Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 2, February 2018, pp. 280-289
DOI: 10.23940/ijpe.18.02.p9.280289

Relief Feature Selection and Parameter Optimization for Support Vector Machine based on Mixed Kernel Function

Wei Zhang[a,b,*], Junjie Chen[b]

[a]*Information Center, Shanxi Medical College for Continuing Education, Taiyuan, 030012, China*
[b]*College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, China*

Abstract

In order to improve the classification performance of Support Vector Machine (SVM), Relief feature selection algorithm was used to obtain the most relevant feature subset and remove redundant features. The mixed kernel function, which combined the global kernel function with the local kernel function, was proposed to strengthen the learning ability and generalization performance of SVM. In addition, the parameter optimization of SVM, which combined Genetic Algorithm (GA) with grid search, was performed to reduce computation time and find optimal solutions. Finally, the methods presented in this paper were used in the Heart disease data set and the Breast cancer data set in the UCI. Compared with KNN and BP neural network, the classification result of SVM model with Relief algorithm and mixed kernel function significantly outperformed the other comparable classification model and the experimental results demonstrate the validity of the proposed model.

Keywords: relief; mixed kernel function; support vector machine; parameter optimization

(Submitted on November 13, 2017; Revised on December 19, 2017; Accepted on January 8, 2018)

1. Introduction

Feature selection refers to selecting an optimal feature subset from an input feature set according to certain criterion and it is widely used in many fields such as machine learning [15], pattern recognition [3,15], bioinformatics [18], text categorization [21], signal processing [13] and so on. To select beneficial feature reasonably and effectively and reduce the dimensions of feature not only eliminates the redundancy and speeds up computational speed, but also improves the efficiency of classification and reduces classification error. Due to the above advantages, feature selection has been an important research direction in the field of machine learning, which raises the interest of more and more scholars in this field.

According to the relation between feature selection and learning algorithms, the methods of feature selection can be divided into Embedded Method [23], Wrapper Method [5,25] and Filter Method [1,9]. The process of feature selection of Embedded Method is embedded into model training of learning algorithm; it selects in a local space so the effect is more limited. The biggest feature of Wrapper Method is concerned with learning algorithm; the training data set needs to be divided into training subset and testing subset. While it has the advantage of having stronger learning ability through the feature subset constructed by the method, Wrapper Method is time consuming and has poor generalization ability. Feature selection of Filter Method is independent of learning algorithm and is obtained by the original feature set. The method can be combined with any training methods, and has high flexibility and high operation efficiency. Therefore, the study selects to use Relief feature selection of Filter Method [22,29].

Support Vector Machine (SVM) [26] has a complete theory system and better performance, and is widely used in many fields [4,6,17]. The two most important factors to influence the classification performance of SVM are [8,14]: (1) the choice of kernel function and the parameter of g. (2) the penalty parameter C. In addition, the original data set has also

* Corresponding author.
E-mail address: zhangwhhx@163.com

important influence on classification results. Thereby, data preprocessing, which includes de-noising and feature selection, is needed.

The present common methods of the parameter optimization of SVM include grid search method [16,24], Genetic Algorithm(GA) [20], Particle Swarm Algorithm (PSO) [10,27], ant colony algorithm [30,31] and so on. Grid search method is to divide the parameters that will be searched into grid within a certain scope and traverse all points of grid for optimal value. Therefore, this method is can easily find the global optimal solutions. Its disadvantage is that it is time consuming since all points need to traverse. To overcome this disadvantage, researchers have developed intelligent heuristic methods such as GA, PSO.

The heuristic methods set the relevant heuristic rules corresponding to practical problems, which refer to telling us how to search the rules of answers. Compared with blind grid search method, the heuristic methods are obviously more efficient. Therefore, the heuristic methods are suitable for high real-time requirements. However, the heuristic algorithms depend on practical problems, experience and skill level of experimenters. GA is a classical intelligent heuristic method that does not need to experiment with a lot of values, but it can also find values of optimal parameters. On the other hand, the disadvantage of GA is that it easily falls into local optimization and is more complex when making rules.

To overcome the disadvantage of grid search and GA, the study tried to combine grid search with GA. Firstly, GA was used to search relatively optimal solutions of large range. Then, grid search was employed to perform a second precise search to determine the final optimal solutions. By the combinational methods, computation time can be reduced and optimal solutions with high precision can also be found.

