Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2017) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Connected & Intelligent Mobility (CIM 2017)

A Deterministic Feedback Model for Safe Driving based on Nonlinear Principal Analysis Scheme

Minsu Jeon[a], Eunju Yang[a], Eunyoung Oh[a], Jeongkyu Park[b], Chan-Hyun Youn[a] *

[a]*School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, Korea*
[b]*Big Data Planning Team, KIA Motors Corporation, Seoul, Korea*

Abstract

For intelligent driver assistant service, research works have been conducted using sensors to analyze drivers' behavior and score for safe driving. There have been many studies on safe driving, but there was a limit on providing the driver with a comprehensive solution for safe driving. The studies mainly focused on a method to quantify the driving characteristics or the indicators that can be used to assess the driving risk. In order to practically utilize those indicators, it is necessary to provide the driver with the feedback on how to change which controllable factors to fix the indicators to desired values. However, there has been no study to provide feedback on interpretable factors. To improve the problem, in this paper, we propose a model that provides comprehensive feedback to the driver by offering quantitative driving improvement instructions. In the proposed model, K-means clustering is used to classify the safe driving level and the non-linear principal component analysis (NLPCA) model is trained by the classified low-risk data to analyze arbitrary driving data and provide feedback. To evaluate proposal model, we collected sensor data while driving vicinity of Daejeon Metropolitan City in Korea, and analyzed the principal component extracted using the NLPCA. We evaluated the classification accuracy of the principal components to verify the validity of the proposed model, and showed that the characteristics of safe driving can be represented through the proposal feedback model.

Keywords: Advanced Driver Assistance Systems; Intelligent Vehicle; Driving Risk; Machine Learning; Safe Driving Analysis;

* Corresponding author. Tel.: +82-042-350-3495.
E-mail address: chyoun@kaist.ac.kr

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.

1. Introduction

Detecting driver behavior and assessing the risk of accidents are important parts of the advanced driver assistance system(ADAS)[1]. If an accident caused by an abnormal driving behavior is detected and reacted in advance, an accident can be prevented. In the case of drowsy driving, a driver may avoid an accident by a warning from the assistance system[2]. These kinds of driving assistance systems have been studied extensively over the last decades. Driving risk models were studied to classify the driving hazard[3] and make a numeric model for time-to-x (TTX) to assess criticality[4]. In addition to analysis of driving risk model, general driving patterns was also studied, e.g., visualization of driving pattern analysis was conducted using extracted hidden features[5]. In addition, mobile application[6] was developed that can detect inattentive driving by using various sensor data and then reflect the score to the driver.

Even though the previous studies[3,4,7,8,9] provided driver assistance system by detecting the risk of accidents and scoring the driving behavior, how to improve the drive for safety cannot be provided since only the assessing indicators were provided and the factors that can be manipulated were not offered. As an example, the quality of the lane change behavior was analyzed using image data and kinematics data[9], the criticality was assessed[4,7], and driving risk was analyzed[3,8]. However, if the indicator, which can be manipulated and guide drivers to drive more safely, is given, then driving behavior can be improved to reduce accidents by reflecting the practical driving feedback. The numeric driving feedback to improve safety is not only helpful to the driver, but can also be reflected in intelligent automotive systems. Therefore, in this paper, we propose a model detecting driving risk level based on collected mobile sensor data and providing quantitative driving indicators about how to alleviate accident risks. In the proposed model, measured sensor data are pre-processed to classified by the three risk-levels through the K-means clustering. The clustered low-risk data are fed into the nonlinear principal analysis (NLPCA)[10] to train it and the trained model evaluates the numeric safe driving feedback for driving behavior inputs.

The remainder of this paper is organized as follows. Section 2 describes the model of our work. Evaluation for what we propose is provided in Section 3 and we finally conclude the paper in Section 4.

2. A Deterministic Feedback Model for Safe Driving

To handle the numerical feedback for safe driving, we proposed a quantitative feedback model, which consists of two main steps: model training and inference analysis (refer to Fig. 1). Firstly, in model training step, in-vehicle sensed data is collected and pre-processed. To analyze driving behavior and provide feedback to the driver, we used the NLPCA auto-encoder model. Before training the model, the pre-processed data is classified to label the risk level, and then the low risk labeled data as a safe driving data is used to train the NLPCA model. After training the model, in the inference analysis step, the reconstructed values can be obtained from the observed input data through the trained model. Analyzing the observed data and its reconstructed values, the quantitative feedback for safe driving is provided to the driver.

