

Anomalous Behavior Detection in Trajectory Data of Older Drivers

Seyedeh Gol Ara Ghoreishi
Florida Atlantic University
Boca Raton, USA
sghoreishi2021@fau.edu

Sonia Moshfeghi
Florida Atlantic University
Boca Raton, USA
smoshfeghi2021@fau.edu

Muhammad Tanveer Jan
Florida Atlantic University
Boca Raton, USA
mjan2021@fau.edu

Joshua Conniff
Florida Atlantic University
Boca Raton, USA
fau.jconniff@health.fau.edu

KwangSoo Yang
Florida Atlantic University
Boca Raton, USA
yangk@fau.edu

Jinwoo Jang
Florida Atlantic University
Boca Raton, USA
jangj@fau.edu

Borko Furht
Florida Atlantic University
Boca Raton, USA
bfurht@fau.edu

Ruth Tappen
Florida Atlantic University
Boca Raton, USA
rtappen@health.fau.edu

David Newman
Florida Atlantic University
Boca Raton, USA
dnewma14@fau.edu

Monica Rosselli
Florida Atlantic University
Boca Raton, USA
mrossell@fau.edu

Jiannan Zhai
Florida Atlantic University
Boca Raton, USA
jzhai@fau.edu

Abstract—Given a road network and a set of trajectory data, the anomalous behavior detection (ABD) problem is to identify drivers that show significant directional deviations, hard-brakings, and accelerations in their trips. The ABD problem is important in many societal applications, including Mild Cognitive Impairment (MCI) detection and safe route recommendations for older drivers. The ABD problem is computationally challenging due to the large size of temporally-detailed trajectories dataset. In this paper, we propose an Edge-Attributed Matrix that can represent the key properties of temporally-detailed trajectory datasets and identify abnormal driving behaviors. Experiments using real-world datasets demonstrated that our approach identifies abnormal driving behaviors.

Keywords—Trajectory data mining, abnormal detection and spatio-temporal network database.

I. INTRODUCTION

Given a road network and a set of trajectory data, the anomalous behavior detection (ABD) problem is to identify drivers that show significant directional deviations, hard-brakings, and accelerations in their trips. Consider an example input of ABD in Figure 1a. Each graph represents a trip on the road network and the table shows the temporally detailed driving information (i.e., speed, length, direction and the number of accelerations and brakes) for edges. Figure 1b shows an example output of ABD containing the abnormality score for each driver. One of the most appealing properties of our model is the ability to capture geometrical (e.g., longitude and latitude), directional, and topological features of the trajectories. Analyzing this data presents a considerable challenge due to the large size of temporally-detailed trajectory datasets.

A. Application Domain

ABD problem could be linked to identifying cognitive impairment in the elderly drivers¹. Older drivers face many challenges during their trips. They might miss intended routes or exits, make incorrect turns, and go the wrong way due to reduced spatial awareness or the need to remember the way when trying to reach their destination. Some trips had cyclic patterns and deviated significantly from directed paths. However, not all deviations are necessarily abnormal. It is possible for drivers to make turns and take alternative routes due to various reasons such as roadblocks, traffic congestion, or emergencies [1]. Furthermore, some drivers might choose the correct routes during their journeys, but they exhibit frequent sudden accelerations. These abrupt speed changes could also signal confusion or a lack of awareness regarding the road network conditions.

B. Problem Definition

In our problem formulation, a trip is represented as a directed graph consisting of nodes, edges and edge attributes. Each node represents a spatial location on the road network and each edge represents a road segment. Every edge has temporally-detailed driving attributes. The ABD problem can be formalized as follows:

Input:

- A road network consisting of a set of nodes V and a set of edges E ,
- a set of trip attributes on a road segment $A : E \rightarrow \mathbb{R}^+_0$, and

