

An Evolutionary Approach to Driving Tendency Recognition for Advanced Driver Assistance Systems

Jong-Hyun Lee and Chang Wook Ahn

Department of Computer Engineering, Sungkyunkwan University (SKKU), Suwon 440-746, Republic of Korea

Abstract. Driving tendency recognition is important for constructing Advanced Driver Assistance Systems (ADAS). However, it had not been a lot of research using vehicle sensing data, due to the high difficulty to define it. In this paper, we attempt to improve the learning capability of a machine learning method using evolutionary computation. We propose a driving tendency recognition method, with consideration of data characteristics. Comparison of our classification system with conventional methods demonstrated the effectiveness and accuracy over 92% in our system. Our proposed evolutionary approach is confirmed that improve the classification accuracy of the learning method through evolution in the experiment.

1 Introduction

Automobile safety driving support apparatus is one of the major issues in the world automobile industries. The technology has been actively developed and commercialized. Some technologies such as pre-crash safety [1], forward vehicle collision warning [2], brake assist [3], blind spot warning [4], and lane keeping assist [5] are being commercialized and applied to luxury cars. The products are expected to be extended to all cars as well as luxury cars.

Safety of cars driving is significantly affected by driving tendencies of the driver as well as the safety support systems because the support of driver assistance systems is limited and human is still driving manly. However, the driver's inclination is difficult to define as objective information. The existing research has been carried out to quantify subjective driving through the driver's inclination Experience Index rating [6]. However, the researches rarely used the combined information with the sensor and driving data. Also, there are additional barriers to develop the safety driving support system that can recognize aggressive driving behavior of the driving tendency. The problems are that people have too many type of driving patterns, additional equipment such as camera and bio-signal sensors increase the cost, the driving tendency may change as situation and time, and so on.

In this paper, we design a driving pattern recognition algorithm for the driver assistance systems that use only basic signal data of electronic control units (ECU) in the vehicle collected by controller area network (CAN). The main idea is that the signal combinations in the many factors such as vehicle acceleration, engine speed, torque, and steering wheel angle reflect the driver's inclination.

The algorithm requires a high accuracy and flexibility due to the complex and changing the objective function. Also, there are many dozens of data sources (i.e., features) that can be influence by the driver's inclination among thousands of the sources. The data consisting of many features are collected dozens of times every second in the real driving environment. It is difficult to define the optimal pre-processing function that is able to preserve the important information of the driver, but the raw data have too huge size to learn the pattern using existing methods in this environment. To solve this problem, we design high performance machine learning algorithm based on evolutionary computation and random forest algorithm for the big and complex dataset.

The method we propose in this research is novel in terms of technology convergence of evolution and machine learning technique. Also, it is meaningful that the challenge issue about the driving tendency is attempted to solve by applying advanced information technology. In this work, random forest which is widely used for pattern learning is tried to improve by evolutionary computation for adapting to the tendency changes and obtaining the high accuracy of driving pattern classification. For the validation of the proposed algorithm, we collect real driving data for ten expert drivers and perform the comparison experiment with existing algorithms.

The rest of the paper is organized as follows: Section 2 describes the proposed evolutionary approach for the recognition system in detail and in Section 3 the experimental results are depicted. Conclusion is placed in Sections 4.

2 Proposed Approach

Operation tendency of a vehicle reflects the emotion states of the driver. The tendency recognition is the main point of the driving assistance systems, particularly for the intelligent active safety alert systems. Also, it is significantly important in the improvement of effectiveness and stability of intelligent safety systems that alert a collision to driver.

2.1 Data Acquisition

The tendency unpredictability appears in the operation. Sometimes the signals may be clearly distinct and be too minuteness to distinguish. For this reason, the recognition system uses the sensor information as possible that associate with operations of the driver.

In this study, vehicle data were collected in Korean K9 model of Hyundai Motors by CAN on-board diagnostics (OBD) device with universal serial bus. About a hundred sensor data such as height, velocity, steering, gear, lane keeping, and fuel consumption were selected among a thousand data in ECU based on the relationship with the driver. The frequency for data collection was 1000~2500Hz. The raw data was used to prevent loss of information. In order to collect a variety of drivers' patterns, ten expert drivers travelled the same course of approximately 15km at the similar time of day as shown in Figure 1. In order to ease tensions on the unfamiliar environment, the course driving for each driver was performed twice; the second data has been used in the experiments. As the result, the huge size of data was several gigabytes were collected. The new method to handle and learn the data was required for the big and complex dataset.



Figure 1. Ten expert drivers travelled the same course.

2.2 Driving Tendency Recognition

Random forest learning algorithm [7] is widely used for handling big data, pattern classification, and recognition. The ensemble principle is the key point of random forest. The algorithm constructs many weak-classifiers consisting of decision tree for randomly selected feature or sample subset in a training dataset. The each decision tree is learned in the given data subset to produce a decision making model. In other words, the feature points in their dataset are examined to divide the conditions for deriving an effective decision. In the classification, the

algorithm collects the predicted decision results that are obtained by the models, respectively. The final decision is made by a majority rule of the results.

