Transformer-Based Prediction of Freeway Vehicle Lane Change Intentions

1st Jiayi Song
College of Literature, Science, and the Arts
University of Michigan
Ann Arbor, United States
jiayiso@umich.edu

***Abstract*—Accurate lane change intention prediction can help vehicles reduce the probability of collision, which is crucial for maintaining traffic safety. In this paper, a data-driven vehicle lane change intention recognition model is proposed. The target vehicle is based on Transformer to build a lane change intention recognition model considering the driving data of the target and surrounding vehicles, the size of the surrounding vehicles, and the type of the target vehicle. The NGSIM dataset is used for training and validation. The results show that the proposed model recognizes the lane change intention of a vehicle with 85.33% accuracy. This study can be used to predict the lane change intentions of surrounding vehicles, provide assistance to highway drivers, and help intelligent vehicle driver decision-making systems to ensure driving safety.**

***Index Terms*—lane change intention recognition,transformer, time series**

I. Introduction

Vehicle lane changing is one of the major causes of traffic congestion and traffic accidents. The core task of this project is to predict whether a vehicle will change lanes or not by analyzing vehicle trajectory data and using time series models. This problem has important applications in traffic safety and efficiency optimization. Through this project, we practice our knowledge of data analysis and modeling, explore how to apply advanced deep learning models to real-world problems, and provide new ideas for data analysis in the transportation field.

In recent years, the development of data collection technologies and the release of open datasets have driven the research on data-driven vehicle behavior prediction models. From traditional statistical methods to deep learning models that have emerged in recent years, the application of different techniques in vehicle behavior modeling has gradually expanded the possibilities in this field. Yang[1]based on the naturalistic highway driving data proposed a novel integrated bi-directional recurrent neural network (RNN) model with long short-term memory(LSTM) units to process time-series driving sequences and temporal behavior patterns. Liu[2] proposed an end-to-end approach in the lane detection problem, using a network built by Transformer to learn richer structures and contexts. Lin[3] used an LSTM model with a spatio-temporal attention mechanism to explain the vehicle trajectory prediction method. Zan[4] accurately identifies vehicle lane-changing behavior using a LSTM network considering highway mobile bottleneck environments.Yu[5] combines the advantages of Convolutional Neural Network (CNN) and LSTM to accurately predict the trajectories of self-driving vehicles.

Identify applicable funding agency here. If none, delete this.

II. Methods

A. Learning problem formulation

The input to the model used in the project is time series data with a time step of 0.1 seconds, i.e., 10 samples per second. The data content includes vehicle trajectory data for a random 4-second consecutive time period within 8 seconds before the lane change, including a total of 28 features of vehicle type, speed, acceleration, and headway distance data from the target vehicle and the 6 vehicles around it. The model outputs are lane change category labels including 0: no lane change, 1: lane change to the left, and 2: lane change to the right. The goal of the model is to be able to predict a vehicle's intention to change lanes based on the input time series features and to improve the prediction accuracy.

B. Dataset description

The NGSIM(Next Generation Simulation) dataset is a collection of publicly available traffic data initiated and funded by the U.S. Department of Transportation's Federal Highway Administration at the beginning of the 21st century. These data are obtained from multiple high-precision cameras mounted at elevated locations and provide detailed microscopic vehicle trajectory information. This project will mainly use the US-101 freeway dataset, which is a detailed high-density vehicle trajectory data collected on the US-101 freeway in the Los Angeles area of California, USA, and contains detailed trajectory information such as time, vehicle number, location, speed, acceleration, and so on. The roadway section is shown in Fig1.

The NGSIM dataset includes the following fields: Vehicle ID, Frame ID, Total Frames, Global Time, Local X, Local Y, Global X, Global Y, Vehicle Length, Vehicle Width, Vehicle Class, Vehicle Velocity, Vehicle Acceleration, Lane Identification, Preceding Vehicle, Following Vehicle, Spacing, Headway. The data cover different time periods and traffic flow conditions, and the total number of records reaches hundreds of thousands, which is highly representative.