¹<https://www.fau.edu/newsdesk/articles/older-drivers-research.php>

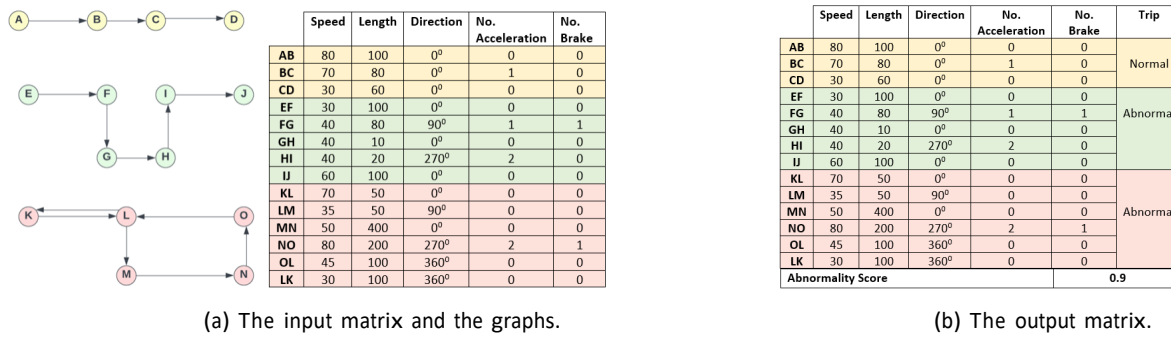


Fig. 1: An example of the input and output of ABD -three trips with its corresponding input and output matrices.

- the minimum length of a trip α

Output: Abnormal driving behaviors

Objective:

- Compute the abnormality scores related to direction deviations, hard brakings, and hard accelerations.

Constraints:

- Every length of the trip should be greater than α .

C. Outline

The next of this paper is organized as follows: Section 2 describes the related work. Section 3 introduces the proposed method. Section 4 illustrates the experimental dataset and results. Section 5 discusses future directions and concludes the research work.

II. RELATED WORK

Numerous researchers have analyzed the GPS trajectories of vehicles in a road network and calculated the similarity between different trajectories. However, most of their attention has been on vehicles traveling from one specific starting point to a particular destination. Researchers apply various methods to identify abnormal behaviors in these driving patterns. In paper [2], GPS points are aggregated into spatiotemporal units, where each unit represents the maximum value of speed, acceleration, and direction deviation observed within a specific road segment. The authors compared these units among drivers to identify those showing significant deviations from typical traffic patterns and proposed a variational autoencoder method named STDTB-AD to quantify the level of abnormality for each driver. The researchers in [3] used trajectory data to predict the traffic context on road networks and created a model to represent road segments and capture characteristics of road networks. Another paper, [4] proposed a partition and detect approach to achieve anomalous sub-trajectory identification. The process begins by splitting each trajectory into t-partitions. Then, it detects segments that significantly differ from those in other trajectories in terms of distance and shape. A trajectory with significant abnormal segments contains an outlier sub-trajectory and is classified as an outlier. In [5], a self-supervised proposed approach creates low-dimensional vectors from raw trajectories. The method first presents raw trajectory data and road networks as road segment vectors

and then combines these vectors into a single representation called trajectory vectors. This transformation is achieved by maintaining spatial-temporal characteristics of the trajectories. In [6], each trajectory is segmented to a minimum number of homogeneous segments with similar spatiotemporal characteristics, such as location, heading, speed, velocity, sinuosity, curviness, and shape. Authors in [7] clustered trajectories by using the iVAT algorithm and employing a two-stage process. In the first stage, a similarity measure is applied to group trajectories that follow similar paths but may have opposite starting and ending points. These similar trajectories are assigned to the same cluster. In the second stage, directional similarity is used within each cluster to separate trajectories going in opposite directions. From these clustered trajectories, the proposed approach identify trajectory outliers by finding trajectories significantly distant from others in the same cluster or by detecting clusters with too few trajectories. This process helps identify unusual or isolated trajectory patterns within the dataset. In [8], authors proposed a distance-based method to establish local clusters for continuous trajectory streams and use pruning techniques to monitor anomalous behavior. In [9], the main idea is to detect fraud in taxi driving by combining density and distance characteristics. The similarity is that first, they compute the expected distance of the most common routes and the distances a trajectory differs from the norm. In Graph-based methods for identifying outliers in trajectories, in [10], the researchers depict various trajectory trips but only for trips with the same starting and ending points. They analyzed the common nodes and edges between these graphs to detect trajectory outliers. If a graph has edges that are notably different from the usual patterns observed in other graphs or contains a different subgraph between the source and target graph, it is considered as an outlier. A machine learning algorithm employs the extracted features to classify the trajectory. In [11], researchers detected newly generated cycles in a dynamic graph that is constantly changing. Regarding detour detection, in [12] for the same start and end point, a detour is defined as taking much time or driving long distances. The authors in [13] proposed a graph-based method to detect loop closure for localization and mapping. For abnormal taxi trajectory, authors of [14] used detour detection, local shape

