



Trajectory Anomaly Detection Based on the Mean Distance Deviation

Xiaoyuan Hu¹, Qing Xu^{1(✉)}, and Yuejun Guo²

¹ College of Intelligence and Computing, Tianjin University, Tianjin, China
qingxu@tju.edu.cn

² Graphics and Imaging Laboratory, Universitat de Girona, Girona, Spain

Abstract. With the development of science and technology and the explosive growth of data, there will be a lot of trajectories every day. However, how to detect the abnormal trajectory from many trajectories has become a hot issue. In order to study trajectory anomaly detection better, we analyze the Sequential conformal anomaly detection in trajectories based on hausdorff distance (SNN-CAD) method, and propose a new measurement method of trajectory distance Improved Moved Euclidean Distance (IMED) instead of Hausdorff distance, which reduces the computational complexity. In addition, we propose a removing-updating strategy to enhance the conformal prediction (CP). Then, we also put forward our Non-conformity measure (NCM), Mean Distance Deviation. It can enlarge the difference between trajectories more effectively, and detect the abnormal trajectory more accurately. Finally, based on the technical measures mentioned above and under the framework of enhanced conformal prediction theory detection, we also build our own detector called Mean Distance Deviation Detector (MDD-ECAD). Using a large number of synthetic trajectory data and real world trajectory data on two detectors, the experimental results show that MDD-ECAD is much better than SNN-CAD in both accuracy and running time.

Keywords: Conformal prediction · Non-conformity measure · Trajectory distance

1 Introduction

With the rapid proliferation of closed circuit television cameras, satellites and mobile devices, massive trajectory based on different kinds of moving objects such as people, hurricanes, animals and vehicles [11] have been generated (Fig. 1). Undoubtedly trajectory data analysis plays a vital role and abnormal trajectory detection is one of the most key issues for this topic. There are more mature

This work has been funded by Natural Science Foundation of China under Grants No. 61471261 and No. 61771335. The author Yuejun Guo acknowledges support from Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya and the European Social Fund.

methods based on distance similarity [10]. Although it is relatively high in complexity in terms of large-scale data, it works well in small and medium-scale trajectory anomaly detection. The more famous method, Sequential conformal anomaly detection in trajectories based on hausdorff distance (SNN-CAD) method, is proposed by Laxhammar et al. [6]. Their detection method is mainly based on the conformal prediction (CP) theory [8]. Firstly, the Hausdorff distance [1] is used to calculate the distance between the trajectories as the trajectory similarity. Then Non-conformity Measure (NCM) [9] is given in the light of K-Nearest Neighbor [2], and finally uses conformal prediction detection theory to determine whether the trajectory is abnormal. However, the NCM can not distinguish the abnormal trajectory very well, and the detection accuracy is not high.

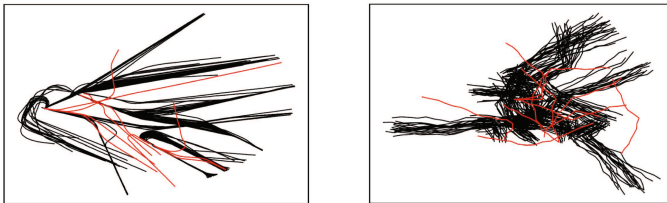


Fig. 1. A display of aircraft trajectories (left) and synthetic trajectories (right). Red trajectories are abnormal and black trajectories are normal. (Colour figure online)

In the view of this, we take into account the new NCM, Mean Distance Deviation (MDD), and present a removing-updating strategy to enhance conformal anomaly detection. Accordingly we propose the Mean Distance Deviation based on Enhanced Conformal Anomaly Detector (MDD-ECAD), which can deal with trajectory anomaly detection very well.

Also importantly, in this paper, we propose a new distance measure by improving Euclidean Distance (ED), which is called Improved Moved Euclidean Distance (IMED). It can characterize the trajectory distance efficiently. What is more, IMED does not require that the length of the trajectories must be same and its computational complexity is small.

The rest of this paper is organized as follows. Section 2 introduces the relevant background knowledge of our work. Section 3 presents the details of our method MDD-ECAD. Experimental data and results are described in Sect. 4. Finally, the paper is concluded in the Sect. 5.

2 Background

In this section, we will introduce the basic concept of trajectory and the specific details of CP theory.

2.1 Trajectory Type

In general, a trajectory data we study in this paper is a sequence of coordinate points in Cartesian coordinate system. Speaking ahead of time, a trajectory can be simply represented as $T = (a_1, a_2, \dots, a_n)$.