SVM kernel function can be divided into local kernel function and global kernel function. The local kernel function has strong learning capacity and weak generalization ability; on the contrary, the global kernel function has strong generalization capacity and weak learning ability. To further improve learning capacity and generalization ability, the study will combine the advantages of the two kernel functions to construct the hybrid kernel function, which can promote the learning capability and generalization capability [7,12,28].

Firstly, the study adopted Relief feature selection to compute the contributing weights to classification. Then, key features with greater influential factors were elected and redundant features were removed. Secondly, the study used nonlinear computational capabilities of constructed combined kernel function to approximate the function relation implied by the trained data and implemented the construction of classification model. Finally, we conducted experiments and verified our methods.

2. Relief Feature Selection Algorithm

Relief is a kind of famous filter feature selection algorithm which designs a “relevant statistics” to measure the important degree of the feature. This statistic is a vector whose every component corresponds to an initial feature respectively. Moreover, the importance of feature subset depends on the component sum of relevant statistics which every feature of the subset matches. At last, the top k features of bigger component of the relevant statistics were selected. Therefore, the key of Relief algorithm is to determine the relevant statistics.

$D=\{(x_1,y_1),(x_2,y_2),...,(x_m,y_m)\}$ is a given dataset, where x_i represents an input vector and y_i represents a class label corresponding to the input vector. The steps of Relief algorithm are presented as follow:

(1) Randomly select a sample x_i. Firstly, search its nearest sample $x_{i,np}$ from samples of its same class. Repeat by searching its nearest sample $x_{i,nq}$ from samples of its different class. Then, the component of attribute j that the relevant statistics corresponds to is Equation (1)

$$\delta^j=\sum_i -diff\left(x_i^j,x_{i,np}^j\right)^2+diff\left(x_i^j,x_{i,nq}^j\right)^2 \tag{1}$$

where x_i^j represents the value of the sample x_i in the attribute j, and $diff(x_a^j,x_b^j)$ depends on the type of the attribute j: if the attribute j is discrete, then when $x_a^j=x_b^j$, $diff(x_a^j,x_b^j)=0$, otherwise $diff(x_a^j,x_b^j)=1$; if the attribute j is continuous, then $diff(x_a^j,x_b^j)=|x_a^j-x_b^j|$.

(2) If the distance between x_i and $x_{i,np}$ in the attribute j is less than one between x_i and $x_{i,nq}$, which illustrates it is beneficial for the attribute j to distinguish samples of the same class from samples of different class, increase the component of statistics that the attribute j corresponds to; otherwise, if the distance that is between x_i and $x_{i,np}$ in the attribute j is greater than one between x_i and $x_{i,nq}$, which illustrates it is negative for the attribute j to distinguish samples of the same class from samples of different class, decrease the component of statistics that the attribute j corresponds to.

(3) Average the estimation results obtained based on different drawn samples, then obtain the relevant component of statistics of each attribute. The bigger the component value, the stronger the classifying ability of the corresponding attribute is.

(4) Repeat the above random sampling steps for M times and repeatedly operate for N times, then average the relevant statistics of each attribute. At last, employ the top k attributes of bigger component of the relevant statistics to classify.

3. Mixed Kernel Function SVM Model

3.1. SVM Algorithm

SVM is a supervised classification model, which is very classical in the field of machine learning and is usually used in pattern recognition, classification and regression. The principle of SVM is to apply the classification hyperplane to separate two kinds of sample points in the space and get the maximum classification margin [2,11,19]. The optimal classification hyperplane is represented as Equation (2)

$$f(x) = \boldsymbol{w} \bullet \boldsymbol{x} + b \tag{2}$$

where **w** is a normal vector of the optimal classification hyperplane and **x** is an input vector. To transform the nonlinear problem into linear problem to solve, the nonlinear problem is needed to map in a high dimension space, then the decision function of classification in the high dimension is Equation (3)

$$f(x) = w \bullet \varphi(x) + b \tag{3}$$

The constraint condition is as Equation (4)

$$f_i\left[w \cdot \phi(x_i) + b\right] - 1 + \xi_i \geq 0 \tag{4}$$

where i=1,2,…n, x_i refers to ith trained data. In addition, ξ_i is introduced as a slack variable which can prevent big influence caused by rare wrong data to the classification model, and $\xi_i \geq 0$. If the samples were classified correctly, ξ_i was set as 0, otherwise $\xi_i > 0$.