The main point of this research is that the subset sampling process of features and samples can be improved by evolutionary computation [8] because it is perform by the random rule. Although the random rule is an advantage to relieve the overfitting, it is bring up the decrease the learning capability. In this research, it is important to obtain the search capability of the learning algorithm due to the characteristic of the complex objective function. Thus, the random forest algorithm is designed as evolutionary approach.

2.2.1 Overall process of the proposed method

The proposed algorithm is performed as shown in the Figure 2. Selected data subsets by the rules that are randomly generated at the initial state are assigned to each weak classifier. The classifiers are constructed based on the subsets, and then they predict the class (i.e., each driver) of training samples and make a decision result by the majority rule. The results are evaluated based on the classification accuracy to give the fitness for evolutionary individuals. The population consisted of the individuals pass though the genetic process such as selection, crossover, mutation, and again evaluation iteratively. When the individuals are converged, the algorithm stops repeating and then derives one of the best in individuals as the result that means the optimal subsets for the classification. In the next step of using the trained model, it performs ensemble classification using the random forest that generated from the data of the optimum individual.

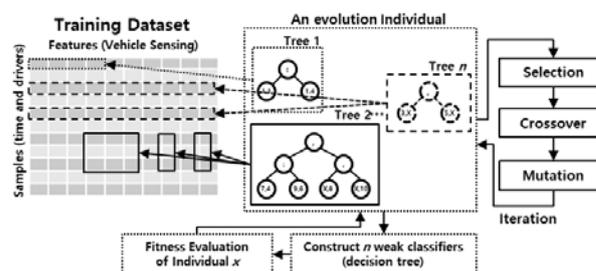


Figure 2. The evolutionary process of the proposed learning algorithm.

2.2.2 Structure of an evolution individual

An individual is designed for data subset selection in the tree structure. The tree has two types of nodes branch and leaf. The leaf nodes are assigned the coordinates of the subset in the full dataset. The m-th column n-th row is represented as $\{n, m\}$. Also, it can have an operator that selects all in a row or column. In the initial, the node is assigned a random value within the data size. Then, that can be changed by the genetic operators such as crossover and mutation. The branch nodes have the range selection operator of the comma and colon. The comma is union operator, which combines multiple references into one reference. The colon is range operator, which

produces one reference to all the cells between two references, including the two references. The operators may constitute a region with reference to the value of the child node. Sometimes the subset region can cause incomplete problems for the features and samples. Then, the classifier uses the dataset in complete area.

2.2.3 Genetic operators for the evolution

The crossover operator consisting of internal and external crossover as shown in Figure 3 leads the interactions between the individuals or the trees in an individual. These crossover methods are designed in consideration of the characteristics of a number of tree-based individual. In the external crossover, the two subtrees are selected in different two individuals and are swapped. The internal crossover is performed in the same individual.

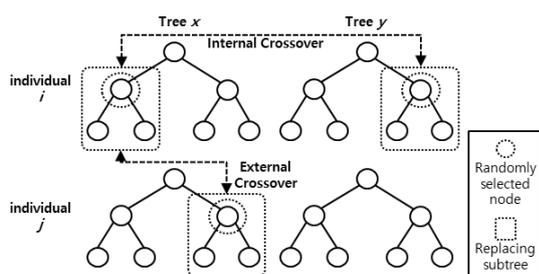


Figure 3. A genetic operator: internal and external crossover.

Selection and mutation process follow the rule of existing tree based evolutionary algorithms. The selection replicates the higher fitness object to the next generation through a competitive. The mutation performs such as changing the value of any node of an individual for increasing the search capability.

2.3 Use of Driving Data

Each driver has a specific driving tendency. The tendency may be evaluated by the statistical values that appear in the operation such as involvement lane, over speeding, and sudden starts and stops. In this research, we gave the risk ratings to the experts based on the safety driving instructions of traffic agency for the full driving data of them. In other words, some drivers in a lot of the violation had a high-risk rating. Other drivers keeping regulations had a good safety rating. It means that can generate a learning model for the risk driving based on the tendency. They can be used for evaluation of diver's inclination in the driving. For this, the accuracy of the learning model is more important than anything. Thus, the experiments in the following section perform with respect to the learning accuracy.

3 Experimental Results

The simulation environment of the proposed method was constructed in C++ with parameters including the number of evolution individuals, depth limitation of tree in the individuals, iteration, probabilities of crossover, and

mutation set at 100 individuals, 4 depths, 100 generation, 0.7, and 0.1, respectively. The maximum depth of decision trees is unlimited. The number of decision trees is 100.

