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 Kędzia[1][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

## 1.1 Anomaly Detection Methods in Vehicle Trajectories

你好，世界！ The detection of abnormal vehicle trajectories is crucial for applications ranging from fraud detection to traffic management. Various methods have been developed, each with distinct characteristics, data requirements, and feasibility levels. This section organizes these approaches by method type, discussing their technical foundations, data dependencies, and implementation feasibility.

*Clustering-Based Methods*

- **Characteristics**: Groups trajectories based on spatial/spatiotemporal similarity, assuming normal behavior forms dense clusters
- **Key Algorithms**:
  - *DBSCAN/RDBSCAN* for density-based clustering [7,9]
  - *CFSFDP* for density peak identification [7]
  - *K-Means* for comparative analysis [1]
- **Data Types**: Taxi GPS data (7,600 vehicles/month [10]), requires grid mapping and trajectory indexing
- **Feasibility**: Moderate. Requires sufficient historical data and careful parameter tuning [9]

*Distance/Similarity-Based Methods*

- **Characteristics**: Identifies outliers through geometric or behavioral dissimilarity metrics
- **Key Metrics**:
  - Edit distance for Pathlet sequences [1]
  - Enhanced *DTW* and *Hausdorff* distances [7]
  - *MBR* for geographic distribution analysis [7]
- **Data Types**: High-frequency GPS data (502 SF trajectories [9])
- **Feasibility**: High for basic metrics (e.g., detour detection), Moderate for sequence-based analysis [7]

*Model-Based Methods*

- **Characteristics**: Uses machine learning to learn normal patterns from data
- **Key Models**:
  - *LSTM-AE-Attention* with data augmentation [8]
  - *TSA + MCNN* for time-series analysis [4]
  - *DiffTAD* diffusion models [5]
- **Data Types**: Diverse sources including BeiDou GPS [4] and traffic trajectories
- **Feasibility**: Low-Moderate. Requires substantial training data and computational resources [8]

*Density-Based Anomaly Detection*

- **Characteristics**: Focuses on low-density trajectory regions
- **Key Approaches**:
  - *DENM* density values [7]
  - *iBAT* isolation forests [10]
- **Data Types**: Grid-mapped taxi GPS (Beijing/Shanghai datasets [7])
- **Feasibility**: Moderate. Requires spatial discretization tuning [10]

*Video Analysis-Based Methods*

- **Characteristics**: Analyzes visual patterns from surveillance footage
- **Key Techniques**: Background subtraction, Mean Shift tracking [2]
- **Data Types**: Video datasets (UCF-Crime [6])
- **Feasibility**: Low for GPS-based projects due to domain mismatch

*Data Quality Methods*

- **Characteristics**: Detects sensor reliability issues
- **Key Indicators**: Temporal gaps [1,11], position jumps [3,11]
- **Data Types**: Raw GPS logs with timestamps
- **Feasibility**: High. Requires minimal computation [1]

*Hybrid/Specialized Methods*

- **Characteristics**: Combines multiple approaches for specific anomaly types
- **Key Examples**:
  - Density-length outlier fusion [9]
  - Spatiotemporal *Pathlet* analysis [1]
- **Data Types**: Contextual data (e.g., POI types) with trajectories
- **Feasibility**: Moderate. Depends on data integration complexity [9]

The methodological landscape shows clear trade-offs: distance-based methods offer computational efficiency for basic anomalies, while model-based approaches enable complex pattern recognition at higher resource costs. Recent advances in deep learning [5] and real-time detection [3] continue to expand the feasibility frontier for different application scenarios.

### 1.2 Data and Preprocessing

### 1.3 Data

### 1.4 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

1. Boyang, H., Zhaoyang, W., Beihong, J.: 一种基于轨迹大数据离线挖掘与在线实时监测的出租车异常轨迹检测算法. **46**(3), 247–252 (2016). https://doi.org/10.3969/j.issn.0253-2778.2016.03.010
2. Hongpeng, Y.: . Master's thesis, Beijing Jiaotong University (2014)
3. Hu, W., Li, M., Kwan, M.P., Luo, H., Chen, B.: Real-time taxi spatial anomaly detection based on vehicle trajectory prediction. Travel Behaviour and Society **34**, 100698 (2024). https://doi.org/https://doi.org/10.1016/j.tbs.2023.100698
4. Jian-dong, Z., Xiao-dong, Z., Yu, L., Wei-wei, W.: . **22**(1), 284–291 (2021)
5. Li, C., Feng, G., Li, Y., Liu, R., Miao, Q., Chang, L.: Difftad: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection. Knowledge-Based Systems **286**, 111387 (2024). https://doi.org/https://doi.org/10.1016/j.knosys.2024.111387, https://www.sciencedirect.com/science/article/pii/S0950705124000224
6. Lu, Y., Chen, W., Yu, Z., Wang, J., Yang, C.: Vehicle abnormal behavior detection based on dense block and soft thresholding. CMC **79**(3), 5055–5067 (2024). https://doi.org/10.32604/cmc.2024.050865
7. Ming, H., Yuting, C., Qiang, L., Bo, Z., Gongda, Q.: . (6), 49–54 (2019)
8. Shi-chen, H., Chun-fu, S., Juan, L., Zong-tao, D.: . **21**(3), 47–54 (2021)
9. Yuan, H.: . In: . pp. 54–58. No. 6 (2019)
10. Zhang, D., Li, N., Zheng, Y., Ramamohanarao, K., Zhao, Z.: ibat: Detecting anomalous taxi trajectories from gps traces. In: UbiComp'11: Proceedings of the 13th international conference on Ubiquitous computing. pp. 99–108. ACM (2011). https://doi.org/10.1145/2060091.2060106
11. Zheng, Y.: Trajectory data mining: An overview. ACM Trans. Intell. Syst. Technol. **6**(3) (May 2015). https://doi.org/10.1145/2743025, https://doi.org/10.1145/2743025

## A Data Preprocessing Details