The classification hyperplane should make the minimum distance of the two kinds of samples from the hyperplane maximize, then the optimal objective function is as Equation (5)

$$\min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right) \tag{5}$$

where C is a penalty factor which represents the degree of punishment for misclassified samples. The Lagrange multipliers and quadratic programming optimization method were adopted to transform the problem as a dual problem as Equation (6).

$$\max\left[\frac{1}{2}\sum_{j=1}^{n}\alpha_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i, x_j)\right] \tag{6}$$

where α_i is the Lagrange multiplier, then the optimal decision function can be represented as Equation (7):

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{I} y_i \alpha_i (x_i, x) + b \right) \tag{7}$$

where I refers to the number of support vectors. The nonlinear kernel function $K(x_i,x)$ can be used to map the data to the feature space of the high dimension, then the decision function is as Equation (8):

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{I} y_i \alpha_i K(x_i, x) + b \right) \tag{8}$$

Therefore, the two keys most factors which influences the classification capability of SVM are: (1) the penalty parameter C; (2) the kernel function and the parameter g.

3.2. Problem Formulation

The local kernel function of SVM has strong learning capacity and weak generalization ability. On the contrary, the global kernel function has strong generalization capacity and weak learning ability. To further improve the fitting precision and the generalization ability of SVM, two kinds of kernel functions were combined to construct the mixed kernel function.

Four kinds of kernel function of SVM commonly used are:

(1) Linear kernel function as Equation (9)

$$K(x_i, x) = x_i \cdot x \tag{9}$$

(2) Polynomial kernel function as Equation (10)

$$K(x_i, x) = (x_i \cdot x + 1)^d \tag{10}$$

(3) Radial basis kernel (RBF) function as Equation (11)

$$K(x_i, x) = \exp\left(-\|x - x_i\|^2 / \sigma^2 \right) \tag{11}$$

(4) Sigmoid kernel function as Equation (12)

$$K(x_i, x) = \tanh\left(a \langle x_i, x \rangle + r \right) \tag{12}$$

The combinations of commonly used kernel function can construct mixed kernel function. The study combined local kernel function used broadly, i.e. RBF function with global kernel function, and polynomial kernel function to construct mixed kernel function of SVM as Equation (13):

$$K_{mix} = \lambda K_{rbf} + (1 - \lambda) K_{poly} \tag{13}$$

where $\lambda \in [0,1]$. After the kernel functions were mixed, the weight coefficient needed to optimize was added. λ was compared and determined by many experiments.

4. Parameter Optimization

4.1. the Method of Grid Search

The method of grid search is a traditional parameter optimization method and its idea is to divide the parameters, which will be optimized into grid, in a certain interval within a certain range. Then, substitute the values of all points of grid into the classifier to obtain the training accuracy; in other words, traverse all points of grid. The optimization principle of the method of grid search can obtain the mean square error CVmse in this set of values by the method of K-CV (K fold cross-validation) for a certain combination of the parameters (c,g). CVmse is shown as (14). At last, the set of values which made CVmse highest was determined as the optimal parameter.

$$CVmse = \frac{\sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2}{n} \tag{14}$$

where y_i,y_i' respectively represents the actual value and predicted value of the classification labels which the training data correspond to, and n represents the number of training data as Equation (14).

Because the method of grid search can traverse all points of the grid, it can always find the optimal parameter as long as the search scope of the parameter is big enough and the search interval of one is small enough. But, in the meantime such approach brings an issue that the computation time wastes since the classification accuracies of most points of grid are lower, and the points that can reach higher classification accuracy are within smaller interval range.

4.2. Genetic Algorithm (GA)

Genetic Algorithm is a heuristic method derived from biologically genetic rules. When organisms produce offspring, there will be genetic cross and mutation. Individuals whose fitness is low will be eliminated and ones whose fitness is high can survive, and their numbers become more and more. After the natural selection of N generations, all preserved individuals have high fitness. Genetic algorithm is based on such an idea and produces the next generation of individuals by the operation of replication, cross and mutation, and gradually weeds out the solutions with function values of low fitness, while increasing the solutions with function values of high fitness.

