

Research Article

Hybrid Model for Early Onset Prediction of Driver Fatigue with Observable Cues

Mingheng Zhang,^{1,2} Gang Longhui,³ Zhe Wang,² Xiaoming Xu,²
Baozhen Yao,^{1,2} and Liping Zhou⁴

¹ State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China

² School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China

³ School of Navigation, Dalian Maritime University, Dalian 116026, China

⁴ Wuxi Mingda Traffic & Technology Consulted Co., Ltd., Wuxi 214125, China

Correspondence should be addressed to Baozhen Yao; yaobaozhen@yahoo.cn

Received 5 May 2014; Accepted 10 June 2014; Published 13 July 2014

Academic Editor: Rui Mu

Copyright © 2014 Mingheng Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a hybrid model for early onset prediction of driver fatigue, which is the major reason of severe traffic accidents. The proposed method divides the prediction problem into three stages, that is, SVM-based model for predicting the early onset driver fatigue state, GA-based model for optimizing the parameters in the SVM, and PCA-based model for reducing the dimensionality of the complex features datasets. The model and algorithm are illustrated with driving experiment data and comparison results also show that the hybrid method can generally provide a better performance for driver fatigue state prediction.

1. Introduction

Fatigue driving is a significant contributing factor to road related crashes worldwide, which will endanger lives of drivers and passengers and can cause serious accidents along major roads [1]. Studies show that 15–20% of car crashes may be attributable to driver sleepiness in high income countries [2]. If symptoms of fatigue could effectively warn the drivers and corrective measures could be taken, then disastrous outcomes would have been prevented. Researchers have made great efforts in this field. At present, the techniques for preventing driver's fatigue can be classified into the following categories.

- (1) Technologies to assess the vigilance capacity of an operator before the work is performed [3].
- (2) Mathematical models of dynamic alert [4].
- (3) Vehicle-based performance technologies that detect the behavior of the driver by monitoring the transportation hardware systems such as steering wheel movements and acceleration [5, 6].

- (4) Real-time technologies for monitoring the driver's state, including intrusive [7, 8] and nonintrusive monitoring systems [9–11].

The method assessing the vigilance before the work is performed investigates driver's subjective feeling of tiredness to determine whether the fatigue happens or not. This method is greatly influenced by individual differences, so it is only taken as a comprehensive qualitative evaluation method for driver fatigue analysis. The method based on vehicle performance uses visual sensor, speed sensor, and angle sensors to monitor vehicle running state such as speed, steering angle, or other features. The driver's spirit state can be inferred from this detected information. This method can make full use of the vehicle devices, so few additional equipment are required, but it is limited with the driver's driving habits, vehicle type, and road conditions. Experiments have demonstrated that the driver in fatigue would exhibit some visual cues [12]. Therefore, at present, the method based on machine vision has been a leading technique due to its hardware's characteristics such as small size, low cost, and nonintrusiveness to drivers.

Various visual parameters, such as eye blinking [9], head movement [13], face orientation [10], and other features, can be extracted when the driver's fatigue occurs. But, it is inaccurate when prewarning is triggered according to a certain specific aspect, because the driver's fatigue is not directly observable but can only be inferred from the information available. There are numbers of reasons for adopting the information fusing method to resolve the driver fatigue detection issues. Firstly, the driver's fatigue derived from the contextual features contains much subjectivity that cannot always reflect the real objectivity; secondly, inferring the driver's fatigue from his/her facial expression is not always reliable because of the following two limitations: (a) the current techniques for image processing cannot always ensure the recognition accuracy; (b) an introverted person might have a tendency to control his or her displays of emotions, which leads to an inaccurate interpretation of the facial expression. Thus, to fuse as many as possible features from uncertain events is necessary to make an accurate inference for monitoring fatigue effectively.

Since there are various nonlinear, stochastic, time-varying factors during the driver fatigue prediction, it is very difficult to predict the fatigue state accurately. Recently, SVM, a novel machine learning algorithm, has been proved, that is, a promising tool for both data classification and regression [14]. It shows high resistance to the over fitting problem, achieving high generalization performance in solving various time series forecasting problems, which has been applied in prediction of time series [15]. These successful applications motivate us to apply SVM in the driver fatigue state prediction.

The parameters selection in SVM, which greatly impacts the performance of SVMs, need to be optimized and set by users. Many literatures have proved that the parameters in SVM play an important role in its performance [16]. Improper selection of the parameters could cause either the overfitting or underfitting of the training data points. These parameters mainly include the penalty factor C and the parameters of kernel function (for instance, parameter γ of RBF kernel function). At present, grid search algorithm [17] is the most reliable method for the offline data training process. However, for large scale or real-time feature practice application, the considerable searching time cannot be accepted. Many literatures suggested that heuristic algorithms were good choice to solve this kind of problems [18–28]. Selakov et al. [29] propose a model for short term electrical load forecasting based on particle swarm optimization (PSO) and SVM. The obtained results show better accuracy compared to results generated with classical methods. A discrete gravitational search algorithm is combined with SVM which is presented for improving classification accuracy with an appropriate feature subset in binary problems [30]. Hou and Li [16] presented evolution strategy with covariance matrix adaptation to identify the parameters in SVM.