and speed anomaly detection by using spatial location, sequence, and behavioral features in vehicle trajectories. In their approach, detour detection is based on how much a trajectory changes direction at a specific point in its path. In [15], a distance metric is proposed to measure the similarity between anomalous and normal trajectories in different driving patterns, such as long-distance detours. They establish definitions for global detour and local detour. A global detour refers to a considerably longer trajectory that significantly deviates from the norm, while a local detour is a longer trajectory with partial deviations from the norm. Anomalous behavior like detours in [1] are based on the density of heading changes and the combination of sequential turns in the same direction. Their method automatically identifies a detour's start and end points and computes the detour factor while retaining the dynamic aspects of trajectory data. To achieve this, they employ a recursive Bayesian filter to estimate a single unobserved state and make predictions.

III. METHODOLOGY

We propose a framework to identify unusual driving patterns or behaviors that deviate from the expected norms during a trip. The primary objective is to evaluate sudden speed changes (like rapid slowing down or speeding up) and the paths a driver chooses in a trip. This evaluation relies on individual trajectory data points and subsequently assigns a numerical score reflecting the abnormality of the trip. In our study, a trip characterized by a long-distance cyclic pattern or a looping behavior should receive a high abnormality score. However, it is evident that all deviations, such as roundabouts or short-distance U-turns, are not abnormal or unusual behavior. As well as, many driving scenarios require drivers to make long detours or take alternative routes for various reasons, such as roadblocks or obstacles, traffic congestion, or emergencies [1]. In these cases, in our framework, apart from the length of the deviation paths, we also need to consider the speed changes during these detours. This is because an increase in speed during deviation paths suggests that the drivers have an acceptable reason. Besides, deviations allow them to drive at higher speeds and reach their destination more quickly. Our method includes data preprocessing, spatial graph modeling, and anomaly detection.

A. Data preprocessing

Since our main goal is to concentrate on the drivers' paths, it is necessary to preprocess temporally-detailed trajectory data within the road network. The temporally-detailed trajectory data contains GPS points, and we project them onto their nearest road segments to identify which points are located on which roads. This process enables us to explain the movement patterns of a trajectory throughout time and space. To model the trajectory data and their topological connectivity, we create a spatial graph for each driver's trip. Fig. 2a shows an edge-attributed graph consisting of 6 nodes and 5 edges, and each edge is associated with 3 attributes that provide essential indications about the associated road segments. The features

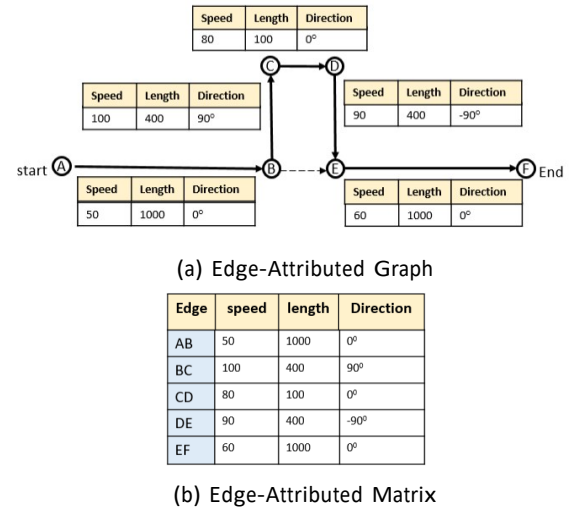


Fig. 2: Example of a trip with a set of trajectory attributes

include the average direction and speed of trajectory data and the segment's length. To simplify the example, only three attributes are selected for this illustration. These attributes are further presented in matrix format in Fig. 2b.