The detailed operation steps to optimize the parameters of SVM by GA can be summarized as follows.

Step 1. Set the initialization (the size of initial population of GA, the maximum of genetic generations T, crossover and mutation probability).

Step 2. Encode the parameters needed to be optimized with binary format according to setting range and generate initial population randomly. The chromosome was composed of every parameter, which ranked in the order of binary format, and the length of the chromosome is the sum of binary length of every parameter. In addition, the counter of genetic generations is set as t=0.

Step 3. Compute the fitness of every individual in the population. Set the classification accuracy of SVM as the value of target function, namely the fitness of individuals. The higher the correct rate of classification that the individuals matched was, the bigger the fitness of individual was.

Step 4. Based on the fitness of individuals, elect individuals to enter the next generation from current population in terms of to certain rules (use the method of roulette).

Step 5. Select two individuals x_1, x_2 of population as father to cross at a probability (crossover probability) and produce two new individuals. Here, adopt single-point crossover and set the crossover probability as 0.8.

Step 6. Randomly select the individuals of population to execute the mutation operation at a probability (mutation probability). Generate new individuals by changing some genes of individuals randomly and set the mutation probability as 0.05.

Step 7. Judge termination conditions. If $t \leq T$ then go to (2); if $t > T$ or the change of average fitness keep smaller than certain constant above certain generations, then output the individual with the maximum fitness as optimal solutions and the algorithm terminates.

Step 8. Decode the obtained optimal solution and obtain optimal parameters.

4.3. Combinatorial Optimization of Kernel Parameter

The traditional grid search and GA described above have its own advantages and disadvantages respectively, and the two kinds of methods have complementary advantages. The method of grid search can ensure that the optimal solutions can be found, but it needs to traverse all points of grid and spends too much time. On the other hand, GA is more intelligent and has fast searching speed, but it can only find the optimal solution with certain probability, and cannot ensure to obtain the absolute optimal solution.

In the study, the two kinds of searching algorithms were tried to combine. Firstly, GA was taken advantage of to search cursorily in initial population of parameters. The relative optimal solution was obtained in a large range by evolution and iteration, which can provide evidence for the range of the parameter optimization of grid search method and lighten the workload of grid search method. Next, the grid search method was employed to perform a second search accurately in the adjacent area of optimal solution determined in the first step to determine ultimate optimal solution. Not only can the computing time of GA and grid search method be shortened, but the optimal solution with higher precision can be found.

5. Experimental Results and Discuss

5.1. The Description of Experiment

The experiment was conducted in the environment of MATLAB R2009, and the open source toolkit LIBSVM are employed. Besides, UCI dataset of Breast Cancer and Heart Disease were also used. There are 683 samples and 9 features in the dataset of Breast Cancer, and there are 270 samples and 13 features in the dataset of Heart Disease.

5.2. Determination of λ in the Mixed Kernel Function

Table 1 presents the changes of training accuracy for the two datasets when λ was different. After comparison with many experiments, the accuracy of training set and testing set were all highest when λ=0.6. Thus, the weight coefficient in the mixed kernel function of SVM was set as 0.6.

Table 1. The training accuracy when λ was different

λ / dataset	The training accuracy of Breast Cancer	The training accuracy of Heart Disease
0	87.66%	87.15%
0.1	89.81%	88.26%
0.2	89.96%	88.63%
0.3	90.25%	89.74%
0.4	90.69%	91.22%
0.5	91.27%	93.44%
0.6	92.64%	95.29%
0.7	91.86%	94.19%
0.8	91.03%	91.96%
0.9	89.23%	90.11%
1.0	88.25%	89.00%

5.3. Relief Feature Selection

Relief parameters were set as follows: the sampling numbers m=80, and operating numbers N=20. Figure 1 and Figure 2 show the value of the average weight by using Relief algorithm to compute the feature of Breast Cancer set and Heart Disease set for 20 times respectively. It can be seen that the weights among features differ significantly. Finally, the first 5 features with bigger weights in the dataset of Breast Cancer and Heart Disease were selected as the key feature subset for the next classification, and the features with smaller weights were removed.