The purpose of this paper is to build a hybrid model for driver fatigue state prediction, which is named SVM-GA model. Within the proposed model, several cues are considered including PERCLOS AECS and PNS. The structure of this paper is organized as follows: the problem description

and the architecture of the solution for the driver fatigue state prediction are given out in Section 2. In Section 3, experiments and results are discussed; and lastly, the conclusions are given together with suggestions for further study.

2. Problem Description

As discussed above, fusing as many as possible features from uncertain events is necessary to make an accurate inference for predicting fatigue effectively. According to this problem, fatigue state is obviously the target hypothesis variable that we intend to infer while other visual cues, which are symptoms of fatigue, are information variables. The main purpose of a PSO-SVM model is to predict the unobserved events from the observed or contextual data. Due to there are so many factors reflecting human fatigue as discussed, it is impossible to include all of them into a SVM model. Hence, only the most significant ones are incorporated.

2.1. Observable Cues Analysis. As mentioned above, the driver fatigue can be inferred from the facial expression or observable variables. Various visual cues/parameters can be extracted when the driver's fatigue occurs, as discussed in Section 1. In this research field, the challenge is that we cannot infer the fatigue level through a certain specific aspect for a number of reasons. Therefore, fusing as many information that we can get may be the only way out to resolve the driver fatigue detection issues. But, unfortunately, it is impossible to include all of the visual parameters in the fatigue model. Hence, in this paper, only the following factors that play more contributions to the fatigue are incorporated.

2.1.1. PERCLOS Analysis. PERCLOS is the percentage of eyelid closure over the pupil over time and reflects slow eyelid closures (“droops”) rather than blinks. A PERCLOS drowsiness metric was established in a 1994 driving simulator study as the proportion of time in a minute that the eyes are at least 80 percent closed [31]. Based on research by Fairbanks and colleagues, the federal highway administration (FHWA) and National Highway Traffic Safety Administration (NHTSA) consider PERCLOS to be among the most promising known real-time measures of alertness for in-vehicle drowsiness-detection systems [32].

2.1.2. AECS Analysis. AECS refers to the average eye closure speed, which is usually described via the time elapse from eyelids fully open to close completely. Research shows that the closure speed has obvious difference between the driver in awake and fatigue state [33], and this difference can reach more than eight times. In addition, there exists a higher correlation between AECS and PERCLOS. Therefore, the AECS may be a good visual observable variable for the driver fatigue model design.

2.1.3. PERLVO Analysis. PERLVO refers to percentage of mouth large vertical open over time. Research shows that this variable is obviously different when a driver's yawning fatigue or distraction state due to talking or conversation occurred [34]. Yawn may not occur with everyone when he/she falls

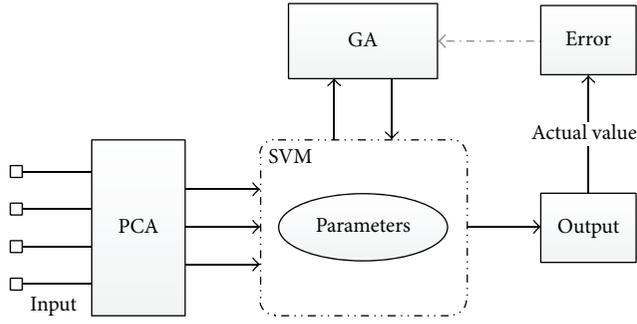


FIGURE 1: Flow-chart of the hybrid model.

into fatigue, but the PERLVO may be necessary supplement for the driver fatigue model design.

2.1.4. PERTITL Analysis. PERTITL refers to percentage of head tilt over time, that is, the percentage that the time of driver's looking up or down accounts for a certain time. It is used to describe the head movement of a driver when he/she falls into fatigue.

2.1.5. PNS Analysis. PNS refers to the percentage of non-steering, that is, the level of the steering wheel without any movement during a period of time. Research result show that, relative to the wake state, the PNS value is significantly increased when the driver is in the fatigue or sleep state. But, unfortunately, it is very difficult to distinguish the state between the fatigue and the sleep [35].

2.1.6. SDLP Analysis. SDLP refers to the standard deviation of lane position, that is, the level of a vehicle transverse the lane mark during a period of time. When the fatigue occurs, before the accident happens, vehicle sometimes could cross the lane. Therefore, researchers worldwide have been developing vision-based systems for lane keeping on unintended lane departures [5]. In these researches, two warning criteria are usually considered: the lateral offsets and the time to lane-crossing (TLC). Therefore, the SDLP variables are useful cues for reflecting the driver's fatigue. It should be considered into the driver fatigue model design.

2.2. Architecture of the Solution. Based on the above analysis, the architecture of the solution can be proposed. The proposed hybrid model consists of three modules: PCA unit, SVM unit, and GA unit. The hybrid model can be described as in Figure 1. In subsequent sections, three submodels will be discussed, respectively.

