

# Driving Style Recognition Incorporating Risk Surrogate by Support Vector Machine



Qingwen Xue, Jianjohn Lu, and Kun Gao

**Abstract** Accurate driving style recognition is a crucial component for advanced driver assistance systems and vehicle control systems to reduce potential rear-end collision risk. This study aims to develop a driving style recognition method incorporating matching learning algorithms and vehicle trajectory data. A risk surrogate, Modified Margin to Collision (MMTC), is proposed to evaluate the collision risk level of each driver's trajectory. Particularly, the traffic level is considered when labelling the driving style, while it has a great impact on driving preference. Afterwards, each driver's driving style is labelled based on their collision risk level using the K-means algorithm. Driving behavior features, including acceleration, relative speed, and relative distance, are extracted from vehicle trajectory and processed by time-sequence analysis. Finally, Supporting Vector Machine (SVM) is applied to recognize driving style based on the extracted features and labelled data. The performance of Random Forest (RF), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP) are also compared with SVM. The "leave-one-out" method is used to validate the performance and effectiveness of the proposed model. The results show that SVM over performs others with 91.7% accuracy. This recognition model could effectively recognize the aggressive driving style, which can better support ADAS.

**Keywords** Driving style recognition · Vehicle trajectory · Risk surrogate · Key feature extraction · SVM

## 1 Introduction

Driving style refers to how drivers choose to drive habitually and the driver states that represent the common parts of varied driving behavior [1]. Recognition of a driver's

---

Q. Xue · J. Lu

The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China

Q. Xue · K. Gao (✉)

Department of Architecture and Civil Engineering, Chalmers University of Technology, Goteborg, Sweden

e-mail: [gkun@chalmers.se](mailto:gkun@chalmers.se)

driving style based on rear-end collision risk is of great significance to improve driving safety. It is important to guarantee the safety and adequate performance of drivers and essential to meet drivers' needs, adjust to the drivers' preference, and ultimately improve the driving environment's safety. Driving style recognition also has potential value to help agencies effectively design control strategies [2, 3].

The paper proposed a driving style recognition model to consider the impact of traffic flow levels on driving behavior. The traffic flow level could be classified into normal and congested traffic, and the driving style is labelled concerning the road condition. The difference of risk surrogates between normal and congested traffic would be explored. The decision about driving style ignoring the traffic level is not acceptable. Therefore, the traffic flow levels would be taken into consideration when labelling and recognizing the driving style. The trajectory data extracted from the video are studied in the paper, which contains the identification, GPS, velocity, and acceleration.

## 2 Literature Review

In recent years, the studies about new modes of travel transportation [4, 5] and innovative approaches have been developed. Some research concentrates on the cooperative schedule to achieve the optimization [6, 7]. Machine learning algorithms applied to driving behavior recognition have been studied in some previous works. Different types of neural network (NN) algorithms have been used [8–11]. However, a larger size of the network could lead to a long training time [12]. The tree-like algorithm [13, 14] and Hidden Markov Model (HMM) [15, 16] are also adopted to detect the driving behaviors according to the extracted features. Some researchers also combined the HMM with dynamic Bayesian networks or ANN to predict the driving behavior by learning the driving data [17, 18]. While HMM requires a long training time, especially for a high number of states, the recognition time also increases with the number of states [19]. Therefore, a more suitable and effective method should be found to identify the driving style. SVM has been widely applied to various kinds of pattern recognition problems, including voice identification, text categorization, face detection [20, 21]. In addition, SVM performs well with a limited number of training samples, and SVM has fewer parameters to be determined [22, 23]. Therefore, many studies employed SVM to build driving style recognition models.

Except for unsupervised machine learning algorithms, for example, clustering, other machine learning algorithms require labelled or partially labelled driving behavior data. Some research adopted behavior-based or accident-based method to label the driving style [20, 24]. Driver self-reported questionnaire [25] and expert scoring [13] are also adopted to evaluate driving style. However, these two methods rely on drivers or experts' subjective judgments and can be very time-consuming when the number of drivers in the sample is huge. This paper proposes a new driving data label method based on collision surrogates incorporating traffic level.

### 3 Methodology

Three collision risk surrogates are used to determine the risk level of the car-following process for each following pairs. The threshold value to classify the risk level is different for normal and congested traffic. The K-means algorithm is applied to group the drivers as calm or aggressive, based on their trajectory risk levels. As the traffic flow has a great impact on driving behavior, it is considered when labelling the driving style.