2.1. Data Preparation and Training the Model

- Data preparation

To implement the feedback service for safe driving, local dynamic status information of the vehicle is required as an input for the model. Therefore, we defined the sensor feature vector $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ to describe the behavior of the driver and be used as the input to the NLPCA model. In the sensor feature vector $\mathbf{x}$, each element x represents the characteristic indicator value of the sensor data. The collected raw sensor data is pre-processed to construct the sensor feature vector. In this paper, as the elements of the sensor feature vector, we employed 4 sensing types (acceleration in x-axis and y-axis, yaw rate, and velocity), each has 10 characteristic indicators as shown in Table 1.

- Risk labeling

To train the NLPCA model using the safe driving data, the criteria for safe driving is required. Therefore, we applied the method that classifies data to the several risk levels[3]. In the previous study[3], it was shown that the driving risk level could be classified by the minimum acceleration, the average acceleration, and kinetic energy reduction ratio that are shown to be correlated to the crash occurrence rate using the K-means clustering method. In this study, using those 3 indicators we classified the data into 3 risk level: low risk, moderate risk, and high risk through the K-means

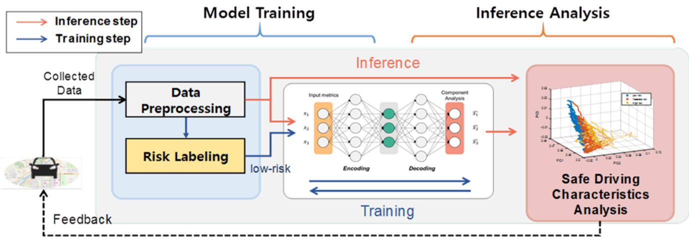

Fig. 1. NLPCA-based feedback model.

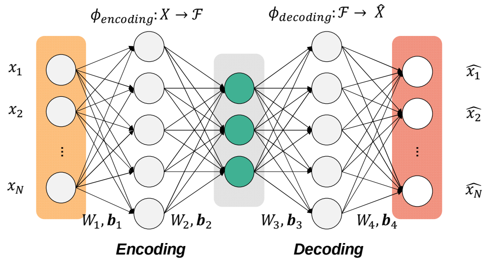

Fig. 2. NLPCA auto-encoder structure.

Table 1. Characteristic indicators for constructing the sensor feature vector.

Characteristic Indicators									
Minimum value on each sensing type	Maximum value on each sensing type	Average value on each sensing type	Standard deviation value on each sensing type	Peak to peak value on each sensing type	Minimum value of differentiation on each sensing type	Maximum value of differentiation on each sensing type	Average value of differentiation on each sensing type	Standard deviation value of differentiation on each sensing type	Peak to peak value of differentiation on each sensing type

clustering method. We defined the safe driving as the driving behavior represented on driving data labeled with the low risk, and used pre-processed data labeled with the low risk to train the NLPCA model.

- Training process of the NLPCA model

The NLPCA[10] is a model used to extract nonlinear key components using an auto-encoder with a nonlinear layer. As shown in Fig. 2, as the NLPCA model, the auto-encoder network consists of an encoder that extracts the main component from the input data and a decoder that reconstructs the input data from the main component. When the input of the auto encoder $\mathbf{x}$ is encoded, the model is configured to extract the characteristics of the input data and reconstruct the original data through decoding. The output of encoder which contains the extracted characteristics of the input data is called principal components. The output of each neuron, which makes up the auto-encoder, is given as the value of the activation function f for $wx+b$, where w is weight, x is input of the neuron and b is bias. All weights and biases for one layer are defined as the matrix $\boldsymbol{W}$ and the vector $\boldsymbol{b}$ for convenience. As the estimate of $\mathbf{x}$ based on the features extracted through the neuron layer, the reconstructed vector $\hat{\mathbf{x}}$ through each layer can be expressed as $\hat{\mathbf{x}} = f \circ f \cdots \circ f(x)$.

In the proposal model, NLPCA model learns the characteristics of safe driving training by the low risk labeled data. As shown in Fig. 1, the NLPCA model receives the pre-processed sensor feature vector data $\mathbf{x}$ as an input and finally derives reconstructed value $\hat{\mathbf{x}}$ from the auto-encoder.