2.2.1. Support Vector Regression. As a new and promising technique for classification and regression problems, SVM can be adjusted to map the complex input-output relationship for the nonlinear system without being dependent on the specific functions. Unlike other nonlinear optimization methods, the solution of SVM always can achieve the global optimal solution without limitation to a local minimum point and it shows the strong resistance to the overfitting problem and the high generalization performance. Support vector

regression is a common term for implementation of SVM in regression problems.

Given the training data set $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$ ($x_i \in X \in R^n, y_i \in Y \in R, x_i \in X \subseteq R^n, y_i \in Y \subseteq R$), n is the number of samples. $\{x_i, y_i\}$ are pair of input and output vectors. SVM make use of a nonlinear mapping function to map x into a high-dimensional feature space H in which linear approximation is conducted to find the mapping function so that we can get a better approximation for the given data samples. Based on the statistical learning theory, we can get the function as follows: $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$

$$f(x) = \omega\phi(x) + b, \quad (1)$$

where ω is weight factor and b is the threshold value. The input samples are mapped to higher dimensional space by using kernel function ϕ . The kernel function is responsible for nonlinear mapping between input and feature space. The previous researches [36, 37] suggested that radial basis function (RBF) kernel had less numerical difficulties with application and was efficient for traffic state prediction. Thus, RBF kernel function is used for the SVM model in this study.

Regression can be defined as a problem that minimized the risks for a loss function. The optimal regression function is the minimum and regularized generic function under certain constraints:

$$\begin{aligned} \min_{w, b, \xi, \xi^*} \quad & \frac{1}{2}w^T w + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \\ \text{subject to} \quad & w^T \phi(x_i) + b - z_i \leq \epsilon + \xi_i \\ & z_i - w^T \phi(x_i) - b \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (2)$$

The dual is

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2}(\alpha - \alpha^*)^T Q (\alpha - \alpha^*) \\ & + \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l z_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ & 0 \leq \alpha_i, \quad \alpha_i^* \leq C, \quad i = 1, \dots, l, \end{aligned} \quad (3)$$

where $Q_{ij} = K(x_i, x_j) \equiv \theta(x_i)^T \theta(x_j)$.

The approximate function is

$$f(x) = \sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(x_i, x_j) + b, \quad (4)$$

where $K(x_i, x_j) = \phi(x_i)$ denotes the inner product of the vector in the feature space. All of the kernel functions can be directly computed in the input space.

2.2.2. GA for Parameters Optimization. Although SVM is feasible and applicable in predicting the driver fatigue state, there are some parameters, which greatly impact the performance of SVMs, need to be optimized. In general, for the RBF kernel, as a nonlinearly kernel function, the parameters such as C , γ , and ε are the key elements and they directly decide about the prediction performance of SVM. Thus, the parameter optimization is important for improving the prediction accuracy. In this paper, genetic algorithm is used for searching the best parameters for the presented SVM model. In general, the process of GA can be briefly described as follows.

(1) *Encoding of Chromosome.* In GA, a standard representation of each candidate solution is as a chromosome that is composed of “genes.” For the SVM parameters optimization problem in this paper, the real encodings were adopted since the parameters C , γ , and ε are continuous-valued. Each chromosome consists of gene_1^g , gene_2^g , and gene_3^g , which represent the three parameters, respectively. Here g is the current generation. In order to reduce the search space, the previous literature has given out the recommended searching space which, respectively, attribute to the range $C \in [2^{-5}, 2^5]$, $\varepsilon \in [2^{-13}, 2^{-1}]$, and $\gamma \in [0, 2]$.

(2) *Fitness Function.* A fitness function is a particular type of objective function that is used to summarize how close the possible solutions is to achieving the set aims. For the SVM parameters optimization problem in this paper, considering that GA is always finding the maximum fitness of the individual chromosome, mean squared error (MSE) is adopted as follows:

$$\text{fitness} = \frac{1}{l} \sum_{i=1}^l (f(x_i) - y_i)^2, \quad (5)$$

where $f(x_i)$ is the prediction value by the SVM model; y_i is the observed value; l is the number of observation variables.

(3) *Selection Operation.* During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. In this paper, the roulette selection strategy is adopted. Based on the fitness calculation results, the sum fitness value of the entire population and then the ratio corresponding to each chromosome are obtained. In the next step, a random number (range from 0 to 1) is used for determining the range of the cumulative probability. The chromosome falling within the expected range is selected out.

(4) *Genetic Operators.* For each new solution to be produced, a pair of “parent” solutions is selected for breeding from the pool selected previously. A second generation population of solutions is generated from those selected through a combination of genetic operators: crossover (also called recombination), and mutation.

