DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Robotics, Cognition, Intelligence

Anomaly Detection for the behavior of drivers based on Structural Temporal Graph Neural Networks

Hanxi Jiang

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Robotics, Cognition, Intelligence

Anomaly Detection for the behavior of drivers based on Structural Temporal Graph Neural Networks

Erkennung von Anomalien im Verhalten von Autofahrern auf der Grundlage struktureller temporaler neuronaler Netze

Author:	Hanxi Jiang
Supervisor:	Supervisor
Advisor:	Advisor
Submission Date:	Submission date

I confirm that this master's thesis in informatics: robotics, cognition, intelligence is my own work and I have documented all sources and material used.

Munich, Submission date Hanxi Jiang

Acknowledgments

Abstract

Kurzfassung

Contents

1. Introduction

Despite the fact that High Driving Automation(SAE Level 4) is achievable in the foreseeable future[1], the majority of drivers nowadays would still prefer comprehensive control over their own vehicles. Therefore, increasingly more studies have been focusing on driving safety[2][3]. According to these papers, driver behavior represents the majority cause of accidents while driving. In that case, several methods have been developed to avoid potential danger. Such methods involve either improving monitoring of the vehicle's inside situation, which analyzes the driving parameters as well as the driver him or herself to determine whether there's no abnormality[4], or proposing a vehicle detection and tracking system from an outer view that estimates time-to-collision (TTC) and warn the driver for a possible collision[5].

On the other hand, only few research focus on driving behavior prediction, which may be due to the lack of application for the undetermined decision model in behavior prediction, specifically under the topic of autonomous driving. Scientists have succeeded in years of generating dynamic graphs based on videos. However, there is still a gap in the rational use of these forms for behavioral prediction.

In this work, we would like to focus on constructing a comprehensive behavior prediction model based on graph neural networks (GNNs). A Graph Neural Network is a novel type of neural network architecture that can be applied to graph-like inputs. As we are now expecting to train the behavior model for the participators, the training data would contain interaction between several objects and participators while they are driving. GNN is, therefore, prioritized due to its unique structure. Along with that, we would specifically focus on building dynamic graphs as training datasets, as our model would be trained based on sequences of behaviour descriptions extracted from videos. Besides, to ensure the compatibility of the dataset and training model, we made some adaptions based on JODIE[6], which is the model based on the dynamic evolution of users and items. In order to make the predictions as detailed as possible, we expanded the output catalogue to ensure that not only the interaction itself but also the type of it would be described.

In a nutshell, We would first gather all the necessary information with the help of the Large language model (LLM), convert it into dynamic graphs and insert these graphs into the model we have adapted from model JODIE. These graph could contain either one specificl participator or all the concerned people. By learning how dynamic graphs change under time series, like the appearance and vanish of all these edges and nodes in the grap, the model should be able to absorb the feature behind it and come up with the prediction for behaviour in the future. Therefore, anormaly detection could also be achieved by comparing the predicted graph with the real one and alert once when any unsuitable behaviour during the driving is detected. we believe that this model provides a new perspective for the prediction of driving behavior detected from videos and allows for more diversified and targeted forecasting.

1.1. contribution

The main contributions of this thesis are summarized as:

Dataset Collecton From the Dataset *drive & act*, we acquire the hierarchical activity labels of given video data and rewrite them in the form of time sequences. We also cluster the behavior types into several categories with the help of the Large Language Model.

Dynamic Graph Construction By reassembling the nodes and edges in the video data, we construct time sequenced dynamic graphs that could be used as training data for the model.

Learning Model Adaption To enrich the diversity of the prediction, we expand the output from binary to catalogue description to emsure that not only the interaction itself but also the type of it would be described.

anormaly detection By comparing the predicted model with the real one, we could detect any unsuitable behavior during the driving and alert the driver.

1.2. Structure

This thesis is structured as follows. In Chapter2 we would introduce and explain concepts and definitions concerned to this document. Chapter3 reveiews related work in the field of anormaly detection and dynamic link prediction models, the dataset this work refers to and the model we have adapted. Chapter4 describes the methodology of this work, including the dataset collection, dynamic graph construction and model adaption. Chapter**??** presents the evaluation of the model and the results of the prediction. Chapter**??** discusses the potential future work that could be done based on this work. Chapter**??** concludes the thesis and gives a summary of the work done.

2. Background

2.1. Large Language Model

Large Language Models (LLMs) are highly complex artificial intelligence systems that have the capability to learn from the vast amounts of available text data [104]. Discriminative LLMs (e.g., BERT) have been well-studied for answering questions in a classification manner.