2.2 Conformal Prediction

Conformal prediction (CP) makes use of the past of experience to determine precise levels of confidence in new prediction. Generally speaking, assume a training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ where x_i ($i = 1, \dots, l$) is the input data, that is some data observed or collected by some means. And y_i ($i = 1, \dots, l$) is the output data, that is the label predicted by some method. For exame, x_i is a trajectory data collected by sensor and y_i is the label with only abnormal or normal type in the trajectory anomaly detection. Given a new observed data x_{l+1} , the basic idea of conformal prediction to estimate the p-value p_{l+1} of x_{l+1} by designed NCM according to training data. Finally, the p_{l+1} is compared with the pre-defined threshold ϵ to determine the label of x_{l+1} .

If $p_{l+1} < \epsilon$, x_{l+1} is identified as conformal anomaly. Otherwise, x_{l+1} is determined as normal. However, the key to estimate the p-value of the new example is how to design effective NCM. Next, we will introduce the concept of a Non-Conformity Measure (NCM) whose purpose is to measure the difference between the new example and a set of observed data.

Formally, NCM is a mathematical function. We can get a score α_i about the difference between the example x_i and the rest of dataset by a certain NCM. The score of x_i is given by

$$\alpha_i = A(X_{j \neq i}, x_i) \quad (1)$$

where X is a set of data; x_i is a example of dataset X ; $A(\cdot)$ is a form of NCM.

Based on formula (1), the score $(\alpha_1, \alpha_2, \dots, \alpha_{l+1})$ is gained. Then the p-value of x_{l+1} , p_{l+1} , is determined as the ratio of the number of trajectories that have greater or equal nonconformity scores to x_{l+1} to the total number of trajectories. The p-value is defined as follows:

$$p_{l+1} = \frac{|\{\alpha_i | \alpha_i \geq \alpha_{l+1}, 1 \leq i \leq l+1\}|}{l+1} \quad (2)$$

where $|\{\cdot\}|$ computers the number of elements in the set. CP will estimate a set of p-value to predict the lable of the new example and work excellently by using an effective NCM, especially when ϵ is close to the proportion of abnormal data in the dataset.

The Sequential Hausdorff Nearest Neighbor Conformal Anomaly Detector (SNN-CAD) method was developed by laxhammar et al. [6]. Their main contribution is to use Hausdorff distance to calculate the trajectory distance and use k-nearest neighbor as NCM. Suppose there are two sets of $T_a = \{a_1, a_2, \dots, a_m\}$,

$T_b = \{b_1, b_2, \dots, b_n\}$. The Hausdorff distance can refer to this article. As for NCM, it is defined as follows:

$$\alpha_i = \sum_{T_b \in \text{Neig}(T_a)} d(T_a, T_b) \quad (3)$$

Where $d(\cdot)$ is a kind of trajectory distance, $\text{Neig}(T_a)$ represents the k -nearest neighbor of T_a .

3 Our Method

3.1 Improved Moved Euclidean Distance

In order to measure the distance between two trajectories effectively, researchers have put forward various methods to calculate the distance. The most commonly used and famous ones are ED, HD, and DTW. However, comparing the advantages and disadvantages of the above three distances, we come to the following conclusions: (1) DTW and HD can handle the unequal length trajectory data, but the computational complexity is too high to deal with large and medium-sized data. (2) ED calculates the trajectory distance quickly with the simple implementation, but it can not do anything for the unequal trajectory data.

After our discussion, we can't help thinking about how to calculate quickly and deal with unequal data. For this purpose, based on ED, we propose a new distance measure Improved Moved Euclidean Distance (IMED) to enlarge the difference between trajectories for better performing trajectory anomaly detection. The proposed distance measure can manage both equal and unequal length trajectories. The basic idea is to fix the longer trajectory, moving the shorter trajectory backward until the longer trajectory is completely matched. Given two trajectories, $T_a = \{a_1, a_2, \dots, a_m\}$, $T_b = \{b_1, b_2, \dots, b_n\}$. Assuming $n \geq m$, the IMED is defined as follows:

$$d_{IME}(T_a, T_b) = \frac{\sum_{j=0}^{n-m} \sqrt{\sum_{i=1}^m \|b_{i+j} - a_i\|^2}}{n - m + 1} \quad (4)$$

especially, when $n = m$, T_a and T_b have the same trajectory length.

