Anomalous Human Trajectory Detection Using Clustering Methods

Doi Thi Lan, Seokhoon Yoon
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan 44610, Republic of Korea
doilan151188@gmail.com, seokhoonyoon@ulsan.ac.kr

***Abstract*—Anomalous human trajectory detection is a critical task in security surveillance in working places. Many studies have been proposed and achieved specific results in abnormal trajectory detection over recent years. In this paper, a framework is proposed for detecting anomalies in human trajectories using clustering methods. In particular, we propose two anomaly detection methods using two different clustering algorithms: spectral clustering-based anomalous trajectory detection (SC-ATD) and DBSCAN-based anomalous trajectory detection (D-ATD). Firstly, clustering methods are used to find normal trajectory clusters in the dataset. Then input trajectory is detected whether it is an anomaly using found clusters. Besides, determining the input parameters of clustering methods is investigated in this work. With spectral clustering, we choose the number of clusters using the WB-index. With DBSCAN, a new cluster quality index (CQI) is proposed to find an appropriate value of the Eps parameter, directly affecting DBSCAN's quality. The proposed methods are evaluated on a real trajectory dataset: MIT Badge. The results show that both proposed methods detect anomaly trajectories with 77.89 % for spectral clustering and 80.83 % for DBSCAN in terms of F1-score.**

***Index Terms*—Anomalous trajectory detection, LCSS, Spectral clustering, DBSCAN**

I. Introduction

Anomalies in human trajectories often relate to urgent situations (e.i., accidents, violent attacks, terrorism and fire). Therefore, detecting anomaly trajectories may improve safety and instantly solve risks in working spaces. Many different methods have been proposed for detecting trajectory anomaly (e.i., distance-based, density-based, clustering-based methods).

Anomaly trajectory detection using clustering methods aims to discover normal behaviour clusters in historical trajectories of humans. The abnormality of a new trajectory is evaluated based on its relationship and found normal clusters. In order to identify abnormal trajectories from a taxi GPS dataset, Wang et al. [1] designed an anomalous trajectory detection approach applying a hierarchical clustering algorithm. The clusters of trajectories are first determined using the edit distance and the hierarchical clustering algorithm. Then, the clusters with just a trajectory have been labeled as anomalies. The authors of [2] proposed a two-phase framework for detecting abnormal trajectories. The hierarchical clustering method was also used to divide the dataset into clusters in the offline phase. If a trajectory wasn't connected to any clusters during the second phase, it was detected as an anomaly. With clustering methods, a challenge for researchers is determining the input parameters. Some methods have been proposed for selecting the input parameters [3]–[5]. If the input parameters are selected appropriately, the clustering performance is improved, and the anomaly detection efficiency is also boosted. Therefore, we focus on choosing the input parameters of clustering methods in this work. In particular, with spectral clustering [6], we use the WB-index, which finds the appropriate number of clusters based on the compactness within clusters and the separation between clusters [7]. With DBSCAN, two required input parameters are Eps (the radius to find neighbours) and Minpts (the minimum number of points to create a new cluster) [8]. In our work, a new cluster quality index called CQI is proposed to find the suitable value of Eps. Eps is an important parameter, and choosing it is challenging for DBSCAN. In contrast, Minpts can be selected more easily based dataset.

This paper contains three main contributions. Firstly, we proposed a framework for detecting abnormal trajectories in working places using clustering methods. In particular, two clustering-based anomaly detection methods are proposed: spectral clustering-based anomalous trajectory detection (SC-ATD) and DBSCAN-based anomalous trajectory detection (D-ATD). Secondly, we apply the WB-index for finding the number of clusters in spectral clustering and propose a novel index for selecting the Eps value in DBSCAN. Finally, a performance evaluation for the proposed framework is performed using the real trajectory dataset: MIT Badge.

II. Framework For Anomalous Trajectory Detection Using Clustering Methods

This section presents a framework for detecting trajectory anomalies based on clustering methods, as shown in Fig. 1. This framework contains two phases.

