A Deep Neural Network for Driving Style Recognition

Tianlun Li

Minxing Liu

Haoyu Gong

***Abstract*- In this project, GPS data is used to help us study different drivers' driving styles. We study an improved deep neural network and a route data encoding model to extract people's driving styles by analyzing GPS data. In this process, we apply both supervised and unsupervised feature learning within one framework. The work on recognizing drivers and analyzing huge driver number has proved that the improved model can give a better outcome than other methods proposed before: lower estimation error and higher identification accuracy.**

I. INTRODUCTION

Similar to biometric, driving style can be used to recognize different people. Many features can help to describe driving patterns, and GPS data is almost the easiest data set to collect. [1] Therefore, in this project, we decide to use movement features like speed, difference of speed, acceleration, difference of acceleration and angular speed to study different driving styles, and all of these can be extracted from GPS sensor. The result could be used for insurance company to develop personalized insurance plan and solve their business problems. According to personal driving style, company would have unique weight for everyone. If this car was used by many people, the insurance fee will be above average. Therefore, this model is beneficial to company and customer.

What others have done is to follow the supervised learning framework whose input is GPS trip data and labels are driver classes. The aim of every iteration of training is to make the classification loss as low as possible. These features learned in these approaches can predict unknown driving actions well for classified drivers. But for unknown drivers (drivers not seen or classified yet), these methods cannot give satisfying results. Besides, it is kind of difficult to prepare large training data sets for each driver and collect sufficiently large number of drivers.

As a solution to these problems, we improved the existing deep neural network into a new model to carry out driving style feature learning. The architecture of this new model is shown in Figure 1. Compared to other deep neural networks, this model makes supervised feature learning and unsupervised feature learning combined together into one framework, which is realized by output of a Recurrent Neural Network (RNN) [2]. The main target is to regularize different features in a classification network using a particular encoder architecture. In Figure 1, fc1 is a layer used to extract driving pattern features. RNN output is a hidden layer in this architecture which is always changing during the runtime, and this improved model is dedicated to rebuilding it. We can take this structure as a type of regularization on the classification RNN feature, and this feature must be typical and transformable. Layer fc1 in this architecture helps describe and predict driving styles better when it comes to unknown drivers.

The main problem is driver identification. We split the data set, 80% for data training and 20% for testing. Finally, we get an identification accuracy of about 45.95%.

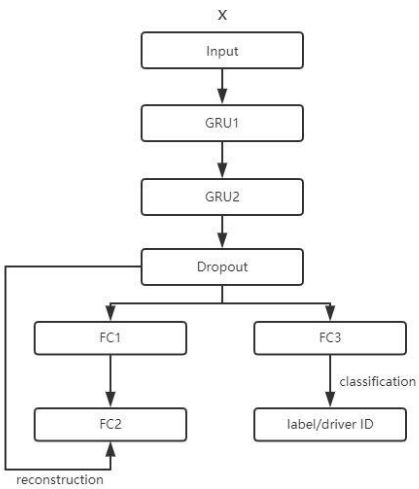
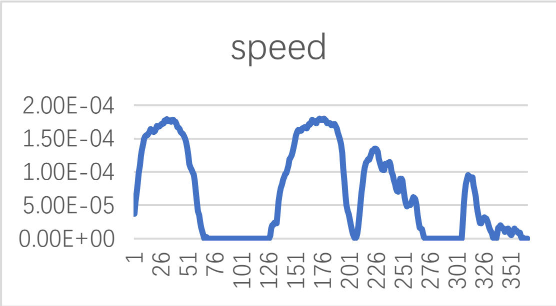


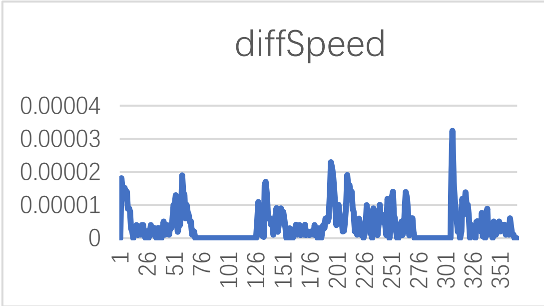
Figure 1: Architecture of the improved deep neural network

II. TRAINING MODEL ILLUSTRATION

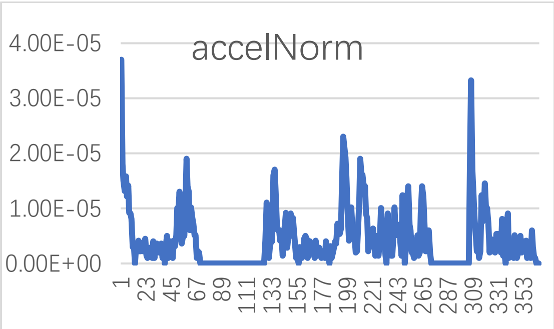
We use a sequence of tuples (u, v, t) to define a GPS trip. (u, v) represents the location and t is time. Now here is the process of data transformation. First we split a trip into fixed-length segments where the length is L_s, these segments have a shift $L_s/2$ and each of them encodes speed, difference of speed, acceleration, difference of acceleration and angular speed. In Figure 2, from (a) to (e) show the distributions of these variables in our data set.