3.2 Mean Distance Deviation

An appropriate NCM is very critical and widely used for general anomaly detection. Generally, if a trajectory is similar to its neighboring trajectories, we can think that it is normal. Otherwise, if a trajectory is not the same as the trajectories around it, we can judge that the trajectory is abnormal. Actually, the employment of the local neighborhood is a fundamental consideration widely used in many anomaly detection methods, such as the classic KNN. In SNN-CAD, they use the sum of k -nearest neighbors of a trajectory as an indicator of comparison with other trajectories. The larger the value of KNN, the greater

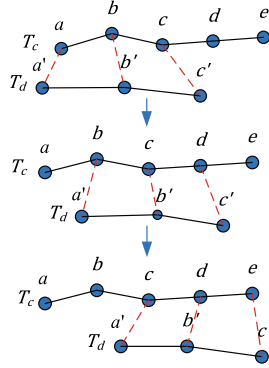


Fig. 2. An example of our IMED for unequal lengths.

the difference between the behavior of the trajectory and the surrounding trajectories. However, it is not ideal to use KNN to judge whether the trajectory is abnormal. For this reason, we propose a new NCM, Mean Distance Deviation (MDD). It is proved by the later experimental data (Sect. 5) that this method is much better than KNN in trajectory anomaly detection. Now we will give its specific definition (Fig. 2):

$$\alpha(T_a) = \sqrt{\frac{\sum_{T_b \in \text{Neig}(T_a)} (MD(T_a) - MD(T_b))^2}{k}} \tag{5}$$

where

$$MD = \frac{\sum_{T_b \in \text{Neig}(T_a)} d(T_a, T_b)}{k} \tag{6}$$

3.3 Removing-Updating Strategy

The process of anomaly detection based on CP is to calculate the p-value of each data, and then compare with the given threshold value ϵ to determine whether the trajectory is abnormal. Because the abnormal data detected last time will interfere with this detection, we propose a removing-updating strategy to CP.

Specifically, when calculating the p-value of all data, the most abnormal data will be removed. Then update the threshold and repeat the above process with the remaining data until the threshold is 0.

4 Experiment

In this section, in order to evaluate the effect of MDD-ECAD and IMED, we compared MDD-ECAD algorithm with SNN-CAD algorithm, as well as several distances IMED, HD and DTW based on the synthetic data and real life data.

4.1 Data Sets

Synthetic trajectories I [5] presented for anomaly detection is created by laxhammar et al. [6] using the trajectory generator software. It includes 100 datasets with 2000 trajectories in each dataset, about 1% of which are abnormal trajectories. In addition, each trajectory is composed of a series of two-dimensional coordinate points. To expand the dataset for experiment, we use another **Synthetic trajectories** [3] including **synthetic trajectories II**, **synthetic trajectories III**, **synthetic trajectories IV**. The three synthetic trajectories each contain 100 trajectory datasets with ϵ equal to 0.05, 0.01, and 0.02. And each dataset has 2000 trajectories with the number of sample points ranging from 20 to 100.

Aircraft trajectories [7] has in all 470 two-dimensional trajectories, involving 450 normal and 20 abnormal ones. And the trajectory length in the set varies from 12 to 171 sampling points.

4.2 Performance Measure

Trajectory anomaly detection is also two classification problem. Therefore, we can use the following evaluation indicators: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Precision (P), Recall (R) and F1 are used to test the classification accuracy. F1 score is used to evaluate the effect of all experiments. The larger F1 value is, the better the algorithm effect is.

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F1 = \frac{2 * P * R}{P + R} \quad (7)$$

4.3 Experimental Results and Analysis

Table 1. The F1 results (%) of synthetic trajectory datasets with two methods.

Synthetic trajectory datasets	Approaches	P	R	F1
Synthetic trajectories I	SNN-CAD	87.15	77.48	79.80
	MDD-ECAD	88.65	91.83	89.11
Synthetic trajectories II	SNN-CAD	83.82	65.04	69.40
	MDD-ECAD	85.8	91.09	86.35
Synthetic trajectories III	SNN-CAD	89.31	74.31	79.31
	MDD-ECAD	92.45	90.96	90.76
Synthetic trajectories IV	SNN-CAD	87.47	83.88	84.62
	MDD-ECAD	91.37	94.61	92.31

In the experiment, we mainly compare the performance of SNN-CAD and MDD-ECAD. It can be seen from Table 1 that the F1 of MDD-ECAD is 89.11%,

Table 2. The F1 results (%) of real-life trajectory datasets with different methods.