#### 3.1 Collision Risk Surrogate

It is essential to find the most effective surrogates to describe the collision risk when driving on the road [26–29]. Vehicle trajectory data such as the vehicle's velocity and acceleration are usually not good enough to estimate the rear-end collision risk. In the paper, the Margin to Collision (MTC) is used to evaluate the risk.

MTC indicates the final relative position of PV and FV if two vehicles decelerate abruptly.

$$MTC = (x_r + v_p^2/2a_p)/(v_f^2/2a_f) \quad (1)$$

where  $a_f$  and  $a_p$  denote the deceleration for FV and PV, respectively. Usually, both are defined as  $0.7G$ .  $v_f$  and  $v_p$  respectively denote the velocity of FV and PV.  $x_r$  denotes relative distance. A modified MTC (MMTC) is proposed in the paper to include the following vehicle's reaction time when the PV abruptly decelerates. The equation is modified as follows.

$$MMTC = (x_r + v_f^2/2a_f - v_p^2/2a_p)/v_f \quad (2)$$

MMTC evaluates the minimum reaction time needed for FV to avoid a collision when PV abruptly decelerates. The collision risk is higher with a lower MMTC value since drivers have little time to react. MMTC can evaluate potential collision risk with the abrupt deceleration of PV.

#### 3.2 Key Features Extraction

In this paper, the vehicle acceleration  $a_f$ , relative distance  $x_r$ , and relative velocity  $v_r$  are adopted to recognize the driving style. The Discrete Fourier Transform (DFT), and Statistical method (SM) are used respectively to extract the effective key features from the vehicle trajectory. The key parameters that can capture most of the distribution information of vehicle trajectory.

### 3.3 Recognition Algorithms

Four machine learning algorithms, i.e., SVM, RF, KNN, and MLP, are adopted to build the driving style recognition model. The inputs of the model are the features extracted in Sect. 3.2. The output of the model is the driving style. The recognition process of driver's aggressive driving style is as follows:

Step 1: Use "leave-one-out" to divide the test set and training set for the model. Select one sample as the training set and the others are test set, ensuring the training set contains calm and aggressive driving styles.

Step 2: In order to avoid the influence of dimension among different trajectory variables and eliminate the differences, the min-max normalization method is used to normalize the sample data.

Step 3: The Differential Evolution algorithm (DE) is applied to optimize the parameters of algorithms, and get the initial structure value of the optimized algorithm.

Step 4: Four algorithms are used to identify the aggressive driving style under normal and aggressive traffics.

Step 5: Model performance evaluation.

## 4 Results and Discussion

In this paper, the I-80 trajectory dataset of Next Generation Simulation (NGSIM) is adopted to study driving style. According to the data analysis, the aggregate flows of HOV lanes are respectively 250 and 398 vph during two periods, indicating two levels of traffic flow. 370 Leader-follower Vehicle Pairs (LVP) are chosen under congested and normal traffic to study the driving style in this paper since fewer interrupting vehicles are from other lanes.

### 4.1 Significant Analysis Considering Traffic Levels

Based on the trajectory data extracted from NGSIM, the significant analysis between trajectory features considering traffic levels has been conducted, shown in Table 1. Table 1 shows that there is no significant difference of ITTC surrogate in different traffic levels. Besides that, the drivers tend to keep higher velocity (17.314 m/s) and lower acceleration (0.040 m/s<sup>2</sup>) when following preceding vehicles. And drivers in congested traffic flow keep higher velocity difference with preceding vehicles to keep safe, while the gap is smaller. However, the ITTC, THW, and MMTC are smaller