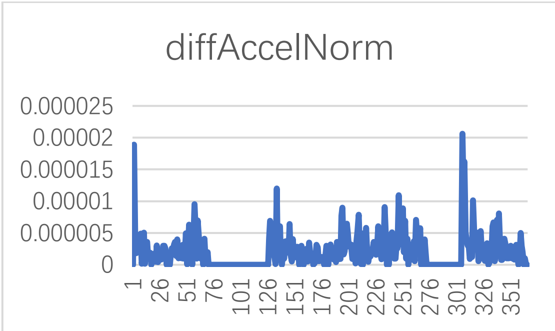
(a) Distribution of speed



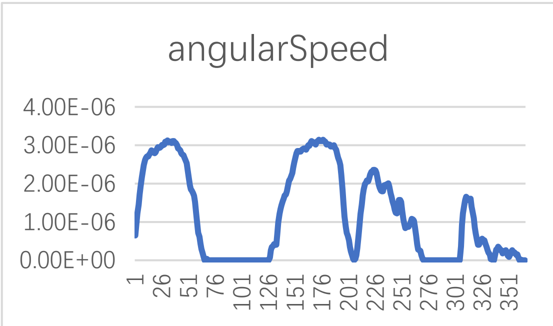
(b) Distribution of difference in speed



(c) Distribution of acceleration



(d) Distribution of difference in acceleration



(e) Distribution of angular speed

Figure 2: Distributions of 5 variables in our data set

After that, each segment is further applied a sliding window of length L_f ($L_f < L_s$) with a shift L_f /2. This is described in Figure 3. Each window generates 7 measurements of the abovementioned 5 features: mean, minimum, maximum, 25%, 50% and 75% quartiles, and standard deviation. Finally, a set of statistical feature matrices of $5 \times 7 = 35$ rows and $2 \times L_s/L_f$ columns are available for each trip. Every input sample of this neural network is a matrix of features recording a trip segment.

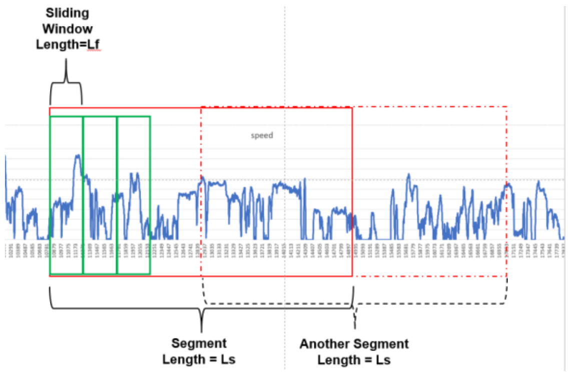


Figure 3: Data transformation sketch

In the following part, the model will be explained. Assuming x is a trip segment which is a 35×128 input, driving style can be defined by the combination of different features (driving movements). The gru1, gru2 and dropout layers in Figure 1 constitute a stacked RNN which takes x to obtain other features. The two GRU (Gated Recurrent Unit) layers can analyze the relationship among a sequence of movements, the dropout layer acts as the bridge between supervised and unsupervised learning, it helps lower overfitting [3]. Let x' represent output of input x.

In Figure 1, the fc3 layer (fully-connected layer) is used to classify samples. A distribution over class labels is generated by a softmax regression. The number of drivers is exactly the number of classes, which is c. Let i $\in \{1, \ldots, n\}, y_i \in \{1, \ldots, c\}$, for a training set $\{x_i,\ y_i\}$, there is a reconstruction loss:

$$\Im_r = \sum_i^n \left| |D_{s_i} - x'_i| |_2^2 + \lambda \right| |s_i||_1 \quad (1)$$

In this equation, $D \in R^{(n\times k)}, s_i \in R^k$. There is an equation which defines classification loss:

$$\Im_c = -\frac{1}{n}\sum_i^n \sum_j^c 1\{y_i = j\} \log \frac{e^{\theta_j^T x'_i}}{\sum_l^c e^{\theta_j^T x'_i}} \quad (2)$$

In this equation, θ is a parameter in softmax regression. In all, we use both Eq(1) and Eq(2) to represent the objective function:

$$min \Im_r + \Im_c \quad (3)$$

The goal of Eq(3) is to obtain a recognition with the clustering attribute and it should be compact to give us better generalization performance. Another equation we want to introduce is K-means clustering algorithm [4]:

$$\min_{\mu} \sum_k (||x_i - \mu_k||^2) \quad (4)$$

In this equation, μ_k is cluster center. The K-means algorithm can be applied to reconstruct x_i [4]:

$$\min_{D,s} \sum_i (||D_{s_i} - x_i||^2)\, , s.t. ||s_i||_0 \leq 1, \forall i \quad (5)$$

Eq(1) is aimed to learn a sparse reconstruction and encode x'_i more efficiently.