Before the inference analysis is performed, the NLPCA model should be trained with safe driving data. The goal of the auto encoder is to extract features that make the inputs of encoding and results of decoding to be equal. Therefore, the training of the auto encoder proceeds in the direction of minimizing the error between $\hat{\mathbf{x}}$, which is the decoding result of all models, and $\mathbf{x}$, which is given as input. In this paper, we applied the mean square error (MSE) $L(x) = \|x - \hat{x}\|^2$ as the loss function which is widely used to minimize the error between input value and reconstructed value. To find the weight w and bias b that minimize the loss function, the model is trained through the backward propagation which updates the weight and bias in the gradient direction of the cost function as a general learning method as shown in Algorithm 1.

2.2. Inference and its Analysis

After training the model, to give driver the quantitative feedback for safe driving, decoding is conducted to calculate reconstructed value $\hat{\mathbf{x}}$. As shown in Fig. 1, the input data and the reconstructed values extracted by the trained model are received to analyze them and provide the driver with the feedback for safe driving. As the NLPCA model was trained by safe driving data, the trained model is expected to extract safe driving hidden features for arbitrary data inputs, and the reconstructed value from the decoding step can represent the nearest safe driving data values in the input dimension. As the reconstructed value $\hat{\mathbf{x}}$ indicates representative low risk input values, distance between the

Algorithm 1. Training algorithm for NLPCA[10] based feedback model.

Initialize weights W and bias b randomly
for *each epoch* **do**
  for *each minibatch of size* N_B **do**
    $g_W \leftarrow 0$, $g_b \leftarrow 0$
    for *each case* k **in** *the minibatch* **do**
      if *label* $\neq$ *low risk* **then**
        $x \leftarrow x_{(k)}$
        $\hat{\mathbf{x}} \leftarrow f \circ f \circ \cdots \circ f(x)$
        $L(x) \leftarrow \|x - \hat{x}\|^2$
      end if
    end
    accumulate $g_W \leftarrow g_W + \frac{\partial L(x)}{\partial W}, g_b \leftarrow g_b + \frac{\partial L(x)}{\partial b}$
  end
  update the parameter $W \leftarrow W - \frac{\lambda}{N_B} g_W, b \leftarrow b - \frac{\lambda}{N_B} g_b$
end
return W, b

sensor feature vector and reconstructed vector $\hat{x} - x$ can be provided to the driver as a quantitative feedback for the safe driving.

3. Evaluation

In this study, in order to verify the validity of the analysis model, we evaluated how accurately principal components extracted from the analysis model can classify the driving risk. Through this evaluation, we are to demonstrate that the NLPCA based analysis model can reasonably analyze the driving risk from input sensor data.

3.1. Experimental Environments

To collect the sensor data, we drove the routes nearby Daejeon Metropolitan City in Korea as shown in Fig. 3, and collected the sensor data using the sensor recording application of the mobile phone while driving. We applied time window with 5 seconds length, and aggregated the data that were interpolated into 20 frames on each 0.1 second step. We collected 14078 steps of pre-processed data in total.

3.2. Results and Discussion

As mentioned in section 2.1, we classified the collected data into low, moderate, and high risk using K-means clustering method, and labeled the collected data. We trained the NLPCA based feedback model only using the data labeled with low risk to obtain the analysis model for low risk which is considered as safe driving. The principal components are extracted by using the trained model, and then we tried to verify the validity of the feedback model by evaluating the risk classification accuracy of the principal components through the support vector machine (SVM) which is widely used classification method for the supervised learning.

We firstly observed the distributions of the collected data in each risk level and values reconstructed through the trained model. The feedback model was trained using 85% of the total collected dataset, and the rest 15% of the dataset was feedforwarded through the trained model for evaluation. The distributions of each risk level data and reconstructed values in the main indicator axes of the sensor feature vector are shown in Fig. 4, and Fig. 5 illustrates the distribution of each risk level data in the axes of the principal components extracted from the model that was trained by low risk data. It is possible to distinguish the low risk level by using only 3 principal components as much as using several indicators of the sensor feature vector. In addition, as shown in Fig. 4, it is shown that the distribution of the reconstructed values for moderate and high risk data is represented in the center of low-risk data distribution, therefore it can be inferred that the reconstructed values of the trained model represent the nearest low risk component values.