In order to expand the sample size and maintain a balance of different sample sizes, the project uses a sliding time window

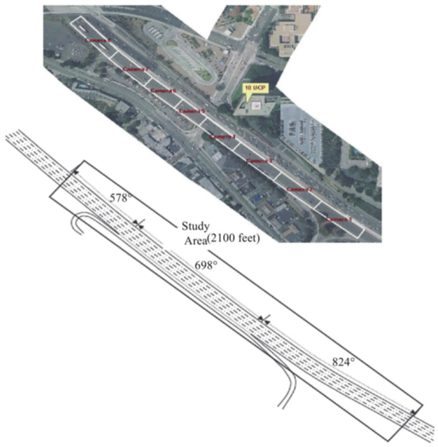

Fig. 1. Detailed US-101 freeway Study Area.

approach. This method uses the point in time when the lane number changes as the lane change point and extracts data by sliding forward from that point. The change in lateral position occurs 4 seconds before the lane change point, while the change in speed occurs before the lateral displacement. Therefore, the length of the time series of the input samples was set to 4 seconds, and the width of the sliding time window was also set to 4 seconds.

About 4500 samples were extracted using the sliding time window method and 1500 samples were extracted for each intention. 80% of the samples were randomly selected as training data and the rest were test data. For vehicles without surrounding vehicles and vehicles in the edge lanes, the vehicle model is uniformly assigned a value of 10, the vehicle characteristics are consistent with the vehicle, and the headway is set to 9999; vehicles in the edge lanes are always assumed to be traveling in close synchronization with the vehicle in the adjacent edge lanes.

III. Model formulation

Since the process of a vehicle deciding to change lanes, initiating the lane change, and completing it takes a certain amount of time, it is necessary to apply the Transformer architecture for time series classification, modeling, and predicting lane-changing behavior. The core of Time Series Classification with a Transformer Model is leveraging the Multi-Head Self-Attention Mechanism to capture global dependencies in the input data without relying on the fixed positional order of the sequence.

The model proposed in this paper is based on Transformer architecture for sequence classification. The model mainly consists of multilayer Transformer encoder block and classification header. The mathematical formulation and detailed description of the model is given below.

A. Transformer encoder block

Each Transformer encoder block consists of two parts: a multi-head self-attention mechanism and a feed-forward neural network (FFN), each followed by the use of residual connectivity and layer normalization.

1) Multi-Head Attention Mechanism: Given an input sequence $X \in \mathbb{R}^{T \times d}$, where T is the sequence length and d is the feature dimension:

1) Compute Query, Key, and Value matrices:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V, \tag{1}$$

where W_Q, W_K, W_V are learnable weight matrices.

2) Multi-head attention extension:

$$\text{head} = \mathit{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V. \tag{2}$$

$$\text{MHA}(X) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O, \tag{3}$$

3) Add residual connections and layer normalization:

$$Z_1 = \text{LayerNorm}(X + \text{Dropout}(\text{MHA}(X))). \tag{4}$$

2) Feedforward Neural Network (FFN): The FFN contains two linear transformations with a ReLU activation function in between:

$$Z_2 = \text{ReLU}(Z_1 W_1 + b_1), \tag{5}$$
$$Z_3 = Z_2 W_2 + b_2, \tag{6}$$

Add residual connections and layer normalization:

$$Z_{\text{out}} = \text{LayerNorm}(Z_1 + \text{Dropout}(Z_3)). \tag{7}$$

B. Classification Head

After passing through N layers of Transformer encoder blocks, the output sequence $Z_{\text{out}} \in \mathbb{R}^{T \times d}$ is processed through a global average pooling layer to aggregate information:

$$\mathbf{h} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_{\text{out},t}, \tag{8}$$

where $\mathbf{z}_{\text{out},t}$ represents the feature at time step t.

The aggregated feature $\mathbf{h}$ is then fed into a Multi-Layer Perceptron (MLP) and normalized using Dropout:

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{h}W_{\text{mlp},1} + \mathbf{b}_{\text{mlp},1}), \tag{9}$$
$$\mathbf{h}_2 = \text{Dropout}(\mathbf{h}_1), \tag{10}$$

Softmax activation function to output class probabilities:

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{h}_2 W_{\text{mlp},2} + \mathbf{b}_{\text{mlp},2}), \tag{11}$$

C. Loss Function

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{n_{\text{classes}}} y_{i,c} \log \hat{y}_{i,c}, \tag{12}$$

IV. Result

In this project, a vehicle trajectory dataset is processed for lane-changing intention recognition and prediction, and the dynamic state during the 8-second period before the lane-changing is recorded. The samples are labeled into 3 categories: no lane change (Class 0), lane change to the left (Class 1), and lane change to the right (Class 2). The features are processed in a sliding window form and a time series model is constructed based on the Transformer architecture and good prediction results are obtained.