Crossover is a genetic operator used to vary the programming of a chromosome or chromosomes from one generation to the next. It is analogous to reproduction and biological crossover. Cross over is a process of taking more than one parent solutions and producing a child solution from them. In literature [38], an arithmetic crossover operator is used. Consider

$$\begin{aligned} \text{gen}_{k,I}^t &= \alpha_i \text{gen}_{k,I}^{t-1} + (1 - \alpha_i) \text{gen}_{k,II}^{t-1} \\ \text{gen}_{k,II}^t &= \alpha_i \text{gen}_{k,II}^{t-1} + (1 - \alpha_i) \text{gen}_{k,I}^{t-1}, \end{aligned} \quad (6)$$

where gen_k^{t-1} is the “parent” chromosomes; gen_k^t is the “child” chromosomes; α_i is a random probability value with range $(0, 1)$.

Mutation is a genetic operator used to maintain genetic diversity from one generation chromosome to the next. Mutation occurs during evolution according to a user-definable mutation probability. A very small mutation rate may lead to premature convergence of the genetic algorithm and a very high rate may lead to loss of good solutions unless there is elitist selection. In general, the mutation rate is defined with the range $[0.001, 0.1]$. In this paper, according to the previous literature, the mutation rate is set to 0.05.

(5) *Termination.* This generational process is repeated until a termination condition has been reached. In this paper, the search loop continues until $\text{MSE}_n - \text{MSE}_{n-1} < 0.0001$ or the number of generation reaches the maximum number of generations T_{\max} .

2.2.3. PCA for Reducing the Dimensionality of Input Datasets. In order to speed the velocity of SVM training and prediction with maintaining the main information of samples and not changing the distribution of samples, in this paper, PCA method is used before the observable variables are input into the SVM model for training and prediction.

In PCA, which kind of data is used will has a great influence for the result of the analysis. Existing literature shows that the analysis results are different when the covariance matrix and the correlation matrix are used for PCA, respectively. In this paper, the various observable variables can reflect the driver fatigue state from different aspects. Thus, there exists some relevance between the variables. Meanwhile, the dimensional of each variable is different. Therefore, based on the above analysis, we consider using the correlation matrix for the subsequent PCA process.

(1) *Calculate the Correlation Matrix.* Given the original data matrix $X = (x_{ij})_{n \times p}$, n is the number of the observable samples and p is the dimensions for each sample. Then the correlation matrix can be calculated with the following formula:

$$r_{x_i, x_j} = \frac{\text{cov}(X_i, X_j)}{\sqrt{D_{x_i} D_{x_j}}}, \quad (7)$$

where, $\text{cov}(X_i, X_j)$, D_{x_i} and D_{x_j} are the covariance and variance of the original data, respectively.

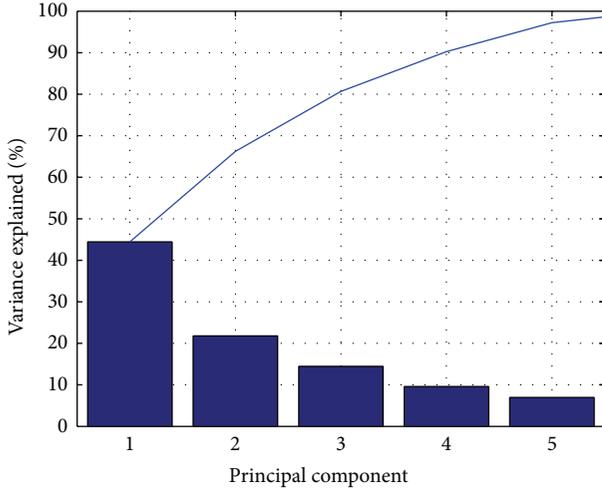


FIGURE 2: Distribution curve for the cumulative energy of PCA with original data.

In this paper, the original data matrix X is composed with the observable variables such as PERCLOS and AECS. Thus, the dimensions of samples are 6 types of the observable variables.

(2) *Calculate the Eigenvectors and Eigenvalues.* Compute the matrix V of eigenvectors which diagonalize the correlation matrix R

$$V^{-1}RV = D, \quad (8)$$

where D is the diagonal matrix of eigenvalues of R . Matrix D will take the form of an $p \times p$ diagonal matrix, where $D(k, l) = \lambda_k$ for $k = l$ is the k th eigenvalue of the correlation matrix R . Matrix V , also of dimension $p \times p$, contains p column vectors, each of length p , which represent the p eigenvectors of matrix R .

(3) *Choosing Components and Deriving the New Data Set.* Once the eigenvectors and eigenvalues of matrix R obtained, the eigenvalues and eigenvectors are sorted in order to decrease eigenvalue. The eigenvalues represent the distribution of the source data's energy among each of the eigenvectors. The cumulative energy content g for the j th eigenvector is the sum of the energy content across all of the eigenvalues from 1 through j . We can use the vector g as a guide in choosing an appropriate value. In this paper, the threshold value for cumulative energy is given as 0.9. Figure 2 is the distribution curve for the cumulative energy of PCA with original data. So the following PCA analysis result can be drawn: with the given threshold value 0.9, the main principal component number is 4. With this operator, the observable variables dimension is changed from 6 to 4.

Based on the above analysis, the new data set can be derived with the following formula:

$$X = V \times R, \quad (9)$$

where X is the new feature dataset matrix which will be input into the GA-SVM model including 4 main principal components with 0.9 cumulative energy loadings.