Although we improved existing deep neural network into this new model, since this model only extracts driving patterns from segment-level trip data, and this highly depends on local factors like road conditions and chosen route. A more reliable architecture should be proposed to deal with this problem. This theoretic model uses the sum of all segment-level attribute vectors to construct a trip-level driving pattern, which is shown in Figure 4 below. This driving pattern feature equation is:

$$S^{tr} = \frac{\sigma^{tr}}{\max_j\{\sigma_j^{tr}\}} \quad (6)$$

In this equation, we assume a trip tr is split into q segments, $\{ s_i^{tr} \}$ represents segment attributes, $i \in \{1, \ldots, q\}$. The sum of vectors is $\sigma^{tr} = \sum_i^q s_i^{tr}$.

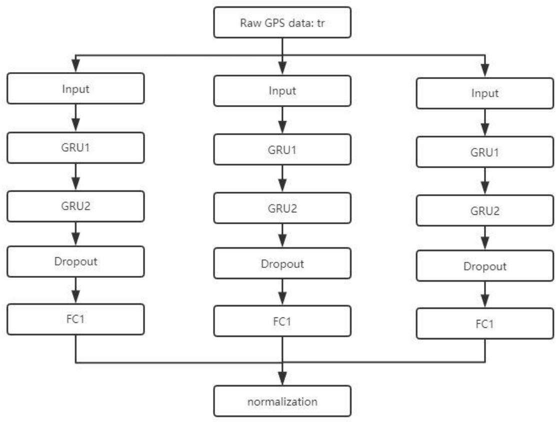


Figure 4: Theoretic model

III. RESULTS

According to our new deep neural network model, we programmed to train our data set. We find this data set from a lab of Korea University (Ref: http://ocslab.hksecurity.net/home), it contains 94381 rows of GPS data from 10 different drivers. For each driver, we randomly divide their trip data into two parts, one is 80% of it and the other is 20% of it. We use the bigger part for data training and the smaller part for data testing. By now, we have reached an accuracy of 45.95%, which is not as good as expectation. Some possible causes: (1) The data set is not big enough; (2) Selected features may not vary much among different people.

IV. SUMMARY

A. Challenges and how we addressed them

1. GPS Data Transformation

The raw GPS data is diverse and tedious. Even though we could read data from geo-location and time directly, it is not enough as the input of neural network. For handling this problem, we did the effective data transformation by using Dong's paper [5], and defined matrices of size 35*128 as network inputs.

2. The selection between LSTM and GRU

The stacked RNN structure is the basement of the whole neural network. It reads data to extract higher-level features. Therefore, the selection between LSTM and GRU is very important. Even though LSTM is a pretty mature and popular RNN architecture, after reading some related papers, we decided to use the lighter and more effective structure GRU. In our project's case, GRU could help us save more time.

3. Our model could be impacted by unavoidable factors

If we want to improve the final accuracy, we must take more and more factors into consideration. Obviously, the present model will be unable to handle those situations. Therefore, we propose a improvement framework for the upcoming challenges.

4. The dataset issue

The real-world problem like our project has already been raised by the auto insurance industry. We looked through the Internet dataset resources as many as possible, and it is all the dataset we've got. The current accuracy might be not so high and satisfactory. But it would get better when having more business-used datasets from industry.

B. Significance of our work

1. Our model combines supervised and unsupervised features

The important part for achieving this function is the dropout layer. It was inspired by Hinton's paper. Beyond that, we could also use it to reduce overfitting. Therefore, it give us the foundation to deal with any other similar problem that need supervised and unsupervised learning.

2. The theoretical model uses the present one as the base encoder

This part was inspired by Fei-Fei and Perona' s paper[6].Our theoretical model will use the Bag-of-Words feature construction strategy. And it gives the theoretical model an ability to "see the whole picture". Based on the present one, the theoretical model will treat a real-world trip by using each segment as a unit. Furthermore, we confirmed the new feature representation function and made it completed.

C. Future work

If continuing this topic, we will focus on our theoretical training model. To do that, a more powerful dataset is inevitable. And with more datasets, we could also improve the accuracy of our present model. We believe that as more datasets and more factors, we need to update our training model over and over again. During that stage, we would change the GRU into LSTM, because of the upcoming huge dataset.

REFERNCES

[1] Lin, N., Zong, C., Tomizuka, M., Song, P., Zhang, Z., & Li, G. (2014). An overview on study of identification of driver behavior characteristics for automotive control. *Mathematical Problems in Engineering, 2014.*

[2] Elman, J. L. (1990). Finding structure in time. *Cognitive science, 14(2), 179-211.*

[3] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580.*

[4] Coates, A., & Ng, A. Y. (2012). Learning feature representations with k-means. *In Neural networks: Tricks of the trade (pp. 561-580).* Springer, Berlin, Heidelberg.

[5] Dong, W., Li, J., Yao, R., Li, C., Yuan, T., & Wang, L. (2016). Characterizing driving styles with deep learning. *arXiv preprint arXiv:1607.03611.*

[6] Fei-Fei, L., & Perona, P. (2005, June). A bayesian hierarchical model for learning natural scene categories. *In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 524-531). IEEE.