [1, 2, 4]. However, a lot of computation will be involved because we have a huge amount of points for trajectories. Compared with these methods, the data transformation can save a lot of computation cost, because we only do computation at the grid level when we compute the outlying scores.

Finally, we apply TOP-EYE for finding top-k evolving trajectory outliers in location traces collected from a parking lot. Experimental results show that TOP-EYE could effectively identify evolving trajectory outliers and reduce the false alarms.

2. EVOLVING OUTLIER DETECTION

Here, we briefly introduce the outlying score calculations and the evolving outlier detection method.

2.1 Outlying Score Calculation

To measure the outlier score for trajectories, we will discretize the continuous direction information as well as the

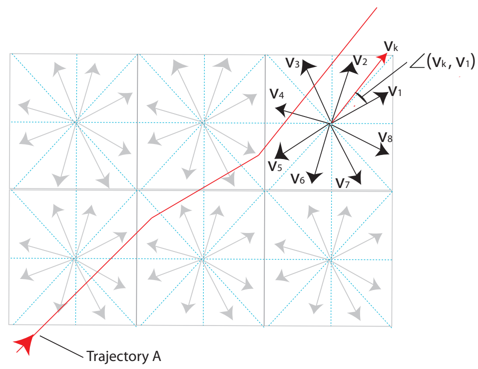


Figure 2: Direction Distance Measure: Illustration

continuous space. Specifically, we divide the monitoring area into small grids. Then we further partition each grid into 8 direction bins, each bin with an angle range of $\pi/4$. The goal of this partition is to summarize the directions within grids with the whole trajectories and represent the summary with a direction vector. Specifically, we can represent each grid with a direction vector:

$$g = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8) \tag{1}$$

where p_i is the frequency of moving trajectories which have passed this grid and have the direction along the direction i, i.e., within the direction bin i. After the monitored area is partitioned into grids, the density of each grid can be computed by scanning each individual trajectory and increasing the density of all grids where this trajectory passes.

In this paper, we generally define the trajectory outlier in terms of direction as :

Definition 1. *Direction-based Trajectory Outlier: A trajectory outlier is the one that deviates from most trajectories within an observed space and time in terms of direction.*

With this definition and the data transformation above, we design the specific direction-based outlier score calculation as follows. Specifically, suppose a new trajectory has $\mathcal{K}$ moving directions within a grid, we similarly represent the directions of this new coming trajectory with a direction vector, e.g., $(q_1, \cdots, q_{\mathcal{K}})$, where $q_k (1 \le k \le \mathcal{K})$ is equal to $1/\mathcal{K}$. Then we can measure the outlying score by measuring the similarity or distance between the summarized direction vector and the direction vector of this new trajectory as:

$$OScoreDir = 1 - \sum_{k=1}^{\mathcal{K}} q_k \sum_{i=1}^{8} p_i \cdot cos\angle(v_k, v_i) \tag{2}$$

where $cos(v_k, v_i)$ is the cosine value of the angle between direction v_i and v_k as shown in Figure 2. Here, v_i is the representative direction for the direction bin i, such as the center direction of the bin.

Also, we consider the density-based outlier score. Specifically, we can simply assign an outlying score for a trajectory if the density of grid, where this trajectory passes, is less than a certain threshold as

$$OScoreDen = \begin{cases} s & \text{if } density < \tau \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