Table 2 shows the statistical indicators that represents how accurately can the principal components extracted from the trained model classify the risk level through SVM. All collected data were feedforwarded through the trained

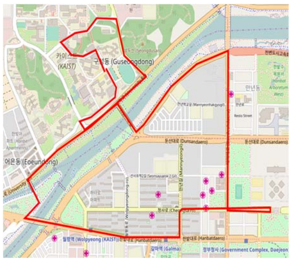

Fig. 3. A sample of driving route for collecting the sensor data.

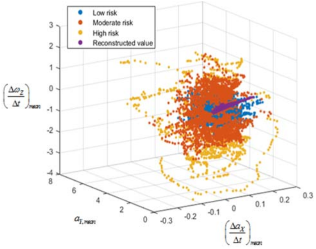

Fig. 4. Risk level distribution for the driving behaviour vector.

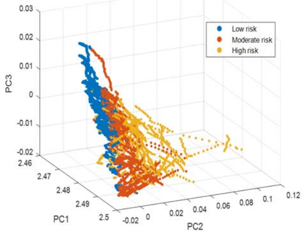

Fig. 5. Risk level distribution for the principal components.

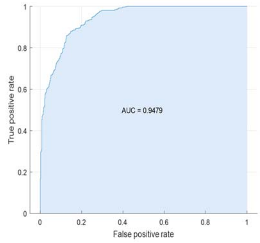

Fig. 6. ROC curve of classifying the risk level from extracted principal components.

Table 2. Confusion matrix and statistical indicators for classifying the risk level.

		Predicted class		F-measure	Total accuracy
		Safe driving	Risky driving		
True class	Safe driving	0.5022	0.0105	0.8554	0.8303
	Risky driving	0.1592	0.3281	0.7945	

Table 3. Quantitative safe driving feedback for risky driving behavior.

	Observed indicators			Reconstructed values			Feedback		
	Average differentiation of a_X	Peak to peak of a_Y	Minimum differentiation of ω_Z	Average differentiation of a_X	Peak to peak of a_Y	Minimum differentiation of ω_Z	Average differentiation of a_X	Peak to peak of a_Y	Minimum differentiation of ω_Z
Case1	-0.1675	0.6747	-0.0053	0.1244	1.0708	-0.0563	**58.89%**	6.31%	-24.55%
Case2	0.0217	5.0288	-0.0015	0.0415	0.9383	-0.0246	3.99%	**-65.15%**	-11.12%
Case3	0.0859	2.3813	-0.1537	-0.0143	0.6157	-0.0056	-20.21%	-28.12%	**71.30%**

feedback model, and the hidden 3 principal components were obtained for the dataset. Using 85% of the obtained principal component dataset, the linear SVM was trained, and the inference was conducted for the rest 15% of the principal component dataset to evaluate the validity of the feedback model. As shown in Table 2, it is observed that the total accuracy is 0.8303, the F-measure values for detecting the safe driving and risky driving are 0.8554, 0.7945 respectively, and the receiver operating characteristic (ROC) curve shown in Fig. 6 indicates that the area under curve (AUC) is 0.9479. The evaluated statistical indicators show that the extracted principal components can accurately classify the risk level, and this implies that the NLPCA based model is precisely trained. This result shows that the proposed model can more precisely extracts features of the safe driving than the decision tree based model[3] that predicts the risk level using the current information of driver characteristics and road environment. However, the accuracy to extract features of safe driving for the proposed model is lower than the accuracy for the maneuver transition probabilities based model[11] that uses additional camera image data as well as the sensor data. Although the proposed model only uses sensor data that can be collected from the smartphone, it achieves slight lower accuracy of the model that uses both sensor data and camera image from the controller area network (CAN) bus. Therefore, the proposed model can take advantages on collecting the data in cheaper cost, but the slight accuracy reduction occurs compared to using the various type of data. Moreover, the critical characteristics of the proposed model is that the model can provide the corrective feedback for safe driving to the driver with a proper accuracy comparing to the conventional methods[3, 11] that can only alarm the risk level.