Figure 4 shows an accuracy variation of the proposed Random Forest in the learning process of the generations. Early generations seem low accuracy due to the incompleteness of the data sets. However, it is improved by the evolution of tree subset selectors to improve the classification accuracy through evolution. also, it seems the average, the maximum, and minimum accuracy of the one hundred individuals increase depending on the progress of the generation. The proposed approach was confirmed that improve the classification accuracy of the learning method through evolution in this experiment.

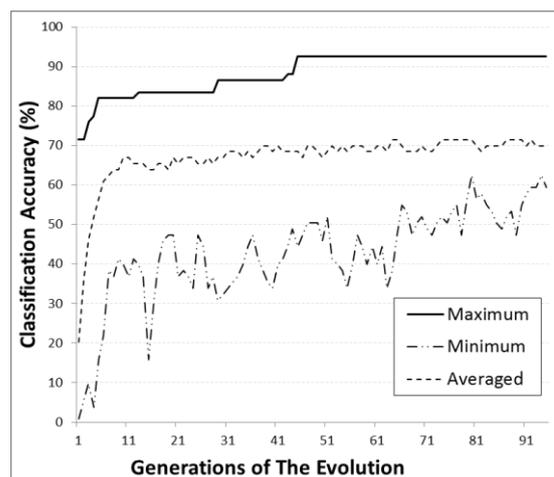


Figure 4. The experiment result of the proposed of the generations progress.

For comparing accuracy, we adopted the classification experiment to the typical classifiers such as support vector machine (SVM), multilayer perceptron (MLP), AdaBoost.M1, classification and regression trees (CART) of decision tree, and random forest. The parameters of the conventional classifiers were set at a typical value of WEKA [9]. The accuracy result in this paper was obtained by 10-fold cross validation.

Table 1 displays the classification results for conventional classifiers (implemented in WEKA) and our proposed method. By referring to the results of Table 1, we see that our classification system resulted in accuracy 7.03% higher than average of the conventions, and also 3.91% higher than SVM that is the best conventional classifier.

Table 1. The comparison experiment results with the conventional algorithms and the proposed evolutionary approach in the classification accuracy (%).

SVM	MLP	Adaboost M1	CART	Random Forest	Evolutionary Approach
88.57	86.71	83.72	82.82	85.43	92.48

4 Conclusions

The recognition of driving tendencies based on basic vehicle sensor data is a challenging task in the safety driving assistance system. This paper targets the improvement of driver recognition accuracy, and proposes a new learning method based on the evolutionary approach and random forest. Our proposed method was promoted the search capability by the evolution. Experimentally, we validated that the proposed mechanism is working well and the superiority in comparison to conventional classifiers in the selected dozens of sensor data in ECU by the relationship with the driver.

In future work, the improvement of the structure of select area operators is required. Adaptive system design has been considered to mitigate the sensitivity of the parameter settings. Also, it has devised a way to verify the operation model based on the risk in the application step. It plans to develop the other applications.

Acknowledge

This research was supported by X-Project Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1A2A1A16074929).

References

1. E. Infantes, M. E. Caspar, S. Kramer, S. Schaub, T. Langner, A. Eggers, and P. Lemmen, *The Effect of Pre-Crash Safety Systems to Occupant Protection in Offset Frontal Impacts*. SAE Technical Paper **2015-26-0164**, (2015)
2. K. D. Kusano, and H. C. Gabler, *Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions*. IEEE Transactions on Intelligent Transportation Systems, **13(4)**, 1546-1555, (2012)
3. A. Badea-Romero, F. J. Páez, A. Furones, J. M. Barrios, and J. L. de-Miguel, *Assessing the benefit of the brake assist system for pedestrian injury mitigation through real-world accident investigations*. Safety science **53**, 193-201, (2013)
4. B. F. Wu, H. Y. Huang, C. J. Chen, Y. H. Chen, C. W. Chang, and Y. L. Chen, *The blind spot warning system for daytime and nighttime.*, 2012 IEEE International Conference on In Industrial Technology (ICIT), pp. 591-597, (2012, March)
5. K. Mineta, K. Unoura, and T. Ikeda, *Development of a lane mark recognition system for a lane keeping assist system*. SAE Technical Paper **2003-01-0281**. (2003).
6. D. Yagil, *Interpersonal antecedents of drivers' aggression*, Transportation research part F: traffic psychology and behaviour **4(2)**, 119-131, (2001)
7. A. Liaw, and M. Wiener, *Classification and regression by randomForest*. R news **2(3)**, 18-22, (2002)
8. T. Bäck, and H. P. Schwefel, *An overview of evolutionary algorithms for parameter optimization*. Evolutionary computation **1(1)**, 1-23, (1993)
9. M. Hall, et al. *The WEKA data mining software: an update*. ACM SIGKDD explorations newsletter **11.1**, 10-18 (2009)