Trajectory datasets	MDD-ECAD	SNN-CAD	iVAT+
Aircraft trajectories	95	75	90

86.35%, 90.76% and 92.31% respectively, higher than that of SNN-CAD. For testing our method on complex real life data, Table 2 shows MDD-ECAD still outperforms SNN-CAD and iVAT+ [4]. The F1 of MDD-ECAD is as high as 95%, while the SNN-CAD is only 75% and iVAT+ is 90%. The reason why SNN-CAD doesn't perform excellent may be that its NCM can't amplify the abnormal trajectory behavior greatly, and the MDD we used can make up for this defect very well. In addition to the problem of detection framework, we use the removing and updating strategy to avoid the secondary interference of obvious abnormal trajectories to others.

Table 3. The F1 results (%) of MDD-ECAD with different distance measures.

Trajectory datasets	IMED	HD	DTW
Synthetic trajectories I	89.11	89.30	79.58
Synthetic trajectories II	86.35	84.80	71.86
Synthetic trajectories III	90.76	82.43	80.01
Synthetic trajectories IV	92.31	90.96	82.65
Aircraft trajectories	95	85	90

Table 4. Runtimes (s) of MDD-ECAD with different distance measures.

	IMED	HD	DTW
Synthetic trajectories I	0.591	6.22	11.86
Synthetic trajectories II	16.88	85.53	97.43
Synthetic trajectories III	17.03	109.06	100.49
Synthetic trajectories IV	17.72	89.03	100.84
Aircraft trajectories	2.39	19.55	18.97

In order to compare the performance of IMED, HD and DTW, we use the three distances in MDD-ECAD method. Table 3 shows that the F1 of IMED is higher than HD and DTW, which indicates that IMED can measure the distance between trajectories better. In addition, the running time of IMED, HD and DTW is given (see Table 4), and it is obvious that the running time of IMED is fewer. From the theoretical analysis, IMED has the minimal computational complexity and no doubt runs the fastest. The experiment just verifies this point.

5 Conclusion

In this paper, in order to improve performance of SNN-CAD, we propose a new method to calculate the trajectory distance. An excellent Non-conformal measurement and a removing-updating strategy are also used for our anomaly detector. Large number of experimental data shows that our detector is better than SNN-CAD.

References

1. Dubuisson, M.P., Jain, A.K.: A modified Hausdorff distance for object matching. In: Proceedings of 12th International Conference on Pattern Recognition, vol. 1, pp. 566–568. IEEE (1994)
2. Güting, R.H., Behr, T., Xu, J.: Efficient k-nearest neighbor search on moving object trajectories. *VLDB J.* **19**(5), 687–714 (2010)
3. Guo, Y., Bardera, A.: SHNN-CAD+: An improvement on SHNN-CAD for adaptive online trajectory anomaly detection. *Sensors* **19**(1), 84 (2019)
4. Kumar, D., Bezdek, J.C., Rajasegarar, S., Leckie, C., Palaniswami, M.: A visual-numeric approach to clustering and anomaly detection for trajectory data. *Vis. Comput.* **33**(3), 265–281 (2015). <https://doi.org/10.1007/s00371-015-1192-x>
5. Laxhammar, R.: Synthetic trajectories (2013)
6. Laxhammar, R., Falkman, G.: Sequential conformal anomaly detection in trajectories based on Hausdorff distance. In: 14th International Conference on Information Fusion, pp. 1–8. IEEE (2011)
7. Leader, D.S.G.: Aircraft trajectories. <https://c3.nasa.gov/dashlink/resources/132/>
8. Shafer, G., Vovk, V.: A tutorial on conformal prediction. *J. Mach. Learn. Res.* **9**(March), 371–421 (2008)
9. Smith, J., Nouretdinov, I., Craddock, R., Offer, C., Gammerman, A.: Anomaly detection of trajectories with kernel density estimation by conformal prediction. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H., Sioutas, S., Makris, C. (eds.) *AIAI 2014. IAICT*, vol. 437, pp. 271–280. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44722-2_29
10. Toohey, K., Duckham, M.: Trajectory similarity measures. *Sigspatial Special* **7**(1), 43–50 (2015)
11. Zheng, Y.: Trajectory data mining: an overview. *ACM Trans. Intell. Syst. Technol. (TIST)* **6**(3), 1–41 (2015)