- Definition 1 (Trajectory data points). In trajectory data, each point p_i represents the trajectory's state at a particular time. This point is characterized by properties, including location, speed, direction and occurrence of hard acceleration and hard brake which is shown as:

$$p_i^t = \begin{matrix} n & o \\ \text{lat}_i^t, \text{lon}_i^t, v_i^t, d_i^t, \text{acc}_i^t, \text{br}_i^t \end{matrix}, \quad (1)$$

where lat is latitude, lon is longitude, v is velocity, d is direction, acc is acceleration, br is brake, superscript t represents the time t , and subscript i represents the id of the point.

- Definition 2 (Edge-Attributed Matrix). Every trip is represented by a directed graph. Each edge has a set of attributes that are listed in a matrix.

- Definition 3 (Edge attributes). The edge attributes are the average speed and direction, length and the total number of hard brakes and hard accelerations observed on its corresponding road segment over time.

B. Spatial graph model

With the Edge-Attributed Matrix, we can examine a variety of spatial graphs with various source and destination points. For a given set of graphs, we compute the anomaly scores using the isolation Forest algorithm to determine which graphs are anomalous and differ from the others in terms of attributes.

C. Anomaly detection using Isolation Forest (iForest)

Isolation Forest is an unsupervised anomaly detection algorithm to find rare and distinct data points within a dataset. It achieves this by constructing multiple binary trees [17]. These