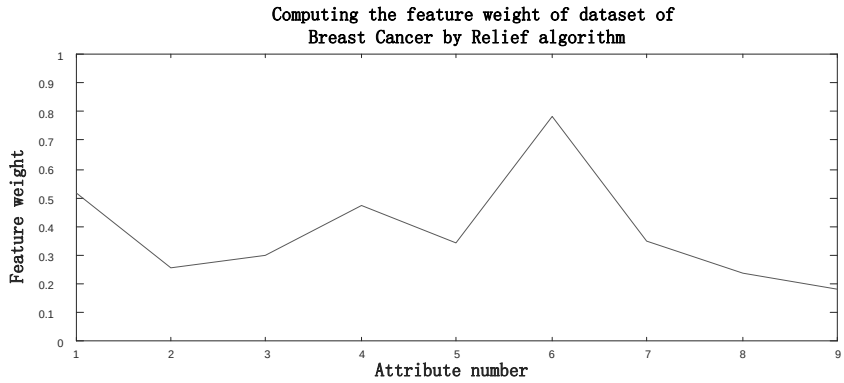

Figure 1. The figure of feature weight of dataset of Breast Cancer

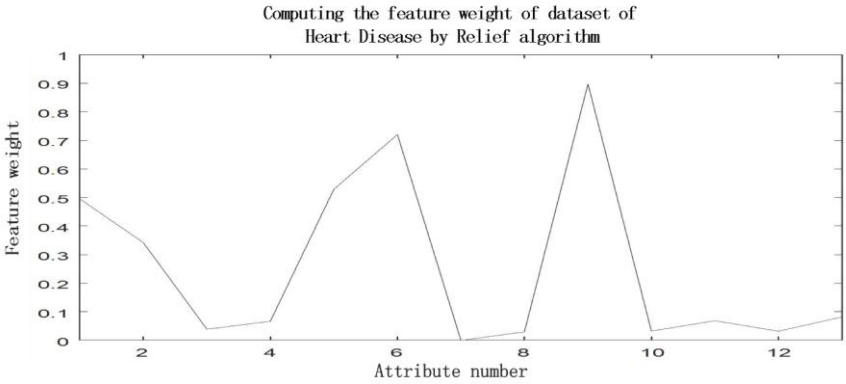

Figure 2. The figure of feature weight of dataset of Heart Disease

5.4. Verification of Model of Combinatorial Optimization

The mixed kernel function proposed in the paper adopted the global polynomial kernel function and local RBF kernel function. Thereby, the combinatorial optimization method of parameters was used in the two kernel functions. The performance of SVM of polynomial kernel function depends on the parameter set (C,d) and the performance of SVM obtained by the different parameter set is also different. Likewise, the performance of SVM of RBF kernel function depends on the parameter set (C,σ) and the classification models obtained by training are different if different C andσare selected.

In the experiment, the maximal number of iterations of GA is usually set as [100,500]; here, it was set as 200. Next, the number of population is usually set as [20,100]; here, it was set as 80. In the end, the crossover probability and the mutation probability was set as 0.6 and 0.05 respectively. After bestc and bestg were determined by GA, the search range of the parameter C and the parameter g in grid search method were set as [0.5*bestc,2*bestc] and [0.5*bestg,2*bestg] respectively. The second optimal parameters were searched accurately in grid.

The final parameter optimization results were shown as Figure 3 and Figure 4; the parameter optimization result of RBF kernel function and polynomial kernel function are shown. Eventually the optimal parameters of polynomial kernel function were bestc=0.57435 and bestg=0.1086, and ones of RBF kernel function were bestc=0.87055 and bestg=0.0625.