2.2. graph

2.2.1. Dynamic graph

2.2.2. Graph Neural Networks

2.3. behavior prediction

3. Related Work

4. Methodology

This work aims to construct a graph neural network-based architecture for predicting, analyzing, and detecting any potentially abnormal behavior regarding the driver during the whole driving process. In particular, The model extracts a description graph, the so-called scene graph, of the driver from the video filmed inside the vehicle and trains itself with these data to learn for future behavior prediction. The result will be used to compare and detect any abnormal behavior. Here we would lay most emphasis on the construction of the training model. To make precious anomaly detection we aim to predict not only if there is a behavior between humans and a specific kind of object but the type of behavior as well, which will cause several adaptions based on existing model *JODIE*.

4.1. scene graph generation

4.1.1. video data extracting

4.1.2. graph generating

4.2. model architecture

After comparing all the training results of the below models we would find that **JODIE** is one coming up with the best prediction. However, the model jodie still fail to predict the state of the predicted edge. In my masterwork I would like to rewrite the embedding function and the loss function of **JODIE to make the state prediction possible.

- function from **JODIE**:
embedding function

$$\mathbf{u}(\mathbf{t}) = \sigma(W_1^u \mathbf{u}(\mathbf{t}^-) + W_2^u \mathbf{i}(\mathbf{t}^-) + W_3^u f + W_4^u \Delta_u)$$

$$\mathbf{i}(\mathbf{t}) = \sigma(W_1^i \mathbf{i}(\mathbf{t}^-) + W_2^i \mathbf{u}(\mathbf{t}^-) + W_3^i f + W_4^i \Delta_i)$$

loss function(BCE)

$$L = -(j_{pos} \log \tilde{j} + j_{neg} log(1 - \tilde{j}))$$

where

$$\tilde{j}(t + \Delta) = W_1 \hat{u}(t + \delta) + W_2 \bar{u} + W_3 i(t + \Delta^-) + W_4 \bar{i} + B$$

- functions adapted in my work:

embedding function

$$\mathbf{u}(\mathbf{t}) = \sigma(W_1^u \mathbf{u}(\mathbf{t}^-) + W_2^u \mathbf{i}(\mathbf{t}^-) + W_3^u f + W_4^u s + W_5^u \Delta_u)$$

$$\mathbf{i}(\mathbf{t}) = \sigma(W_1^i \mathbf{i}(\mathbf{t}^-) + W_2^i \mathbf{u}(\mathbf{t}^-) + W_3^i f + W_4^i s + W_5^u \Delta_i)$$

we will change it from BCE to CE for predictiing state.

$$\tilde{j}(t+\Delta) = W_1 \hat{u}(t+\delta) + W_2 \bar{u} + W_3 i(t+\Delta^-) + W_4 \bar{i} + W_5 s + B$$

A. General Addenda

If there are several additions you want to add, but they do not fit into the thesis itself, they belong here.

A.1. Detailed Addition

Even sections are possible, but usually only used for several elements in, e.g. tables, images, etc.

B. Figures

B.1. Example 1

✓

B.2. Example 2

✗

List of Figures

List of Tables

Bibliography

[1] T. Inagaki and T. B. Sheridan. “A critique of the SAE conditional driving automation definition, and analyses of options for improvement”. In: *Cognition, technology & work* 21 (2019), pp. 569–578.

[2] J. D. Lee. “Driving safety”. In: *Reviews of human factors and ergonomics* 1.1 (2005), pp. 172–218.

[3] J. D. Lee. “Fifty years of driving safety research”. In: *Human factors* 50.3 (2008), pp. 521–528.

[4] M. Karrouchi, I. Nasri, M. Rhiat, I. Atmane, K. Hirech, A. Messaoudi, M. Melhaoui, and K. Kassmi. “Driving behavior assessment: a practical study and technique for detecting a driver’s condition and driving style”. In: *Transportation Engineering* 14 (2023), p. 100217.

[5] B. Aytekin and E. Altu. “Increasing driving safety with a multiple vehicle detection and tracking system using ongoing vehicle shadow information”. In: *2010 IEEE International Conference on Systems, Man and Cybernetics*. IEEE. 2010, pp. 3650–3656.

[6] S. Kumar, X. Zhang, and J. Leskovec. “Predicting dynamic embedding trajectory in temporal interaction networks”. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 1269–1278.