In Phase 1, the normal behaviour clusters are discovered using different clustering methods. In this step, a distance metric is required to determine the similarity between trajectories. In our work, the longest common sub-sequence (LCSS) is chosen [9]. LCSS may be applied to trajectories of different lengths. Besides, this metric is also robust to noise by using thresholds to find close points between two trajectories. A distance matrix of all historical trajectories in the dataset is calculated using LCSS. The cluster quality indices are used to determine the input parameters based on the distance matrix. After the

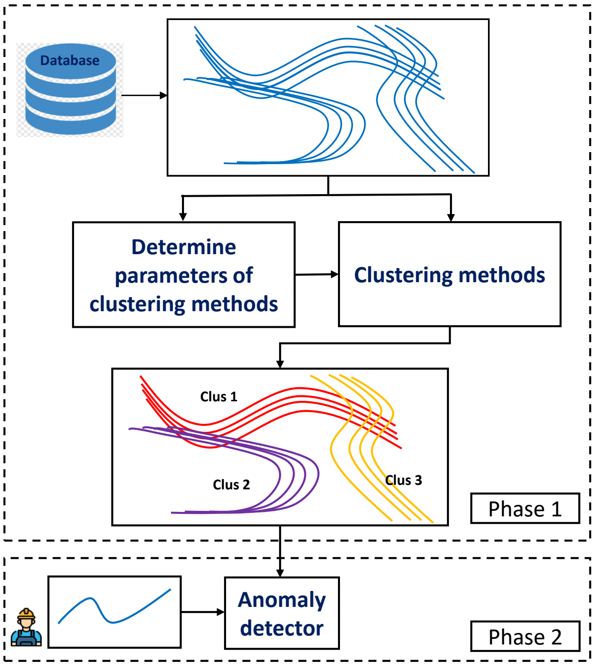

Fig. 1. Framework for anomaly detection using clustering methods.

parameters are selected, clustering methods are applied to find the normal clusters in the dataset.

In Phase 2, anomaly detection is performed when a new trajectory comes. The distances between the input trajectory and all clusters are calculated to determine whether it belongs to clusters. In particular, in SC-ATD, each cluster is modeled by a reference trajectory. If the distance between the input trajectory and the reference trajectory of the cluster is higher than a given distance threshold, the input trajectory does not belong to the cluster. In D-ATD, the neighbour number of the input trajectory is determined using the Eps value. The input trajectory belongs to the cluster if the neighbour number is equal to or higher than Minpts. In both SC-ATD and D-ATD, if the input trajectory does not belong to any clusters, it is detected as an anomaly.

III. Determine Input Parameters In Clustering Methods

A. The Number of Clusters in Spectral Clustering

In spectral clustering, the number of clusters is required to be set before performing the clustering algorithm. Finding the appropriate number of clusters is still challenging for spectral clustering. To address this problem, we use a cluster validation index called WB-index [7]. WB-index validates the performance of clustering methods using sum-of-squares within cluster (SSW) and sum-of-squares between clusters (SSB) as in (1):

$$WB(M) = M \times SSW/SSB, \tag{1}$$

where M is the number of clusters. If SSW is small, the compactness within clusters is strong. If SSB is large, the separation between clusters is high. To find the appropriate value of M, WB-index is calculated over a range of M. When WB-index reaches to the minimum value, M is selected.