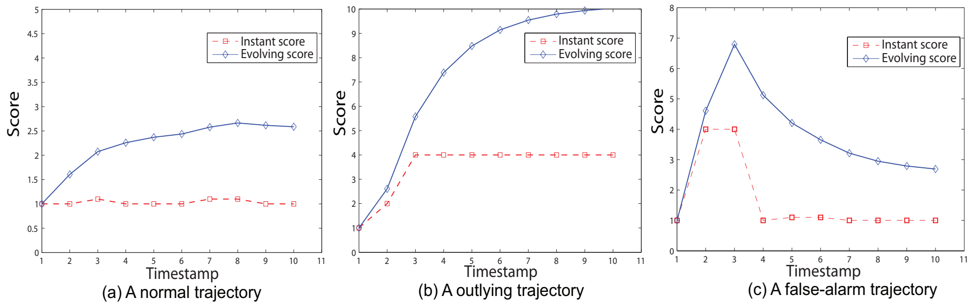


Figure 3: An Illustration of the Use of Evolving Outlying Score

2.2 The Evolving Outlier Detection Model

In this subsection, we introduce an evolving trajectory outlier detection method, named TOP-EYE, which has the ability to continuously compute the outlying score for each trajectory in an accumulating way.

A trajectory is the trace of a moving object. The abnormal behaviors of a moving object can be gradually reflected in its moving characteristics which are encoded in the trajectory of this object. In other words, it is necessary to detect the trajectory outliers in an evolving way. To this end, it is critical to design an evolving trajectory outlier score measure which is able to take both historical outlierness and instant outlierness of a trajectory into the consideration. A key challenge is how to strike a balance between historical influence and instant outlierness for measuring the evolving outlier score. There are two design goals. First, the influence of historical outlierness should be decayed with time. Second, the outlier score measure should be continuously updated in an accumulating way.

In this paper, we exploit an exponential decay function, $exp(-\lambda\Delta t)$, to control the influence of historical outlierness of a trajectory. For example, if one new trajectory Trj passes a grid at initial time instant t_0, its outlying score in terms of direction or density can be obtained by Equation (2) or Equation (3) and is represented as S_{t_0}. Then, at the next time instant t_1, we calculate the current instant outlier score just like at instant t_0, represented as S_{t_1}. Then we accumulate the previous score and the current instant score altogether as:

$$S_{t_1}^{\sum} = S_{t_1} + S_{t_0} * exp(-\lambda\Delta t_0), \quad (4)$$

where λ is a user-specified parameter which affects the decay rate. Also, Δt_0 is the time gap between instant of t_1 and t_0. In general, we can accumulate the evolving outlying score at any time instant as:

$$S_{t_i}^{\sum} = S_{t_i} + S_{t_{i-1}} * exp(-\lambda\Delta t_{i-1}) + S_{t_{i-2}} * exp(-\lambda\Delta t_{i-2}) + \cdots + S_{t_0} * exp(-\lambda\Delta t_0), \quad (5)$$

where Δt_{i-k} is the time gap between instant t_i and instant t_{i-k}. If the time gap between any two neighboring instants is equal to Δt. Equation (5) can be updated as

$$S_{t_i}^{\sum} = \sum_{k=0}^{i} S_{t_k} exp(-(i-k)\lambda\Delta t). \quad (6)$$

As we can see, the evolving score of a trajectory with the decay and accumulation functions can keep updating if this trajectory is gradually evolving to be an outlier. Once the evolving score is above a user-specified threshold, the alarm can be triggered at the early stage. Also, the accidental increase of the evolving outlier score caused by noisy disturbance will not trigger the alarm as long as the threshold is properly specified.

To better illustrate the advantages of this evolving outlier detection method, we provide an example as shown in Figure 3, where $\lambda = 0.5$, $\Delta t = 1$. We assume there are 3 trajectories with 10 time stamps, which correspond to a normal trajectory, an abnormal and a false-alarm trajectory respectively. In Figure 3 (b), the evolving score keeps increasing when the instant outlying score achieves and maintains high values. While for Figure 3 (c), the evolving score quickly decreases once the instant score drops down to normal level. Therefore, if we only consider instant outlying scores, we will certainly decide both (b) and (c) are outlier or normal at a time stamp. However, with the evolving outlier detection model, we can easily use a threshold, e.g., 7, to separate an outlier as (b) from a false alarm as (c).

3. EXPERIMENTAL RESULTS

In this section, we validate the performances of TOP-EYE for top-k evolving trajectory outlier detection.