3. Case Study

In this section, an experiment has been carried out to verify the presented model. Based on the above analysis in Section 2.1, the observable variables used are divided into three types: the eye movement parameters are obtained by eye tracker; the face visual movement parameters are obtained by high-definition camera; and the vehicle running state parameters are collected by a driver simulator. In addition, the parameters obtained from physiological parameter tester are used as the standard that verifies whether the driver is fatigue or not.

A driver simulator equipped with data acquisition sensors is shown as Figure 3. The simulator is equipped with visual and auditory function for real-time simulation. In which, the visual system produces a 3D road scene that can be used for simulating the vehicle running state and the visual effect under the driver's operation such as kinds of traffic flow, road lines, different weather, and other traffic things on the road. The auditory system can simulate the sound of different things perfectly, including the engine, the horn, the environment, and the wind. The tester manipulates the steering wheel and sends command to the simulator. Then, various sensors installed on the simulator can interpret this command and give out the parameters such as the throttle opening and the steering wheel angle. The sampling frequency of the simulator we adopt is 50 Hz.

The experiments lasted from 1:00 P.M to 3:00, sampling at every one minutes, during which each participant was asked to operate the driving simulator at the speed of 90 km/h in highway environment, and his/her dynamic facial image was obtained at the sampling rate of 100 fps. Signals were processed with the corresponding methods to form the observable datasets. The experimental cases and simulation scenario were shown as Figure 4.

3.1. Parameter Identification. In the experiment, there are 100 samples in total. The data is divided into two subsets, which represent training samples and testing samples, respectively. Firstly, about 20% of samples data were set as testing data. Then, 80% of the remaining samples data were assigned to training.

To properly optimize the three parameters C , γ , and ϵ for SVM, GA is used. Before the implementation, four GA parameters, namely, p_c , p_m , p_{size} , and T_{max} , need to be predetermined. In general, p_c varies from 0.3 to 0.9; p_m varies from 0.01 to 0.1; p_{size} is the population size which is set according to the size of the samples. T_{max} is the maximum number of generations which can be determined according to a good convergence of the calculation. Considering the features of this problem and our experiences, the characteristic of GA can be acquired, as can be seen in Table 1. Then GA continues running for searching the best result for parameters C , γ , and ϵ . Figure 5 shows the MSE curve for parameters optimal with given conditions. It can be observed that the optimal fitness

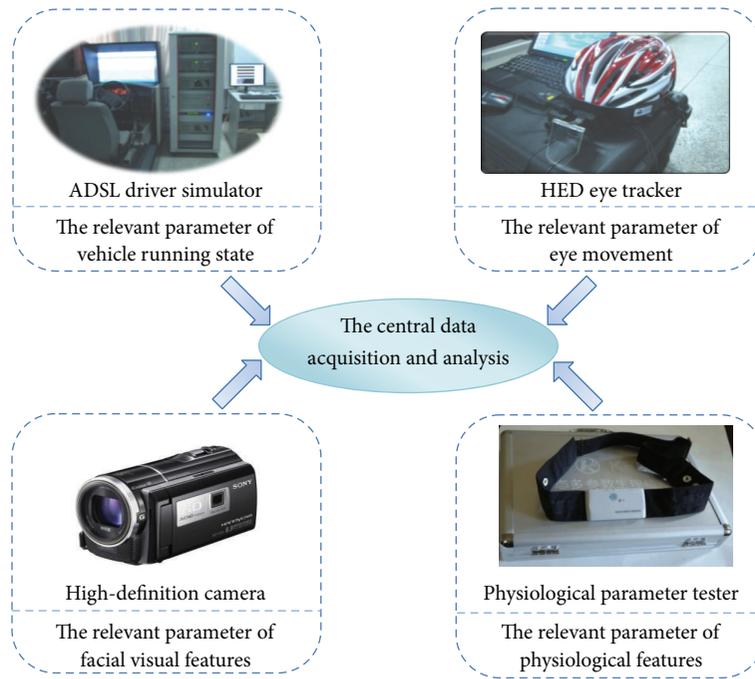


FIGURE 3: The hardware system for driver fatigue experiment.



(a) Participants are driving with simulator



(b) The experiment scenario

FIGURE 4: Experiment cases.

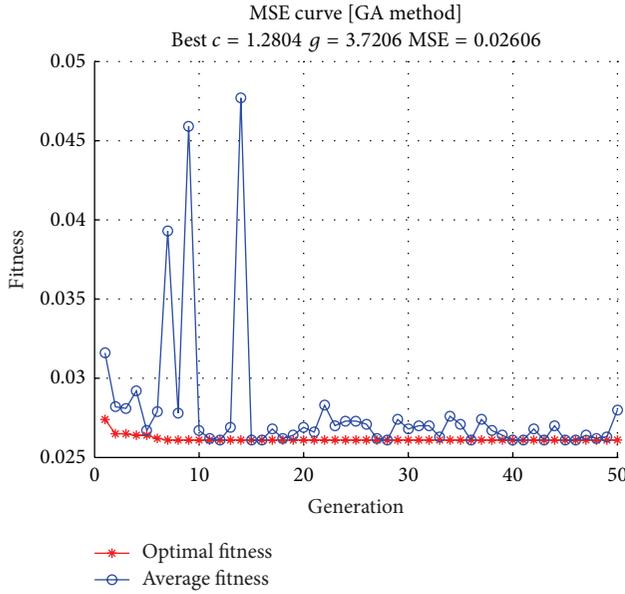


FIGURE 5: MSE curve for SVM parameter searching.