During model training, from the Fig 2, it can be observed that the training loss gradually decreases with the increase of training rounds and eventually stabilizes, indicating that the model is able to fit the data effectively on the training set. The validation loss decreases rapidly at the beginning and then shows small fluctuations, verifying the generalization ability of the model. There is no significant gap between the validation loss and the training loss, indicating that no overfitting occurs.

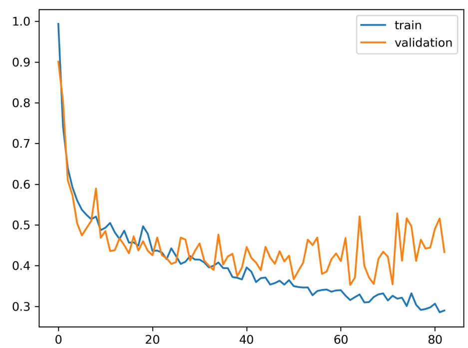

Fig. 2. Training and Verification Loss Curves.

On the test set, the model was evaluated to obtain a test loss of 0.402 and a test accuracy of 85.33%. The test results show that the model performs reasonably well in the actual lane change intention classification task, and is able to effectively distinguish between the three types of lane change behaviors.

V. Analysis and improvement

To gain deeper insights into the model's performance, we computed the confusion matrix on the test dataset, as shown in Fig 3. The confusion matrix reveals that the model achieves a higher classification accuracy for the "No Lane Change" class (Class 0) with relatively low error rates. However, the classification error rates for the "Lane Change to the Left" (Class 1) and "Lane Change to the Right" (Class 2) are slightly higher. These errors are primarily due to misclassifications between these two classes.

To address the classification errors, future work could focus on augmenting the dataset, particularly for Class 1 and Class 2 in regions with overlapping feature spaces, and introducing dynamic features such as relative velocity and surrounding

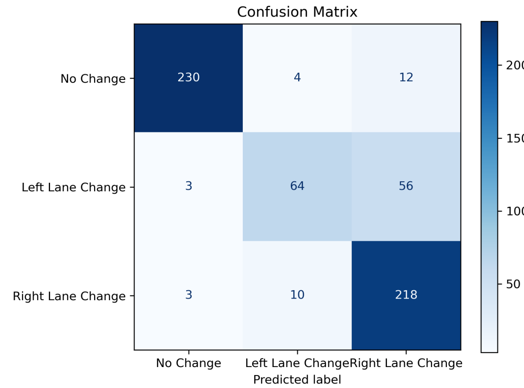

Fig. 3. Confusion matrix.

vehicle density to better distinguish lane-change directions. Additionally, improving the model's representation capability by increasing the number of Transformer blocks and fine-tuning training parameters, such as learning rate and training epochs, could further enhance classification performance.

VI. Conclusion

This paper presents a Transformer-based time series model for identifying vehicle lane-change intentions. By processing spatiotemporal feature data of vehicles and their surrounding environment with a sliding window approach, and leveraging multi-head attention mechanisms combined with feedforward neural networks, the model effectively captures dynamic patterns associated with lane-change behaviors. The model achieved satisfactory performance during training and testing, with a test accuracy of 85.33%.

This program demonstrates the effectiveness of the Transformer architecture in vehicle lane-change intention recognition tasks, providing a novel and powerful approach for vehicle behavior prediction. Future improvements have the potential to further enhance the model's accuracy and generalization, offering reliable technical support for the development of autonomous driving and intelligent transportation systems.

References

[1] Xing Y,Lv C,Wang H, et al. An ensemble deep learning approach for driver lane change intention inference[J]. Transportation Research Part C,2020,115.

[2] Liu Ruijin, Yuan Zejian, Liu Tie, Xiong Zhiliang. End-to-End Lane Shape Prediction With Transformers[J]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021: 3694-3702.

[3] Lin L, Li W, Bi H, Qin L. Vehicle Trajectory Prediction Using LSTMs With Spatial–Temporal Attention Mechanisms[J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14(2): 197-208.

[4] Zan Y, Wang X, Xu B, Feng F, Ling Z, Song J. Lane Change Intention Recognition for Human Driving Vehicles under Moving Bottleneck on Freeway[C]. 2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE). Beijing, China: IEEE, 2022: 586-591.

[5] Yu H, Huo S, Zhu M, Gong Y, Xiang Y. Machine Learning-Based Vehicle Intention Trajectory Recognition and Prediction for Autonomous Driving[C]. 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE). Shanghai, China: IEEE, 2024: 771-775.