**Table 1** Significant analysis considering traffic level

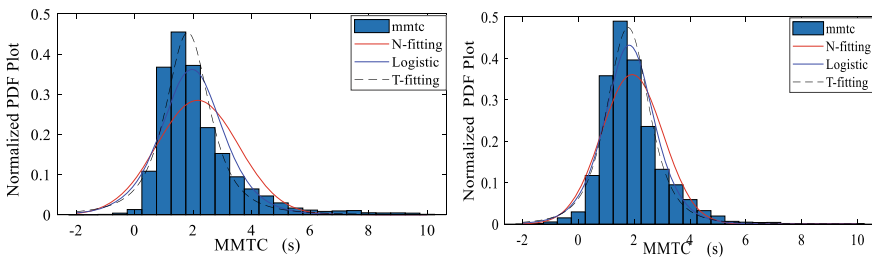
Variables	Description	Normal		Congested		P
		Mean	SD	Mean	SD	
$v_f$	The velocity of following vehicle (m/s)	17.31	3.76	13.80	3.20	0.000**
$a_f$	The acceleration of following vehicle ( $m/s^2$ )	0.04	1.93	0.06	1.69	0.005**
$v_r$	The difference of velocity between two following vehicles (m/s)	0.02	2.52	0.138	2.33	0.000**
$x_r$	The gap between two following vehicles (m)	37.75	28.97	25.90	16.07	0.000**
THW	Time to headway (s)	2.20	1.79	1.88	1.08	0.000**
ITTC	Inversed time to collision ( $s^{-1}$ )	-0.01	2.57	-0.06	15.95	0.373
MTC	Margin to collision (s)	2.24	2.08	1.94	1.46	0.000**

\*\*Significant correlation at 0.01 level (bilateral)

for congested traffic condition. Therefore, the traffic condition should be taken as a reference when labelling the risky following maneuvers.

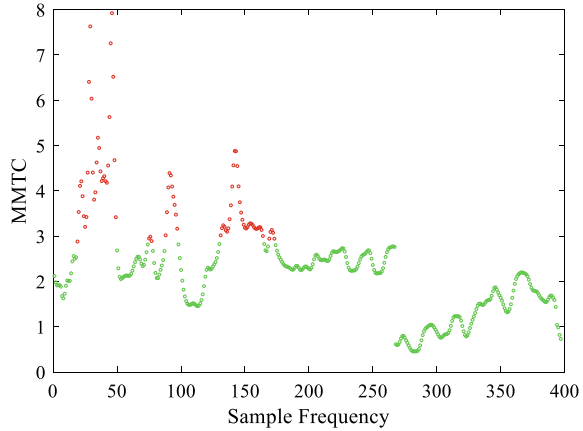
### 4.2 The Sample Data Label

Figure 1 shows the fitting curves of MMTC by adopting three distributions for normal and congested traffic flow, i.e., normal distribution, logistic distribution, and t distribution. The t distribution achieves a better fitting performance than the other two distributions. Therefore, the t distribution is adopted to determine the threshold value of features. The 85% percentile value based on cumulative distribution are also obtained as threshold to classify the car-following maneuvers into several segments with two levels of risk, i.e., safe and risky.



**Fig. 1** Statistical fitting curves for MMTC for normal traffic flow and congested traffic flow

**Fig. 2** Trajectory segments for a driver based on threshold values of MMTC



Each driver's driving trajectory can be divided into several segments, which belong to two different driving risk levels. A driver is selected to show the trajectory segments according to the threshold values of MMTC, shown in Fig. 2.

The proportions of trajectory segments with different risk levels can determine each driver's driving style. The K-means algorithm is applied to trajectory under different traffic flow to classify the drivers as calm and aggressive based on the ratio of risky maneuver. The clustering results show there are 246 calm drivers and 124 aggressive drivers under normal traffic flow, and 200 calm drivers and 170 aggressive drivers under congested traffic flow.

### 4.3 Driving Style Recognition

The SVM method is adopted to recognize the driving style under each traffic level. In this paper, the trajectory data including the vehicle acceleration  $a_f$ , relative distance  $x_r$ , and relative velocity  $v_r$  are adopted to recognize the driving style, respectively. The Discrete Fourier Transform (DFT) and Statistical method (SM) are respectively used to extract the effective key features from the vehicle trajectory. The z-score method is adopted to standardize features before model training.

In the study, the accuracy, precision, and recall rates are assessed to evaluate the model's ability to recognize aggressive drivers among all vehicles. The performance of the recognition model is evaluated using the "leave-one-out" cross-validation method. Driving style recognition results based on different feature extraction methods and SVM are shown in Table 2. Except mentioned, the SVM algorithm uses the linear kernel function.