TABLE 1: The characteristics of GA.

Parameter	P_c	P_m	P_{size}	T_{max}
Value	0.6	0.05	5	500

decreases fast before the 5th generation, and then it changes smoothly. The least prediction error appears in about the 45th generation, and it almost remains unchanged, and the process loop terminates in the 50th generation. This means that GA has a good convergence for the SVM parameters searching problem. Finally, the three parameters C , γ , and ϵ were optimized as (1.2804, 3.7206, and 0.0018) with the best MSE value 0.026 for the practical prediction model of the driver fatigue state prediction SVM-GA.

3.2. Results and Analysis. Based on the parameters selected, the final SVM-GA hybrid model for driver fatigue state prediction is confirmed. In order to verify the actual performance of the prediction model presented above, tests were conducted on the train set and test set; respectively, the result was shown as Figure 6. In this test, 20% of the samples data was used for predicting the driver fatigue state. From the result, it is obviously indicated that the prediction curves agree well with the actual observing data. In addition, the analysis result of prediction error is shown that the mean relative error (MSE) is 1.9% with the test set samples. Thus, it can be seen that the presented SVM-GA model is feasible and applicable for driver fatigue state prediction.

In order to evaluate the impact performance with PCA process and different parameters searching method, tests were conducted on the sample data, respectively. The test results are shown as Figures 7-8 and details information of time consumption and MSE are listed in Table 2.

According to the comparison result with/without PCA process in Figure 7, it is indicated that the preprocesses of

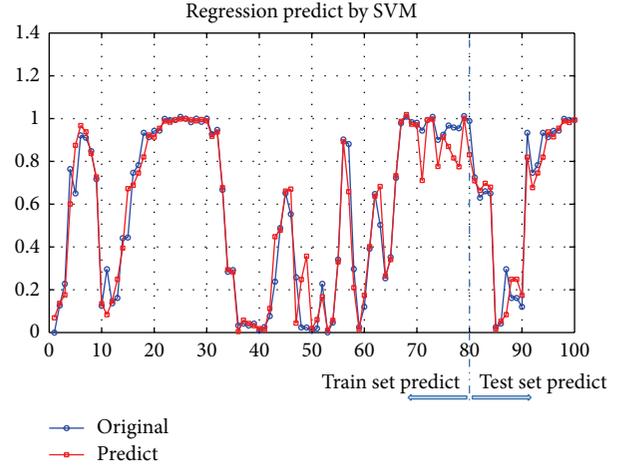


FIGURE 6: Prediction result for SVM-GA model.

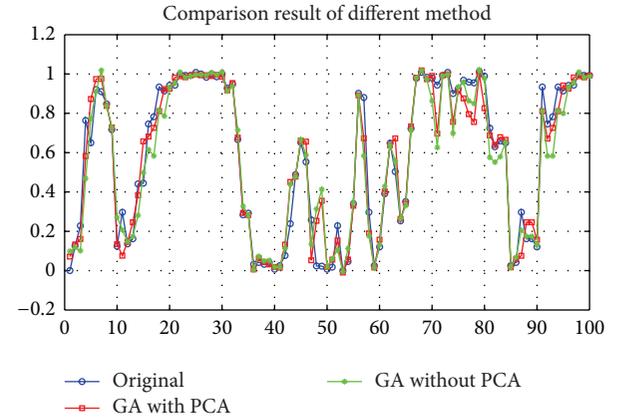


FIGURE 7: Comparison test result with/without PCA process.

TABLE 2: Comparison result for the different methods.

Method	Time elapse (s)	MSE (Test set)	MSE (Train set)	Best MSE
GA with PCA	2.34	0.0094	0.019	0.0284
GA without PCA	2.65	0.0055	0.01	0.0259
CG with PCA	4.62	0.0075	0.0104	0.0281
CG without PCA	5.54	0.0053	0.0094	0.0259

PCA for original data dimension reduction can play a role in improving the prediction result to some extent, that is, in Table 2, the MSE (on test set) for the two methods are 0.0094 and 0.0055, respectively. But, on the contrary, the time consumption for data training and prediction is 2.34 s and 2.65 s. It can be explained in such a way that the PCA process, in fact, is a statistical procedure that uses orthogonal transformation to convert a set of sample data with possibly correlated variables into a set of values of linearly uncorrelated variables. During the transformation, the principal components are extracted for replacing the original sample data and inputting into the SVM model for training and prediction. So there is