3.1 Experimental Setup

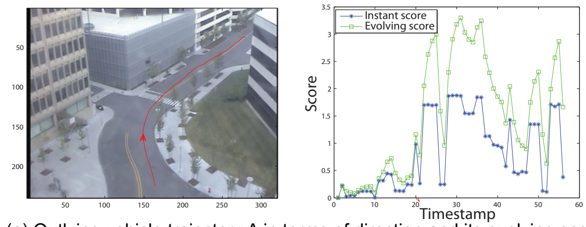
Experimental Data. In the experiments, we used a real-world trajectory data set from Massachusetts Institute of Technology (MIT) [5]. This trajectory data set was extracted from a video surveillance system, which is used to monitor a specific area. This data set contains a large number of trajectories. Each trajectory is a sequence of observations with time-stamped positions, indicating the location of an object at different time stamps. More detailed information about the meta data set can be found in [6].

Experimental Tools and Parameters. In the experiments, since the time gap between neighboring observations of each trajectory is equal, Δt has the same impact on all evolving scores calculated with Equation (6). Therefore, without loss of generality, we assume $\Delta t = 1$. The parameter λ is empirically identified as 0.72 for the direction outlier, 0.84 for the density outlier. Also, we set $\tau = 4, s = 1$ for Equation (3).

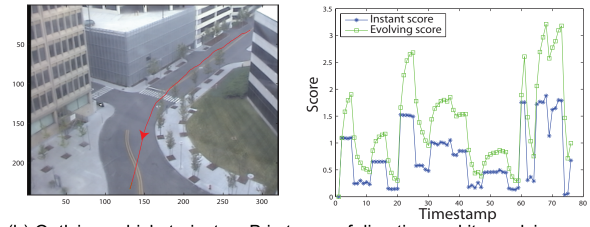
3.2 The Effectiveness of TOP-EYE

In this subsection, we show the effectiveness of TOP-EYE by illustrating how the evolving outlier score can be used to identify the evolving outliers and avoid false-alarms.

First, Figure 4 shows two vehicle trajectory outliers in terms of direction. In Figure 4 (a), the vehicle moves to the left side of the road (wrong way), and the evolving score gradually increased as shown in the figure. Also, in Figure 4 (b), the vehicle takes a short cut to drive to the wrong way.

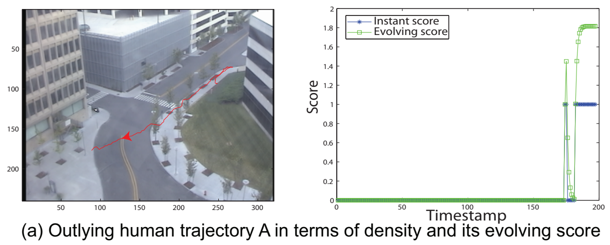


(a) Outlying vehicle trajectory A in terms of direction and its evolving score

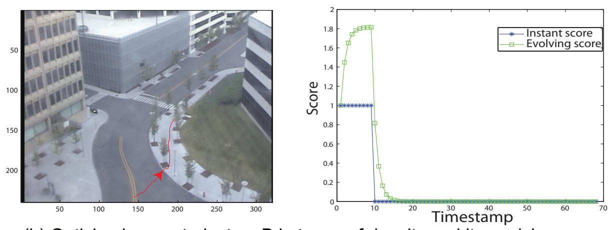


(b) Outlying vehicle trajectory B in terms of direction and its evolving score

Figure 4: Evolving score in terms of direction.



(a) Outlying human trajectory A in terms of density and its evolving score



(b) Outlying human trajectory B in terms of density and its evolving score

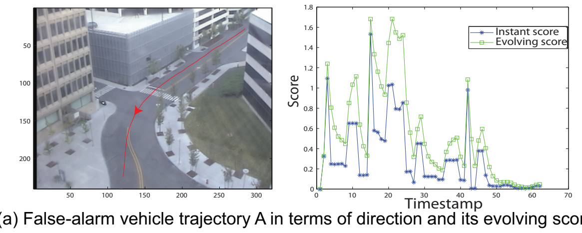
Figure 5: Evolving score in terms of density.

