



Master Thesis

Thesis Title: Concise and Engaging Title

by

Mateusz Kędzia
(2666752)

Supervisor: Ronald Siebes (VU Amsterdam)

Daily Supervisor: Jiancheng Weng (Beijing University of Technology)

Internal Advisor: Zhisheng Huang (VU Amsterdam)

External Advisor: Shuai Wang (VU Amsterdam/Maastricht University)

Second Reader: Name and Surname

June 4, 2025

Submitted in partial fulfillment of the requirements for
the VU degree of Master of Science in Artificial Intelligence

你好，世界！ Contribution Title

Mateusz Kędzia¹[0009–0001–4296–4479]

¹ Vrije Universiteit Amsterdam, Amsterdam

² Vrije Universiteit Amsterdam, Amsterdam, The Netherlands lncs@springer.com
<http://www.springer.com/gp/computer-science/lncs>

³ Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
{abc,lncs}@uni-heidelberg.de

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: First keyword · Second keyword · Another keyword.

1 Literature Review

This literature review examines key contributions to vehicle trajectory anomaly detection, focusing on methods that are particularly relevant to our proposed research on combining isolation forest techniques with knowledge graph representations for enhanced anomaly detection in Beijing taxi trajectory data.

Isolation Forest for Trajectory Anomaly Detection Zhang et al. introduced the iBat (isolation-based anomaly detection for taxi trajectories) framework, which applies isolation forest principles to detect anomalous taxi trajectories in Beijing. Their work demonstrates that isolation forests can effectively identify outliers by isolating anomalous points through random partitioning, requiring fewer partitions for anomalous trajectories than normal ones. This foundational work provides direct relevance to our proposed methodology, establishing the effectiveness of isolation-based approaches for taxi trajectory data in Beijing. The paper validates the approach on real taxi GPS data, showing superior performance compared to density-based methods while maintaining computational efficiency.

Density-Based Clustering Approaches He et al. present an improved density clustering method combined with pattern information mining for anomaly trajectory identification. Their approach uses enhanced DBSCAN clustering with density values (DENM) and incorporates multiple distance metrics including DTW and Hausdorff distances. While this work focuses on density-based methods rather than isolation forests, it provides valuable insights into trajectory representation and similarity measures that could complement our isolation forest approach. The integration of multiple distance metrics offers potential enhancement strategies for our proposed method.

Similarly, Hu et al. develop a density clustering approach specifically for taxi trajectory anomaly detection, focusing on spatial-temporal patterns in urban environments. Their work on combining density and length outlier detection provides comparative benchmarks and alternative preprocessing strategies that could inform our methodology design.

Deep Learning and Advanced Model-Based Methods Huang et al. propose a deep learning framework combining LSTM autoencoders with attention mechanisms for vehicle trajectory reconstruction and anomaly identification. While computationally more intensive than isolation forests, this work demonstrates the importance of temporal sequence modeling in trajectory analysis. The attention mechanism insights could potentially enhance feature selection in our isolation forest implementation.

Li et al. introduce DiffTAD, a denoising diffusion probabilistic model for vehicle trajectory anomaly detection. This cutting-edge approach represents the current state-of-the-art in deep learning for trajectory analysis. Although beyond the scope of our current proposal, this work establishes performance benchmarks

and demonstrates the potential for generative models in trajectory anomaly detection.

Zhao et al. focus on abnormal driving behavior recognition in commercial vehicles using time series analysis and multi-scale CNN. Their work on BeiDou GPS data provides relevant insights for handling large-scale trajectory datasets and could inform our data preprocessing strategies.

Traditional Statistical and Distance-Based Methods Han et al. develop a trajectory big data offline mining approach for taxi anomaly detection, utilizing pathlet-based trajectory representation and edit distance calculations. Their work provides valuable preprocessing techniques and trajectory segmentation strategies that could enhance our isolation forest feature engineering.

Video-Based and Computer Vision Approaches Yin et al. present intelligent video analysis methods for vehicle abnormal behavior detection, focusing on visual surveillance data. While not directly applicable to GPS trajectory data, this work offers complementary perspectives on anomaly definition and multi-modal validation approaches that could strengthen our evaluation framework.

Real-Time Detection Systems Recent work on real-time taxi spatial anomaly detection based on vehicle trajectory prediction emphasizes real-time processing and prediction-based anomaly detection, providing insights into system optimization and deployment considerations relevant to practical implementation of our proposed isolation forest approach.

Spatial and Feature Mixed Approaches Recent work on spatial and feature mixed outlier detection demonstrates hybrid approaches that combine spatial analysis with feature-based detection methods. This approach aligns well with our proposed combination of isolation forests with knowledge graph representations, suggesting potential for enhanced performance through methodological integration.

The reviewed literature establishes isolation forests as a viable and efficient approach for trajectory anomaly detection, while highlighting opportunities for enhancement through improved feature engineering, hybrid methodologies, and knowledge graph integration. Our proposed research addresses gaps in combining tree-based isolation methods with semantic knowledge representations, potentially offering both computational efficiency and interpretability advantages over current deep learning approaches.

1.1 Data and Preprocessing

1.2 Data

1.3 Data Preprocessing

Report with examples on why there is a need for preprocessing. And how these exceptional cases are dealt with.

Report all the statistics before and after the preprocessing.

See Appendix A for the reasoning and choices of all the parameters.

2 Methodology

Below are several ways you can obtain the list of abnormal traj.

2.1 Isolation Tree

Isolation Tree method. And the choice of parameters. - anything that is too detailed goes into the Appendix

2.2 Ratio...

- anything that is too detailed goes into the Appendix

2.3 Improve the results

When examining the real data, we noticed that simply applying the above-mentioned traj. detection algorithm is not enough. Some exceptions should be taken into consideration.
- anything that is too detailed goes into the Appendix
Explain exception 1, 2, 3.

3 Evaluation

Baseline of your naive traj. section algorithm: simple and imperfect.
Then the Isolation tree algorithm
Fine-tuned/improve isolation tree algorithm
I want to see a table of Precision

	Precision	Parametric Setting	Comments
Baseline			
Iso Tree			
Improved Iso Tree			

3.1 Conversion to Knowledge Graph????

...

3.2 Synthetic Knowledge Graph Generation

...

[Fix Chinese chars not displaying]

References

A Data Preprocessing Details