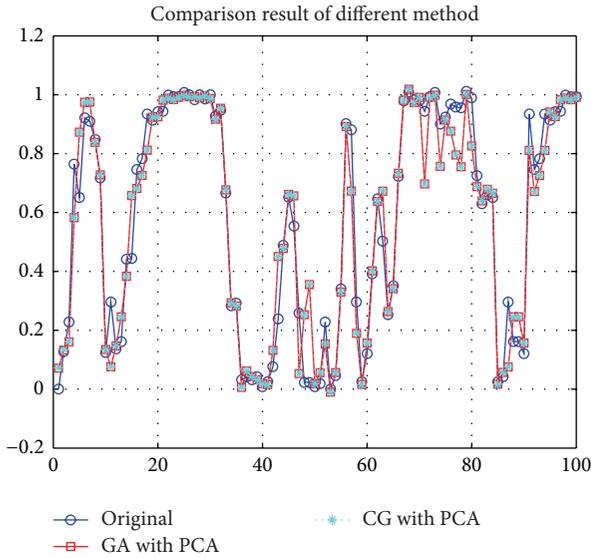


FIGURE 8: Comparison result with GA and CG.

some information loss relative to the original data. This procedure can reduce the data dimension and reveal the internal structure of the data in a way that best explains the variance. But, on the contrary, the loss information may decrease the prediction accuracy in the SVM model applying period. Thus, in practical application, there is a choice must be made between the computation speed and the prediction accuracy.

In order to evaluate the parameters optimization result through GA method proposed in this paper, a cross verification method test, often used for the SVM parameter searching, is conducted. From the result in Figure 8 and Table 2, it is obvious that the results are approximate with the two methods. But the time consumption is significantly different for the SVM model's parameter optimization. There is a great of advantages in SVM-GA method for improving calculation speed.

4. Conclusions

State prediction is an important phase for the driver fatigue early onset prewarning. The state can be inferred from the detected cues such as PERCLOS, AECS, and PNS. Due to the complexity of the information obtained, this paper attempts to develop a hybrid model based on SVM and GA to predict the driver fatigue state during the early onset phases. To evaluate the performance of the proposed method, an experiment with driving experiment data is carried out. The results show that GA has a good convergence and relative stable performance. Furthermore, the comparison of results between different methods suggests that the SVM-GA provides lower prediction errors and time consumption than the other approaches. This indicates that SVM-GA seems to be a powerful tool for driver fatigue state prediction during the early onset phases.

In summary, it is believed that a hybrid model proposed in this paper can provide a better performance for early

onset prediction of driver fatigue. However, there are some unresolved issues to be discussed in the future work. First of all, the driver fatigue has an accumulative property and fatigue is developed over time. Future work is needed to consider this property and extend the current model to be accordant with the practical situation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was jointly supported by grants from the Humanities and Social Sciences Foundation of the Ministry of Education in China (Project no. 12YJCZH280), the Ph.D. Programs Foundation of Ministry of Education of China (Project no. 20112125120004), the National Natural Science Foundation of China (Project nos. 51208079 and 11272075), and the Fundamental Research Funds for the Central Universities (Project no. 3013-852019).

References

- [1] S. K. L. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen, "Development of an algorithm for an EEG-based driver fatigue countermeasure," *Journal of Safety Research*, vol. 34, no. 3, pp. 321–328, 2003.
- [2] J. L. Connor, "The role of driver sleepiness in car crashes: a review of the epidemiological evidence," in *Drugs, Driving and Traffic Safety*, J. Verster, S. R. Pandi-Perumal, J. Ramaekers, and J. Gier, Eds., pp. 187–205, Birkhäuser, Basel, Switzerland, 2009.
- [3] H. Y. Shu, F. Q. Li, S. P. Yi, H. Fujimoto, and Y. Ueno, "Comprehensive evaluation method for automotive driving fatigue," *Journal of Shanghai Jiaotong University*, vol. 42, no. 8, pp. 1338–1343, 2008.
- [4] D. Dawson, N. Lamond, K. Donkin, and K. Reid, "Quantitative similarity between the cognitive psychomotor performance decrement associated with sustained wakefulness and alcohol intoxication," in *Managing Fatigue in Transportation*, pp. 231–256, 1998.
- [5] J.-G. Wang, C.-J. Lin, and S.-M. Chen, "Applying fuzzy method to vision-based lane detection and departure warning system," *Expert Systems with Applications*, vol. 37, no. 1, pp. 113–126, 2010.
- [6] C. Wu, C. Lin, and C. Lee, "Applying a functional neurofuzzy network to real-time lane detection and front-vehicle distance measurement," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 4, pp. 577–589, 2012.
- [7] M. V. M. Yeo, X. P. Li, K. Q. Shen, and E. P. V. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?" *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.
- [8] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by Support Vector Machine," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7651–7658, 2009.
- [9] A. Picot, S. Charbonnier, and A. Caplier, "Drowsiness detection based on visual signs: Blinking analysis based on high frame rate video," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC '10)*, pp. 801–804, Austin, Tex, USA, May 2010.