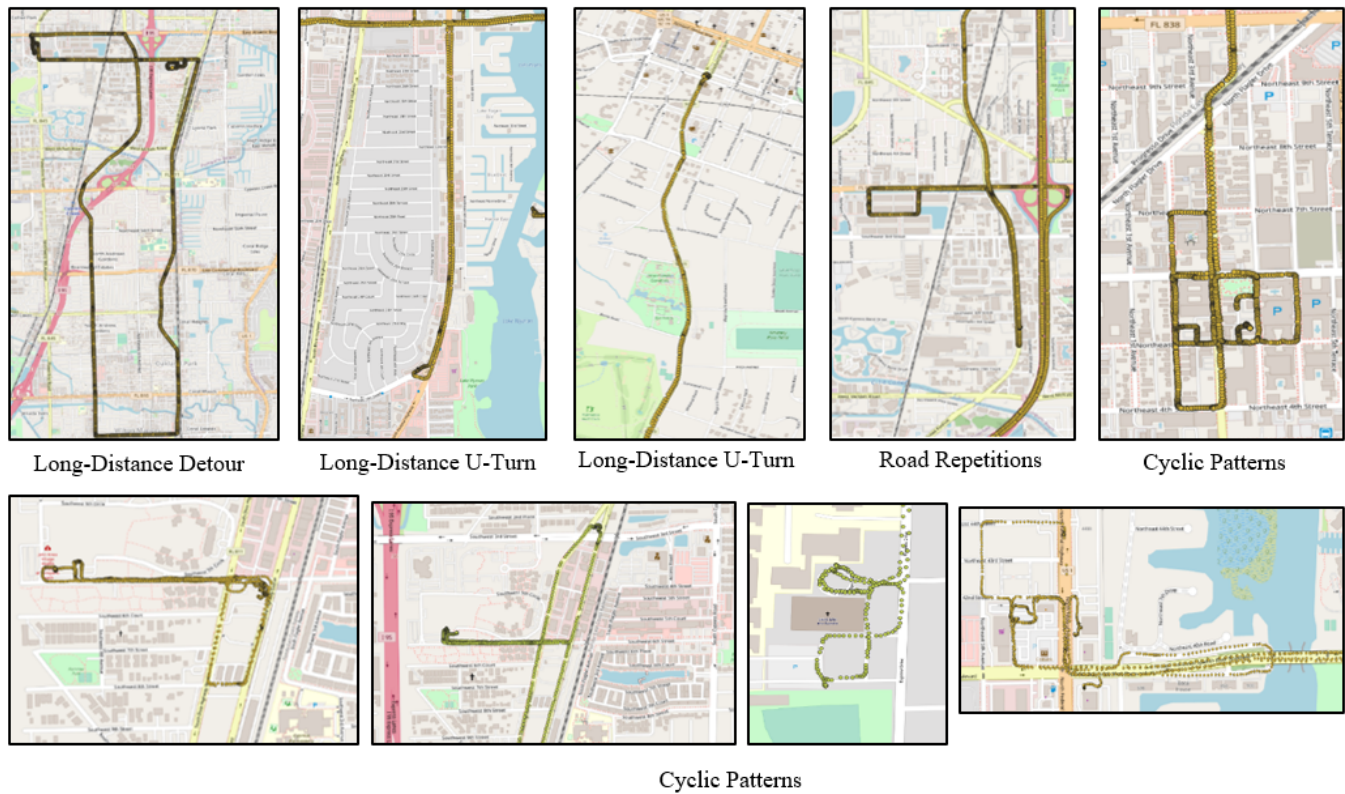


Fig. 3: Snapshot of the Real Dataset - A visual representation of the dataset used in the study, with each picture representing a random trip

trees randomly select features and values to divide the data into smaller groups through a recursive process [17]. Anomalies often require fewer splits to be separated in these trees because they are distinct from regular data. In this algorithm, the contamination parameter serves as a threshold for defining outliers in the dataset. Isolation Forest is useful for finding outliers and anomalies in various datasets, including those with unbalanced anomaly distributions [17].

IV. EXPERIMENTS

In this section, we describe the generated data and display the findings with the iForest algorithm.

A. Data

Our 5-year research project funded by the National Institutes of Health (NIH) comprises drivers with an age group between 65 and 85 years old [16]. In our dataset, the participant's behavior is monitored over 3 years. We collect 18 senior citizens' trajectory data between May 1, 2022 and July 30, 2022. The data is recorded with in-vehicle telematics sensors. The telematics sensor provides Inertial Measurement Unit (IMU) and GPS data. IMU data records the vehicle's dynamic motions and orientations. It consists of two components:

- 3-Axial accelerometer: This device records how the vehicle accelerates or decelerates in three directions (X, Y,

Z). It detects events like hard braking (sudden stops) or hard accelerations (rapid speed increases) ².

- 3-Axial Gyroscope (Angular Velocity): This measures the rate at which the vehicle is rotating or changes in direction ². Fig.4 shows the Accelerometer and Gyroscope ².

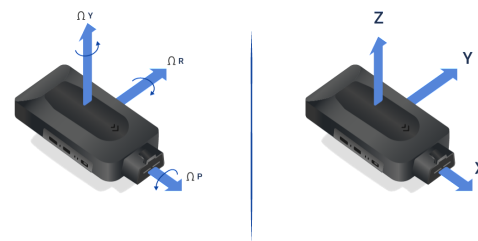


Fig. 4: Gyroscope sensing angular orientation and Accelerometer sensing Axis orientation ²

GPS Data: GPS data provides information about the vehicle's location and movement in the global positioning system. It includes several key components: timestamp, latitude, longitude, and altitude to analyze travel distance and Course Over Ground (COG) to calculate the heading or direction in which the vehicle is moving.

