Identify Abnormal driving behavior using Spatio-Temporal analysis

AU2140057 - Sanket Shah, AU2140125 - Khushal Trivedi, AU2140165 - Saumya Shah, AU2140233 - Harshal Prajapati.

***Abstract*— Understanding and identifying abnormal driving behavior is critical for strengthening road safety and traffic control systems. While it's common knowledge that drivers act consistently on some sections of the road, there are lots of circumstances in which this isn't the case. The purpose of this research is to identify unusual driving behavior by applying spatio-temporal analysis techniques. The study uses a trajectory dataset, to find abnormal tendencies. By building a binary classifier, trajectories are categorized as "normal" or "abnormal" based on their spatial and temporal characteristics. Random Forest Algorithm is the method used for classification. By looking at the trajectories, the model finds patterns that may indicate deviations from normal driving behavior. Temporal components like unexpected stops or irregular direction changes, as well as spatial characteristics like speed, acceleration, velocity are important indicators for anomaly identification. This research contributes to transportation efficiency and safety by providing a systematic approach to identify and address potentially hazardous driving habits.**

***Index Terms*— Normal, Abnormal, Random Forest Algorithm, spatial, temporal, dataset.**

I. Introduction

In the domain of Transportation and Road Safety; the rules that are based on driving are crucial and essential and along with that detecting abnormal driving behavior beholds a major role and impact to it. Abnormal driving behavior has been a serious issue, and it has impacted society tremendously and requires it to be restricted. Any driving pattern which is not in accordance with rules and regulation of Indian Motor Vehicle Act is considered as abnormal driving and it needs to be identified and should be restricted. Henceforth, with reference to this problem; via using Machine Learning, we tend to solve this problem with the help of binary classifiers, which can detect abnormal driving patterns based on various features like trajectory points, height, altitude, time stamp, velocity, acceleration, etc. In our study, we aim to focus on binary classification using Random Forest Algorithm. By considering factors like geographical location (longitude and latitude), height, altitude, and other topological features, our model becomes adept at distinguishing between normal and abnormal driving behaviors. This approach, combined with spatio-temporal analysis, allows us to emphasize abnormalities in driving behavior and classify them accurately. By doing so, we aim to contribute to safer roads and better transportation systems.

II. Methodolgy

In order to find the anomaly in driving behavior, we had a large data set consisting of features like time stamp, latitude, longitude, height, width and altitude which was analyzed using the Random Forest Algorithm. The detail methodology is as follows:

1. Extract Relevant Data from Dataset:
We extracted all the trajectory dataset containing information such as geographical location (longitude and latitude), height, altitude, time stamp, frame etc. We also preprocessed the dataset to ensure data quality and consistency using feature engineering.

2. Spatio-Temporal Analysis:
After extracting the data we performed spatio-temporal analysis on the dataset to identify patterns and anomalies in driving behavior and we also calculated velocity using frame no and left top. In that we analyzed temporal components such as unexpected stops, irregular direction changes, velocity and spatial characteristics.

3. Training an Random Forest Algorithm:
After all the analysis we trained the Random Forest Algorithm using the preprocessed dataset, considering factors like geographical location, height, altitude, time frame, velocity etc. We used a binary classification approach to categorize trajectories as "normal" or "abnormal" based on their spatial and temporal characteristics which we got from the analysis and putting some threshold conditions into it.

4. Predicting Abnormal Tracks:
Driving behavior is distinguished between normal and abnormal driving behaviors by evaluating the performance of the classifier.

III. Results

```
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.89      0.86        44
           1       0.44      0.33      0.38        12

    accuracy                           0.77        56
   macro avg       0.64      0.61      0.62        56
weighted avg       0.75      0.77      0.76        56

Predicted labels: ['normal', 'normal', 'normal', 'abnormal', 'normal', 'abnormal', 'normal',
```

Fig. 1. Results image

Precision for class 0 (normal) is 0.83, indicating that out of all the predicted normal instances, 83% were actually normal. Precision for class 1 (abnormal) is 0.44, indicating that only 44% of total abnormal instances were predicted as abnormal.