B. The Eps Value in DBSCAN

To determine the value of Eps, a classic method is to find the knee point on k-dist graph [8]. In this paper, a new approach is proposed using the evaluation of DBSCAN's clustering performance for choosing the Eps value. The results of DBSCAN contain both clusters and outliers. Therefore, we use three elements for DBSCAN's performance evaluation: the compactness within clusters (CWC), the separation between outliers and clusters (SOC) and the separation between clusters (SBC). We define them as follows:

$$CWC(K) = Min_{k=\{1,\dots,\ K\}} IntraCD(C_k), \tag{2}$$

$$SOC(K) = Min_{o=\{1,2\dots,\ O\},k=\{1,\dots,\ K\}} OCD(o, C_k), \tag{3}$$

$$SBC(K) = Min_{l=\{1,\dots,\ K-1\},k=\{l+1,\dots,\ K\}} InterCD(C_l, C_k), \tag{4}$$

where $IntraCD(C_k)$ is average distance between trajectories in cluster C_k, $OCD(o, C_k)$ is the minimum distance between outlier o and trajectories of cluster C_k and $InterCD(C_l, C_k)$ is the average distance between trajectories in cluster C_l and cluster C_k. A new cluster quality index called CQI is proposed in (5):

$$CQI(K) = SBC(K) + SOC(K) - CWC(K), \tag{5}$$

where K is the number of found clusters by DBSCAN according to a specific value of Eps. To choose an appropriate value of Eps, CQI-index is determined over a range of Eps values. When CQI-index achieves the maximum value, the Eps value is selected.

IV. Performance Evaluation

A. Dataset

In this work, the MIT Badge dataset is used for the methods' performance evaluation. Since the trajectories are not labelled in the dataset, we give a hypothesis for anomalies to estimate the proposed methods, which is presented in the paper [10].

B. Results

This subsection presents the results of choosing input parameters in clustering methods and detecting anomaly trajectories.

Determining clusters' number in spectral clustering using WB-index is presented in Fig. 2. In this experiment, we evaluate WB-index when the number of clusters changes from 2 to 30. From Fig. 2, the selected number of clusters is 5. At this value, WB-index is minimum.

In DBSCAN, the Eps value is chosen in Fig. 3. The maximum value of CQI-index is obtained when $Eps = 0.35$. Therefore, this Eps value is selected for performing DBSCAN. In our work, Minpts is set to 6.

Results of anomaly trajectory detection are presented in Table I . We evaluate the performance of two proposed methods: SC-ATD and D-ATD, and compare them to the density

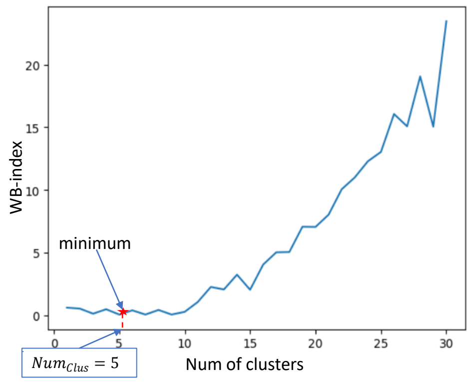

Fig. 2. Determine the number of clusters in spectral clustering.

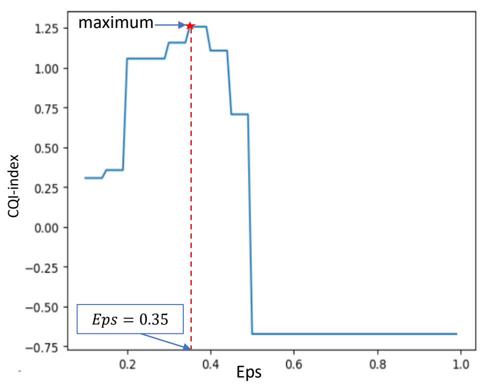

Fig. 3. Determine the value of Eps in DBSCAN.

method. Besides, the Euclidean distance and edit distance on real sequence (EDR) [11] are also estimated in this paper. From Table I, SC-ATD and D-ATD achieve better performance than the density method. This point may be explained that the input parameters of clustering methods used in the two formers are selected using the cluster quality indices. Therefore, their performance may be improved. In contrast, thresholds for anomaly detection in the density method are determined manually [12].