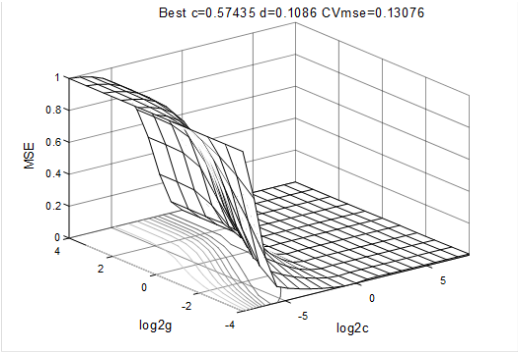

Figure 3. Results of polynomial kernel function parameter optimization

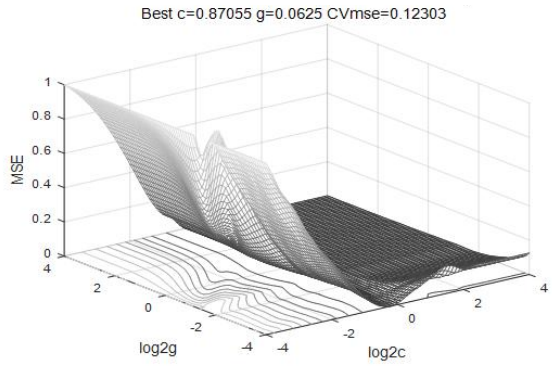

Figure 4. Results of RBF Parameter Optimization

Table 2. SVM classification results of different models

Algorithm	The accuracy of dataset of Breast Cancer		The accuracy of dataset of Heart Disease	
	Training set	Testing set	Training set	Testing set
Unoptimized SVM	88.97%	87.11%	88.44%	86.78%
GA-SVM	90.23%	90.19%	90.67%	90.11%
Grid-SVM	89.35%	88.43%	89.56%	89.00%
GA-Grid-SVM	92.86%	91.51%	92.33%	91.22%

Table 2 presents the comparisons between the results of the parameter optimization of SVM through three different parameter optimization methods including GA, grid search and combinational optimization (GA-Grid) and that of unoptimized SVM classifier. From data of Table 2, we can see that the classification accuracy of the two datasets through combinational optimization of GA and grid search was improved to some degree compared with other three methods. In contrast to the unoptimized classifier, the classification accuracy of training set of Breast Cancer set through combinational optimization method was increased from 88.97% to 92.86% and that of testing set of Breast Cancer set was increased from 87.11% to 91.51%; the classification accuracy of training set of Heart Disease set through combinational optimization method was increased from 88.44% to 92.33% and that of testing set of Breast Cancer set was increased from 86.78% to 91.22%. These results verify the effectiveness of combinational optimization algorithm.

5.5. SVM Classification by the Mixed Kernel Function

The key feature subsets of Breast Cancer set and Heart Disease set, which were elected in the last step, were taken as the input of SVM to train. The method of 10-folds cross validation was adopted to select the training data. Every dataset was randomly divided into ten equal parts. Every time one subset of them was taken as the testing set, the other nine subsets were taken as the training set. The experiment repeated 10 times until all subsets were taken as testing set. The accuracy of classification, which referred to a percentage of the number of samples classified correctly against total samples, was taken as the evaluation index.

To verify the validity of SVM classification model that combined Relief with mixed kernel function of SVM, our method was compared with the other five methods including SVM which did not use the feature selection and instead used single RBF kernel function, Relief-SVM which used Relief to select the key features and used single RBF kernel function, mixed kernel SVM which did not use the feature selection and used mixed kernel function, KNN (K-Nearest Neighbor) and BP neural network. The results were presented in Table 3 and Table 4, where the four evaluation indexes: Mse (mean-square error), R(Relation coefficient), and the accuracy rate of training set and accuracy rate of testing set, were used. The nearer Mse approximates to zero and the nearer R approximates to 1. The higher the accuracy rate becomes, the better the classification effect of the corresponding method becomes.

Table 3. Comparison of classified information by using different methods in Breast Cancer dataset

Method	Mse	R	Accuracy of training set	Accuracy of testing set
SVM of single kernel function	0.123	0.8758	89.87%	87.13%
Relief-SVM	0.102	0.9045	91.41%	90.47%
SVM of mixed kernel function	0.108	0.9106	91.80%	89.89%
SVM of Relief-mixed kernel function	0.064	0.9433	94.27%	93.23%
KNN	0.085	0.9075	91.48%	88.68%
BP neural network	0.076	0.9156	92.07%	89.79%

Table 4. Comparison of classified information by using different methods in Heart Disease dataset