For both cases, TOP-EYE could effectively capture the outliers by thresholding evolving outlying scores. In addition, a similar performance for the density outlier detection can be observed in Figure 5, where TOP-EYE can identify people who are crossing the street illegally.

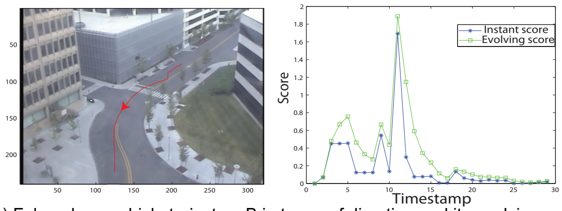
Furthermore, TOP-EYE can also help to reduce false alarms. Figure 6 shows the false trajectory outliers in terms of moving direction. For example, the trajectory in Figure 6 (b) has a direction change for a very short period of time. However, this trajectory is in the normal status for the rest time. This sudden change causes an impulse for instant scores, resulting in the highest instant score (1.6), which is close to the highest instant score of normal outlier in Figure 4. In contrast, the maximum value of evolving scores is always below 2. If we use the same threshold for evolving outlying scores as shown in Figure 4, we can avoid this false alarm. However, we will probably decide it as an outlier if we try to specify a threshold for instant scores. In Figue 6 (a), we can observe the similar results. Thus, TOP-EYE can be more tolerate to noisy disturbance among the trajectories.

4. CONCLUDING REMARKS

In this paper, we provided an evolving trajectory outlier detection method, named TOP-EYE, which calculates the evolving outlying score by accumulating the instant outlying score with the mitigated past outlying scores via a decay function. This evolving score enables us easily to detect outliers by thresholding the evolving scores since only true outliers lead to high values of the evolving score. In contrast, it is much harder to specify a threshold for the instant scores to differentiate outliers and false alarms because instant scores often suddenly reach high values due to the noisy disturbance. Thus, TOP-EYE has the advantages of detecting trajectory outliers at the early stage and reducing false alarms. Experimental results on real-world trajectories showed that TOP-EYE could effectively detect evolving trajectory outliers with relatively low false alarm rate.



(a) False-alarm vehicle trajectory A in terms of direction and its evolving score



(b) False-alarm vehicle trajectory B in terms of direction and its evolving score

Figure 6: False alarm in terms of direction.

5. ACKNOWLEDGEMENTS

The authors were supported in part by National Science Foundation (NSF) via grant numbers CNS-0831186 and CCF-1018151, National Fundamental Research Program of China via grant number 2010CB327903 and the Jiangsu 333 Program. This work was also partly sponsored by WINLAB with industry affiliation program with Panasonic in pursuing data mining to abnormal behavior detection for security and safety applications.

6. REFERENCES

[1] Y. Bu, L. Chen, A. W.-C. Fu, and D. Liu. Efficient anomaly monitoring over moving object trajectory streams. In *SIGKDD*, 2009.

[2] J.-G. Lee, J. Han, and X. Li. Trajectory outlier detection: A partition-and-detect framework. In *ICDE'08*, 2008.

[3] X. Li, Z. Li, , J. Han, and J.-G. Lee. Temporal outlier detection in vehicle traffic data. In *ICDE*, 2009.

[4] D. Pokrajac, A. Lazarevic, and L. J. Latecki. Incremental local outlier detection for data streams. In *IEEE Symposium on CIDM*, 2007.

[5] X. Wang, K. Ma, G.-W. Ng, and W. E. L. Grimson. Trajectory analysis and semantic region modeling using nonparametric bayesian model. In *CVPR*, 2008.

[6] X. Wang, K. Tieu, and E. Grimson. Learning semantic scene models by trajectory analysis. In *ECCV*, 2006.

[7] M. Wu, X. Song, S. R. Chris Jermaine, and J. Gums. A lrt framework for fast spatial anomaly detection. In *SIGKDD*, 2009.