- [10] M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith, "3D face reconstructions from photometric stereo using near infrared and visible light," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 942–951, 2010.
- [11] S. Vitabile, A. de Paola, and F. Sorbello, "A real-time non-intrusive FPGA-based drowsiness detection system," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 4, pp. 251–262, 2011.
- [12] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, 2004.
- [13] A. Doshi and M. Trivedi, "A comparative exploration of eye gaze and head motion cues for lane change intent prediction," in *Proceedings of the 2008 IEEE Intelligent Vehicles Symposium*, pp. 49–54, June 2008.
- [14] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [15] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003.
- [16] S. Hou and Y. Li, "Short-term fault prediction based on support vector machines with parameter optimization by evolution strategy," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12383–12391, 2009.
- [17] C. W. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification*, Technical Report, National Taiwan University, 2003.
- [18] B. Z. Yao, P. Hu, X. H. Lu, J. J. Gao, and M. H. Zhang, "Transit network design based on travel time reliability," *Transportation Research C*, vol. 43, pp. 233–248, 2014.
- [19] B. Z. Yao, P. Hu, M. H. Zhang, and S. Wang, "Artificial bee colony algorithm with scanning strategy for the periodic vehicle routing problem," *Simulation*, vol. 89, no. 6, pp. 762–770, 2013.
- [20] B. Yao, C. Yang, J. Yao, and J. Sun, "Tunnel surrounding rock displacement prediction using support vector machine," *International Journal of Computational Intelligence Systems*, vol. 3, no. 6, pp. 843–852, 2010.
- [21] B. Yao, C. Yang, J. Hu, J. Yao, and J. Sun, "An improved ant colony optimization for flexible job shop scheduling problems," *Advanced Science Letters*, vol. 4, no. 6-7, pp. 2127–2131, 2011.
- [22] B. Z. Yao, P. Hu, M. H. Zhang, and X. M. Tian, "Improved ant colony optimization for seafood product delivery routing problem," *ROMET-TRAFFIC & Transportation*, vol. 26, no. 1, pp. 1–10, 2014.
- [23] B. Z. Yao, J. B. Yao, M. H. Zhang, and L. Yu, "Improved support vector machine regression in multi-step-ahead prediction for tunnel surrounding rock displacement," *Scientia Iranica*, 2013.
- [24] B. Yu, Z. Yang, K. Chen, and B. Yu, "Hybrid model for prediction of bus arrival times at next station," *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 193–204, 2010.
- [25] Y. Bin, Y. Zhongzhen, and Y. Baozhen, "Bus arrival time prediction using support vector machines," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 10, no. 4, pp. 151–158, 2006.
- [26] B. Yu, Z. Z. Yang, and B. Z. Yao, "A hybrid algorithm for vehicle routing problem with time windows," *Expert Systems with Applications*, vol. 38, no. 1, pp. 435–441, 2011.
- [27] B. Yu, Z. Yang, and J. Yao, "Genetic algorithm for bus frequency optimization," *Journal of Transportation Engineering*, vol. 136, no. 6, pp. 576–583, 2010.
- [28] B. Yu, H. B. Zhu, W. J. Cai, N. Ma, and B. Z. Yao, "Two-phase optimization approach to transit hub location—the case of Dalian," *Journal of Transport Geography*, vol. 33, pp. 62–71, 2013.
- [29] A. Selakov, D. Cvijetinović, L. Milović, S. Mellon, and D. Bekut, "Hybrid PSOSVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank," *Applied Soft Computing*, vol. 16, pp. 80–88, 2014.
- [30] S. Sarafrazi and H. Nezamabadi-pour, "Facing the classification of binary problems with a GSA-SVM hybrid system," *Mathematical and Computer Modelling*, vol. 57, no. 1-2, pp. 270–278, 2013.
- [31] R. J. Fairbanks, S. E. Fahey, and W. W. Wierwille, "Research on vehicle-based driver status/performance monitoring," Seventh Semi-Annual Research Report, U.S. Department of Transportation, National Highway Traffic Safety Administration, 1994.
- [32] D. F. Dinges and R. Grace, "Perclous: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance," 1988.
- [33] Q. Ji and X. J. Yang, "Real time visual cues extraction for monitoring driver vigilance," in *Computer Vision Systems*, B. Schiele and G. Sagerer, Eds., pp. 107–124, Springer, Berlin, Germany, 2001.
- [34] W. Rongben, G. Lie, T. Bingliang, and J. Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 314–319, October 2004.
- [35] X. B. Zhang, B. Cheng, and R. J. Feng, "Real-time detection of driver drowsiness based on steering performance," *Journal of Tsinghua University*, vol. 50, pp. 1072–1076, 1081, 2010.
- [36] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, "Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 1156–1166, 2012.
- [37] Q. Li, "Short-time traffic flow volume prediction based on support vector machine with time-dependent structure," in *Proceedings of the 2009 IEEE Instrumentation and Measurement Technology Conference (I2MTC '09)*, pp. 1730–1733, May 2009.
- [38] B. Yu, Z. Yang, and C. Cheng, "Optimizing the distribution of shopping centers with parallel genetic algorithm," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 2, pp. 215–223, 2007.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