Table 3 shows values of the evaluated sensor feature vector for 3 points in high risk data, their reconstructed values, and their quantitative feedback for safe driving. These 3 points are selected from the distribution of high risk data. To provide more comprehensive feedback, we provide the how much percentage to change each factors that is expressed as $(\hat{x} - x)/(x_{max} - x_{min})$. In case1, the average differentiation of longitudinal acceleration was observed to be lower than the reconstructed value, so the feedback for increasing the average differentiation of longitudinal acceleration was mainly provided. Likewise, in case2, feedback on the peak to peak value of the lateral acceleration was offered to decrease the lateral acceleration fluctuation since the distance between observed data and reconstructed value on that factor was long. In case3, feedback for increasing minimum rotation acceleration in z-axis was provided as the observed minimum differentiation of yaw rate was relatively lower than the reconstructed value. To evaluate the validity of the reconstructed value, we roughly calculated risk indicating information by $a_{mean} \approx \sqrt{(a_{X,mean})^2 + (a_{Y,mean})^2}$, $a_{min} \approx$

$\sqrt{(a_{X,min})^2 + (a_{Y,min})^2}$ and classified risk level from the calculated values through K-mean clustering whose parameters were used to label the data. It is observed that 83.14% of reconstructed values are classified as the low risk data, and the rest of them are classified as the moderate risk. Therefore, it is verified that the reconstructed values mostly represent the low risk data. As the validity of the NLPCA-based safe driving feedback model was verified, it can be seen that the quantitative feedback for safe driving using the reconstructed value becomes feasible.

4. Conclusion

In this paper, we proposed an NLPCA-based analytical method that can provide quantitative feedback to the driver. The validity of the trained NLPCA-based model was tested through the accuracy of detecting safe driving data and risky driving data from extracted principal components. Using the reconstructed value, which is one of the critical features of NLPCA, quantitative analysis for safe driving feedback becomes possible by analyzing the difference between observed data and its reconstructed values that indicate the representative values of low risk driving.

Using a machine learning method to analyze the driving risk from the dynamics of vehicle is consistent with the traditional studies that use the other analytic method to detect the driving risk. However, in addition to the traditional methods, our method enables to provide the driver with the quantitative feedback for safe driving.

Our proposed method only considers the behavior of the ego-vehicle and analyzes only the behavior of the one vehicle. Considering the behavior of surrounding vehicles and various factors such as road surface condition, weather, and behavior of driving additionally, it can be possible to construct the more precise model with an expanded scale. Moreover, in order to modeling the safe driving analysis more accurate and detailed, future work could analyze the safe driving patterns separately on each driving behavior such as the line change, left turn, etc. It is also possible to evaluate the feasibility of this service application in the connected car environments using the simulator[12].

Acknowledgements

This work was partly supported by Hyundai Motor Company and KIA Motors Corporation and Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.R7117-16-0235, Development of HPC System for Accelerating Large-scale Deep Learning).

References

1. Paul, Aneesh, et al., “Advanced Driver Assistance Systems.” SAE Technical Paper, 2016.
2. S. Kaplan, M. A. Guvensan, et al., “Driver behavior analysis for safe driving: a survey.” IEEE Transactions on Intelligent Transportation Systems, 2015. pp. 3017-3032.
3. J. Wang, Y. Zheng, et al., “Driving risk assessment using near-crash database through data mining of tree-based model.” Accident Analysis & Prevention, 2015. pp. 54-64.
4. A. Tamke, T. Dang, and G. Breuel., “A flexible method for criticality assessment in driver assistance systems.” Intelligent Vehicles Symposium (IV), 2011.
5. HaiLong Liu, Tadahiro Taniguchi, et al., “Visualization of driving behavior based on hidden feature extraction by using deep learning.” IEEE Transactions on Intelligent Transportation Systems, 2017.
6. L. M. Bergasa, D. Almeria, et al., “Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors.” Intelligent Vehicles Symposium Proceedings, 2014.
7. M. Schreier, V. Willert, and J. Adamy, “An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments.” IEEE Transactions on Intelligent Transportation Systems, 2016. pp. 2751-2766.
8. C. Laugier, I. E. Paromtchik, et al., “Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety.” IEEE Intelligent Transportation Systems Magazine, 2011. pp. 4-19.
9. R. K. Satzoda, P. Gunaratne, and M. M. Trivedi., “Drive quality analysis of lane change maneuvers for naturalistic driving studies.” Intelligent Vehicles Symposium (IV), 2015.
10. M. Scholz, M. Fraunholz, and J. Selbig, “Nonlinear principal component analysis: neural network models and applications." Principal manifolds for data visualization and dimension reduction. Springer Berlin Heidelberg, 2008. pp. 44-67.
11. G. Li, S. E. Li, et al., “Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities.” Transportation Research Part C: Emerging Technologies, 2017. pp. 113-125.
12. Heejae Kim, Jiyong Han, et al., IsV2C: An Integrated Road Traffic-Network-Cloud Simulator for V2C Connected Car Services, SCC, 2017.