Recall: Recall is the ratio of correctly predicted positive observations (true positives) to the total actual positives (true positives + false negatives). It measures the ability of the classifier to find all positive instances. Here, Recall for class 0 (normal) is 0.89, suggesting that the classifier correctly identified 89% of all the actual normal instances. For class 1 (abnormal) it is 0.33, indicating that the classifier only captured 33% of all actual abnormal instances.

F1-score: The F1-score is the weighted average of precision and recall, giving more weight to lower values. It balances both precision and recall into a single metric. In this case, F1-score for class 0 (normal) is 0.86 and for class 1 (abnormal) is 0.38.

Support: Support is the number of actual occurrences of each class in the specified dataset. In our case, it shows the number of instances in each class.

Accuracy: It measures the overall correctness of the model's predictions. The accuracy here is 0.77, which implies that the model correctly classified 77% of the instances.

Macro avg: Macro average is the average of precision, recall, and F1-score across all classes. It provides an indication of overall performance across all classes, without considering class imbalance. In this case, the macro average F1-score is 0.62.

Weighted avg: It is the weighted average of precision, recall, and F1-score across all classes, weighted by the support (number of true instances) of each class. It gives more weight to classes with larger support. Here, the weighted average F1-score is 0.76.

IV. Discussion

A. Performance Evaluation:

- The precision, recall, and F1-score metrics are essential for measuring the classifier's effectiveness in correctly identifying instances of both classes.

B. Precision and Recall Disparities:

- The disparity between precision and recall for the "abnormal" class suggests that the classifier is more prone to false positives, i.e., instances which are predicted as "abnormal" are actually "normal". This indicates potential areas for improvement, especially in reducing false alarms for abnormal behavior detection.

C. Impact of Class Imbalance:

- The lower recall for the "abnormal" class could be influenced by class imbalance, where the number of abnormal instances is significantly lower than normal instances. Addressing class imbalance through techniques like oversampling, under-sampling, or using different evaluation metrics can help mitigate this issue.

D. Model Generalization and Robustness:

- While the overall accuracy of 77% indicates reasonable performance, further validation on unseen data or through cross-validation is necessary to assess the model's generalization ability and robustness across different datasets.
- It would be valuable to investigate if the model's performance varies across different settings or if it remains consistent, indicating its reliability in real-world scenarios.

E. Interpret ability and Feature Importance:

- Exploring feature importance derived from the Random Forest model can provide insights into which spatial features contribute most to classification decisions. Understanding these features can lead to better interpretation of abnormal behavior patterns.

F. Comparison:

- Comparing the performance of the Random Forest Algorithm with the output available then, we will try to improve our model.

G. Future Directions:

- We would convert this output in the form of graph and will visualize the output such that we can predict the movement of the vehicle and can see whether the prediction made is right or wrong.

V. Conclusion

This research presents a robust approach to identifying abnormal driving behavior by integrating spatio-temporal analysis and Random Forest Algorithm. Leveraging trajectory data and various spatial and temporal features, our model accurately distinguishes between normal and abnormal driving patterns, contributing to road safety and traffic control systems.

Our findings underscore the importance of addressing abnormal driving behavior to enhance road safety and transportation efficiency. By detecting deviations from normal driving, our approach enables proactive intervention strategies to mitigate potential hazards and reduce accident risks.

VI. References

1) Ma, Y., Xie, Z., Chen, S., Qiao, F., & Li, Z. (2023). Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals. *Transportation Research Part C: Emerging Technologies*, 146, 103983.

2) Huang, X., Yun, P., Wu, S., & Hu, Z. (2023). Abnormal driving behavior detection based on an improved ant colony algorithm. *Applied Artificial Intelligence*, 37(1).

3) Zhang, Y., He, Y., & Zhang, L. (2023). Recognition method of abnormal driving behavior using the bidirectional gated recurrent unit and convolutional neural network. *Physica A: Statistical Mechanics and Its Applications*, 609, 128317.

4) Abosaq, H. A., Ramzan, M., Althobiani, F., Abid, A., Aamir, K. M., Abdushkour, H., Irfan, M., Gommosani, M. E., Ghonaim, S. M., Shamji, V. R., & Rahman, S. (2022). Unusual driver behavior detection in videos using deep learning models. *Sensors*, 23(1), 311.