TABLE I
RESULTS FOR ANOMALY DETECTION IN HUMAN TRAJECTORY

Method	Distance metric	Recall	Precision	F1-score
Density	**Euclidean**	0.6048	0.6634	0.63
	EDR	0.6742	0.6556	0.6635
	LCSS	0.7115	0.7255	0.7164
SC-ATD	**Euclidean**	0.6535	0.881	0.7489
	EDR	0.7196	0.8016	0.7577
	LCSS	0.7874	0.7715	**0.7789**
D-ATD	**Euclidean**	0.5686	0.9257	0.7029
	EDR	0.6845	0.8632	0.7626
	LCSS	0.7771	0.8474	**0.8083**

With the distance metrics, Euclidean distance obtains the lowest performance over the three methods. Euclidean distance is the most straightforward metric for determining trajectories' distance. This metric requires trajectories of the same lengths and is sensitive to noises.

V. CONCLUSION

In this work, a framework for detecting anomalous human trajectories using clustering methods was proposed. In particular, we proposed two anomalous trajectory detection methods: SC-ATD and D-ATD. In SC-ATD, spectral clustering was applied to find normal trajectory clusters in the dataset. The WB-index is used for choosing the appropriate number of clusters in the spectral clustering. In D-ATD, we proposed CQI-index to select the Eps value of DBSCAN instead of choosing manually. CQI-index was determined based on the DBSCAN's quality evaluation. The proposed anomaly detection methods were estimated on the MIT Badge dataset. The results showed that our proposed methods detected human trajectory anomalies.

ACKNOWLEDGMENT

This work was supported in part by the Institute of Information and Communication Technology Planning and Evaluation (IITP) Grant by the Korean Government through MSIT (Development of 5G-Based Shipbuilding and Marine Smart Communication Platform and Convergence Service) under Grant 2020-0-00869, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2021R1I1A3051364.

REFERENCES

[1] Wang, Y.; Qin, K.; Chen, Y.; Zhao, P,“ Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data,” ISPRS Int. J. Geo-Inf. 2018, 7, 25.
[2] Ghrab NB, Fendri E, Hammami M,“ Abnormal events detection based on trajectory clustering,” in 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV) 2016 Mar 29 (pp. 301-306). IEEE.
[3] Aljarah I, Habib M, Nujoom R, Faris H, Mirjalili S,“A comprehensive review of evaluation and fitness measures for evolutionary data clustering,” Evolutionary Data Clustering: Algorithms and Applications. 2021:23-71.
[4] Liang B, Cai J, Yang H,“ A new cell group clustering algorithm based on validation & correction mechanism,” Expert Systems with Applications. 2022 May 1;193:116410.
[5] Ros F, Riad R, Guillaume S,“ PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” Neurocomputing. 2023 Apr 1;528:178-99.
[6] Jia H, Ding S, Xu X, Nie R,“ The latest research progress on spectral clustering,” Neural Computing and Applications. 2014 Jun;24:1477-86.
[7] Zhao Q, Fränti P, “WB-index: A sum-of-squares based index for cluster validity,” Data & Knowledge Engineering. 2014 Jul 1;92:77-89.
[8] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., “A density-based algorithm for discovering clusters in large spatial databases with noise,” in kdd, vol. 96, no. 34, 1996, pp. 226–231.
[9] M. Vlachos, G. Kollios, and D. Gunopulos, “Discovering similar multidimensional trajectories,” in Proceedings 18th international conference on data engineering. IEEE, 2002, pp. 673–684.
[10] Lan DT, Yoon S., “Trajectory Clustering-Based Anomaly Detection in Indoor Human Movement,” Sensors. 2023 Mar 21;23(6):3318.
[11] Chen L, Özsu MT, Oria V,“ Robust and fast similarity search for moving object trajectories,” in Proceedings of the 2005 ACM SIGMOD international conference on Management of data 2005 Jun 14 (pp. 491-502).
[12] J.-G. Lee, J. Han, and X. Li, “Trajectory outlier detection: A partition-and-detect framework,” in 2008 IEEE 24th International Conference on Data Engineering. IEEE, 2008, pp. 140–149.