As shown in Table 2, the DFT outperforms SM in feature extraction, with an accuracy of 91.7%. With any combinations of features, the accuracy rate of the SM is lower than that based on DFT. In general, the feature  $x_r$  and  $v_r$  perform better than

**Table 2** The recognition results of driving style based on SVM with DTF and SM

Features	DFT (%)	SM (%)
af	67.0	66.2
vr	83.2	79.7
xr	88.9	83.5
af + vr	83.2	78.7
vr + xr	87.8	83.2
af + xr	88.9	84.6
af + vr + xr	91.7	85.7

**Table 3** The recognition results of driving style based on RF

Features	Algorithm	Accuracy (%)	Precision (%)	Recall (%)
af + vr + xr	RF	91.6	89.2	82.0
	MLP	88.1	73.0	87.3
	KNN	87.6	85.7	76.2
	SVM	91.7	92.8	81.8

**Table 4** The recognition results of driving style with and without traffic level

Features		Accuracy (%)	Precision (%)	Recall (%)
af + vr + xr	With context	91.7	92.8	81.8
	No context	76.50	87.5	68.4

$a_f$  in recognizing the driving style. A possible reason is that the rear-end collision risk determines the driving style label, the feature  $a_f$  cannot accurately describe the relative motivation between two following vehicles.

The performance of four machine learning algorithms RF, MLP, KNN and SVM using all features with DTF method are compared. The accuracy, precision, and recall rates are listed in Table 3. It can be seen that SVM outperforms other machine learning algorithms. Random Forest is the second best algorithm. However, MLP gives the highest recall rate among all candidates. KNN, as the simplest classification method, unsurprisingly obtains the worst performance. As seen from Table 4, the recognition accuracy of driving style without traffic levels is 76.50%, lower than that with context. Therefore, the traffic levels should be taken into consideration.

## 5 Conclusion

In this study, a novel driving style label method is proposed to assign calm and aggressive labels based on collision risk incorporating traffic levels, which is critical

to sample data needed in supervised machine learning. The main findings could be summarized as follows:

The rear-end collision risk surrogates, namely MMTTC, are adopted to evaluate the risk during the car-following process. Each driver's trajectory can be divided into two risk levels incorporating traffic levels, and all drivers can be grouped into calm or aggressive using the K-means algorithm.

Two feature extraction methods are compared in the recognition model. Three machine learning algorithms including RF, MLP, and KNN are also adopted to compare with the SVM. The results show that DFT could better capture the characteristics of driving behavior. The driving style can be recognized with the highest accuracy of 91.7% using SVM.

This study offers the possibility of developing more sophisticated driving style recognition methods. For further work, the proposed method can be extended by selecting other features that can reflect the driving style more accurately. As we know, the driving style is also influenced by the road conditions. Such results can also be used to improve driving style recognition.

## References

1. Gao, K., Tu, H., Sun, L., et al.: Impacts of reduced visibility under hazy weather condition on collision risk and car-following behavior: implications for traffic control and management. *Int. J. Sustain. Transp.* **14**(8), 635–642 (2020)
2. Zhou, M., Qu, X., Li, X.: A recurrent neural network based microscopic car following model to predict traffic oscillation. *Transp. Res. Part C Emerg. Technol.* **84**, 245–264 (2017)
3. Zheng, Y., Li, S.E., Wang, J., et al.: Stability and scalability of homogeneous vehicular platoon: study on the influence of information flow topologies. *IEEE Trans. Intell. Transp. Syst.* **17**(1), 14–26 (2015)
4. Gao, K., Yang, Y., Li, A., Li, J., Yu, B.: Quantifying economic benefits from free-floating bike-sharing systems: a trip-level inference approach and city-scale analysis. *Transp. Res. Part A Policy Pract.* **144**, 89–103 (2021)
5. Xu, Y., Zheng, Y., Yang, Y.: On the movement simulations of electric vehicles: a behavioral model-based approach. *Appl. Energy* **283**, 116356 (2021)
6. Gao, K., Yang, Y., Sun, L., Qu, X.: Revealing psychological inertia in mode shift behavior and its quantitative influences on commuting trips. *Transp. Res. Part F Traffic Psychol. Behav.* **71**, 272–287 (2020)
7. Wu, J., Kulcsár, B., Ahn, S., et al.: Emergency vehicle lane pre-clearing: from microscopic cooperation to routing decision making. *Transp. Res. Part B Methodol.* **141**, 223–239 (2020)
8. Molchanov, P., Gupta, S., Kim, K., et al.: Multi-sensor system for driver's hand-gesture recognition. In: *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–8. IEEE (2015)
9. Hashemi, A., Saba, V., Resalat, S.N.: Real time driver's drowsiness detection by processing the EEG signals stimulated with external flickering light. *Basic Clin. Neurosci.* **5**(1), 22 (2014)
10. Wang, H., Zhang, C., Shi, T., et al.: Real-time EEG-based detection of fatigue driving danger for accident prediction. *Int. J. Neural Syst.* **25**(02) (2015)
11. Srinivasan, D., Jin, X., Cheu, R.L.: Adaptive neural network models for automatic incident detection on freeways. *Neurocomputing* **64**(1), 473–496 (2005)