²<https://www.autopi.io/glossary/accelerometer/>

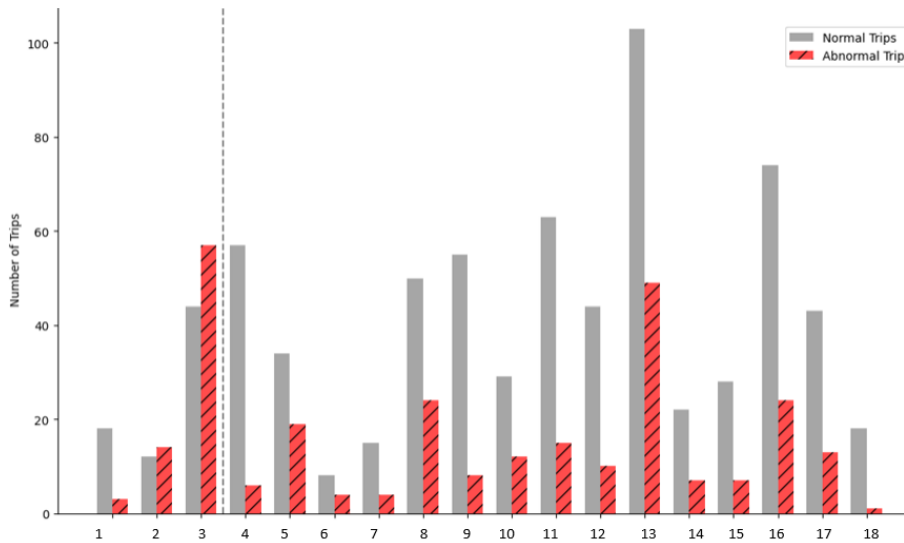


Fig. 5: Number of normal and abnormal trips for each sample

In a road network, the trajectories exhibit various driving patterns, and Fig.3 illustrates several snapshots that correspond to different drivers, each depicting potential abnormal driving behaviors among the elderly. These examples capture situations such as repetition on the road, lengthy detours, U-turns, cyclic patterns, and instances where drivers seem to have lost their way. Road segments comprise their trajectory points across time and by using the raw dataset, we compute statistics. The average velocity and COG, the sum of hard brakes and hard accelerations of data points with road segment length were used as input data. We apply this processing method and iForest to the dataset and obtain a result at different spatial scales.

V. RESULTS

The iForest algorithm calculates trip anomaly scores ranging from 0 to 1. Values close to 1 are considered as anomalous. In this study, the threshold to classify trips is set to 0.6. Our approach extends to driver classification, where the mean anomaly score for all trips informs whether a driver is classified as normal or abnormal. We define the threshold for drivers based on the top 10% and 20% [16] of the total drivers' anomaly scores. The contamination parameter, which indicates the likelihood of outliers in the dataset, is set to 0.2. Fig.5 presents the count of normal and abnormal trips for each driver. The vertical line indicates the top 20% threshold within the drivers' anomaly score distribution and distinguishes the 3 abnormal drivers.

The dataset employed in our experiments contains labels only for drivers, and there are no labels for individual trips. Besides, this dataset is unique and has not been previously examined in the literature. Consequently, making direct comparisons with existing studies is challenging due to this disparity in labeling. To evaluate the effectiveness of our approach in detecting abnormal drivers, we employed evaluation metrics of accuracy(ACC) and F1-score. These metrics encompass True

TABLE I: Anomaly detection metrics

Contamination		Top10%	Top20%
0.2	ACC	0.83	0.77
	F1	0.40	0.50

Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. ACC quantifies the percentage of correctly classified cases throughout the entire dataset, precision (P) assesses the proportion of true positives among all predicted positives and recall (R) measures the proportion of true positives among actual positives. It was shown in the literature that F1-score is more suitable than the accuracy [18].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (5)$$

Table 1 illustrates the relationship between the chosen thresholds and performance metrics. The accuracy is notably higher when the top 10% of drivers are considered. However, for the top 20%, the improvement in F1 score suggests that selecting a broader group of drivers results in a better balance between precision and recall.

VI. CONCLUSION AND FUTURE WORK

We discussed the problem of automatically identifying abnormal driving habits among senior individuals. The problem is challenging because of the large size of the road network

and trajectory data. By utilizing in-vehicle sensor data, edge-attributed matrices, and isolation forest algorithms, we aim to provide a baseline approach for identifying unusual actions, such as cyclic routes and long-distance detours, and high acceleration or braking incidents, which may indicate cognitive decline or impairments in older drivers. However, there is still room for improvement and additional study in this field.

In future work, we plan to expand our dataset by incorporating more samples of subjects for the test. Additionally, we intend to design and evaluate a novel algorithm for the notable characteristics of senior drivers.

VII. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation CAREER under Grant No. 1844565. and the National Institutes of Health under Grant No. 1R01AG068472.