12. Singh, N.K., Singh, A.K., Tripathy, M.: A comparative study of BPNN, RBFNN and ELMAN neural network for short-term electric load forecasting: a case study of Delhi region. In: International Conference on Industrial and Information Systems. IEEE (2015)
13. Bejani, M.M., Ghatee, M.: A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data. *Transp. Res. Part C Emerg. Technol.* **89**, 303–320 (2018)
14. Li, G., Li, S.E., Cheng, B., et al.: Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities. *Transp. Res. Part C Emerg. Technol.* **74**, 113–125 (2017)
15. Berndt, H., Dietmayer, K.: Driver intention inference with vehicle onboard sensors. In: 2009 IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 102–107. IEEE (2009)
16. Meng, X., Lee, K.K., Xu, Y.: Human driving behavior recognition based on hidden Markov models. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics (2007)
17. Kumagai, T., Sakaguchi, Y., Okuwa, M., et al.: Prediction of driving behavior through probabilistic inference. In: Proceedings of the Eighth International Conference on Engineering Applications of Neural Networks (2003)
18. Zong, C.F., Yang, X., Wang, C., et al.: Driver's driving intention identification and behavior prediction during vehicle steering. *J. Jilin Univ. (Eng. Technol. Ed.)* **s1**, 27–32 (2009)
19. Aoude, G.S., Desaraju, V.R., Stephens, L.H., et al.: Driver behavior classification at intersections and validation on large naturalistic data set. *IEEE Trans. Intell. Transp. Syst.* **13**(2), 724–736 (2012)
20. Chen, Z., Wu, C., Huang, Z., et al.: Dangerous driving behavior detection using video-extracted vehicle trajectory histograms. *J. Intell. Transp. Syst.* **21**(11) (2017)
21. Sun, W., Zhang, X., Peeta, S., et al.: A real-time fatigue driving recognition method incorporating contextual features and two fusion levels. *IEEE Trans. Intell. Transp. Syst.* **18**(12), 3408–3420 (2017)
22. Feng, L., Yao, Y., Jin, B.: Research on credit scoring model with SVM for network management. *J. Comput. Inf. Syst.* **6**(11), 1032–1040 (2010)
23. Martinez, C.M., Heucke, M., Wang, F.Y., et al.: Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey. *IEEE Trans. Intell. Transp. Syst.* **19**(3), 666–676 (2017)
24. Teimouri, F., Ghatee, M.: A real-time warning system for rear-end collision based on random forest classifier. arXiv preprint [arXiv:1803.10988](https://arxiv.org/abs/1803.10988) (2018)
25. Ishibashi, M., Okuwa, M., Doi, S., et al.: Indices for characterizing driving style and their relevance to car following behavior. In: SICE Conference (2007)
26. Ambros, J., Turek, R., Paukrt, J.: Road safety evaluation using traffic conflicts: pilot comparison of micro-simulation and observation. In: International Conference on Traffic and Transport Engineering (2014)
27. Kitajima, S., Marumo, Y., Hiraoka, T., et al.: Comparison of evaluation features concerning estimation of driver's risk perception. *Trans. Soc. Automot. Eng. Jpn.* **40**(2), 191–198 (2009)
28. Zhang, T., Gao, K.: Will autonomous vehicles improve traffic efficiency and safety in urban road bottlenecks? The penetration rate matters. In: 2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE), pp. 366–370. IEEE (2020)
29. Yu, Y., Jiang, R., Qu, X.: A modified full velocity difference model with acceleration and deceleration confinement: calibrations, validations, and scenario analyses. *IEEE Intell. Transp. Syst. Mag.* (2019)