REFERENCES

- [1] Huang, Hai. "Anomalous behavior detection in single-trajectory data." *International Journal of Geographical Information Science* 29, no. 12 (2015): 2075-2094.
- [2] Yu, Wenhao, and Qinghong Huang. "A deep encoder-decoder network for anomaly detection in driving trajectory behavior under spatio-temporal context." *International Journal of Applied Earth Observation and Geoinformation* 115 (2022): 103115.
- [3] Chen, Yile, Xiucheng Li, Gao Cong, Zhifeng Bao, Cheng Long, Yiding Liu, Arun Kumar Chandran, and Richard Ellison. "Robust road network representation learning: When traffic patterns meet traveling semantics." In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 211-220. 2021.
- [4] Lee, Jae-Gil, Jiawei Han, and Xiaolei Li. "Trajectory outlier detection: A partition-and-detect framework." In *2008 IEEE 24th International Conference on Data Engineering*, pp. 140-149. IEEE, 2008.
- [5] Li, Shuzhe, Wei Chen, Bingqi Yan, Zhen Li, Shunzhi Zhu, and Yanwei Yu. "Self-supervised contrastive representation learning for large-scale trajectories." *Future Generation Computer Systems* (2023).
- [6] Buchin, Maike, Anne Driemel, Marc J. van Kreveld, and Vera Sacristán. "Segmenting trajectories: A framework and algorithms using spatiotemporal criteria." *Journal of Spatial Information Science* 3 (2011): 33-63.
- [7] Kumar, Dheeraj, James C. Bezdek, Sutharshan Rajasegarar, Christopher Leckie, and Marimuthu Palaniswami. "A visual-numeric approach to clustering and anomaly detection for trajectory data." *The Visual Computer* 33 (2017): 265-281.
- [8] Bu, Yingyi, Lei Chen, Ada Wai-Chee Fu, and Dawei Liu. "Efficient anomaly monitoring over moving object trajectory streams." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 159-168. 2009.
- [9] Ge, Yong, Hui Xiong, Chuanren Liu, and Zhi-Hua Zhou. "A taxi driving fraud detection system." In *2011 IEEE 11th International conference on data mining*, pp. 181-190. IEEE, 2011.
- [10] Ahmed, Usman, Gautam Srivastava, Youcef Djenouri, and Jerry Chun-Wei Lin. "Knowledge graph based trajectory outlier detection in sustainable smart cities." *Sustainable Cities and Society* 78 (2022): 103580.
- [11] Qiu, Xiafei, Wubin Cen, Zhengping Qian, You Peng, Ying Zhang, Xuemin Lin, and Jingren Zhou. "Real-time constrained cycle detection in large dynamic graphs." *Proceedings of the VLDB Endowment* 11, no. 12 (2018): 1876-1888.
- [12] Jiang, Jinsheng, Chong Xu, Jian Xu, Ming Xu, Ning Zheng, and Kaiwei Kong. "Route planning for locations based on trajectory segments." In *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, pp. 1-8. 2016.
- [13] Han, Dongxiao, Yuwen Li, Tao Song, and Zhenyang Liu. "Multi-objective optimization of loop closure parameters for indoor 2D simultaneous localization and mapping." *Sensors* 20, no. 7 (2020): 1906.
- [14] Zhang, Haiyan, Yonglong Luo, Qingying Yu, Liping Sun, Xuejing Li, and Zhenqiang Sun. "A framework of abnormal behavior detection and classification based on big trajectory data for mobile networks." *Security and Communication Networks* 2020 (2020): 1-15.
- [15] Wang, Jingwei, Yun Yuan, Tianle Ni, Yunlong Ma, Min Liu, Gaowei Xu, and Weiming Shen. "Anomalous trajectory detection and classification based on difference and intersection set distance." *IEEE Transactions on Vehicular Technology* 69, no. 3 (2020): 2487-2500.
- [16] Jan, Muhammad Tanveer, Sonia Moshfeghi, Joshua William Conniff, Jinwoo Jang, Kwangsoo Yang, Jiannan Zhai, Monica Rosselli, David Newman, Ruth Tappen, and Borko Furht. "Methods and Tools for Monitoring Driver's Behavior." In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1269-1273. IEEE, 2022.
- [17] Ding, Zhiguo, and Minrui Fei. "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window." *IFAC Proceedings Volumes* 46, no. 20 (2013): 12-17.
- [18] Neghabi, Mehrnoosh, Hamid Reza Marateb, and Amin Mahnam. "Comparing steady-state visually evoked potentials frequency estimation methods in brain-computer interface with the minimum number of eeg channels." *Basic and Clinical Neuroscience* 10, no. 3 (2019): 245.