Method	Mse	R	Accuracy of training set	Accuracy of testing set
SVM of single kernel function	0.118	0.8823	89.61%	86.29%
Relief-SVM	0.106	0.9078	91.09%	89.24%
SVM of mixed kernel function	0.126	0.9092	92.58%	90.71%
SVM of Relief-mixed kernel function	0.084	0.9126	94.56%	92.18%
KNN	0.096	0.9076	92.07%	89.92%
BP neural network	0.091	0.9102	93.36%	91.65%

We can see from Table 3 and Table 4 that the four evaluation indexes of the model of Relief-mixed kernel function SVM proposed in the paper were all better than the other methods by comparing the above other methods. The method proposed in the paper approached maximized performance than other methods. It illustrates some issues as follows. (1) It is of great significance for promoting the accuracy of classification to select the features, and obtain some features that have more influence on the disease and remove some features that have smaller influence on the disease. (2) The SVM with the mixed kernel function used in the study has better fitting precision and stronger generalization capability than one with single kernel function. (3) The method using SVM with the feature selection and linear combination of the kernel function is superior to the methods using traditional KNN and BP neural network.

6. **Conclusions**

In response to the rapid growth of technology of artificial intelligence and pattern recognition, it has become more and more popular to apply SVM to the practical problems. To deal with the redundant information of the classification feature and the drawback that the SVM of the single kernel function exists, the paper puts forward the classification model based on Relief feature selection and SVM with mixed kernel function. In order to improve the performance of SVM, the combinational optimization method combining Genetic Algorithm (GA) with grid search was also used, which can reduce the computation time and find optimal solutions. The experiment on Breast Cancer dataset and Heart Disease dataset of UCI dataset validated the validity of the proposed method and the results have shown that the learning and generalization capability of SVM with mixed kernel function are markedly improved.

Future study to improve the results of this paper mainly includes the following: (1) The construction of mixed kernel function of SVM in this paper is only a linear one; we can try to construct mixed kernel function of SVM with a nonlinear manner and test its effect. (2) The dataset of UCI used in this paper have had more than 20-year history and inevitably has flawed, which will influence the results of experiments. We can try newer dataset for the experiments.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (61672374), soft science research project in Shanxi Province (2016041035-1) and scientific research subject of health and family planning commission of Shanxi province (No.2017124).

References

1. V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, J. M. Benítez, and F. Herrera, "A Review of Microarray Datasets and Applied Feature Selection Methods," Information Sciences, vol.282, no.5, pp.111-135, 2014.
2. Y. C. Chen and C. T. Su, "Distance-based Margin Support Vector Machine for Classification," Applied Mathematics and Computation, vol.283, no.12, pp.141-152, 2016.
3. J. Dhalia Sweetlin, H. Khanna Nehemiah, and A. Kannan, "Feature Selection Using Ant Colony Optimization with Tandem-Run Recruitment to Diagnose Bronchitis from CT Scan Images," Computer Methods and Programs in Biomedicine, vol.145, no.7, pp.115-125, 2017.
4. D. D. Du, X. L. Jia, and C. B. Hao, "A New Least Squares Support Vector Machines Ensemble Model for Aero Engine Performance Parameter Chaotic Prediction," Mathematical Problems in Engineering vol.2016, Article ID 4615903, 8 pages,2016
5. S. Foithong, O. Pinngern, and B. Attachoo, "Feature Subset Selection Wrapper Based on Mutual Information and Rough Sets," Expert Systems with Applications, vol. 39, no.1, pp. 574-584, 2012.
6. T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, and M. Schummer, "Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data," Bioinformatics, vol. 16, no. 10, pp. 906–914,2000.
7. J. B. Geng, L. K. Sun, and S. X. Chen, "Parameters Optimization of Combined Kernel Function for Support Vector Machine," Journal of Computer Applications, vol.33, no.5, pp.1321-1323,1356, 2013.
8. B. Gu, G. S. Zheng, and J. D. Wang, "Analysis for Incremental and Decremental Standard Support Vector Machine," Journal of Software, vol.24, no.7, pp.1601-1613,2013.
9. J. P. Hua, W. D. Tembe, and E. R. Dougherty, "Performance of Feature Selection Methods in the Classification of High-

Dimension Data," Pattern Recognition, vol. 42, no.3, pp. 409-424, 2009.
10. F. Kang, J. S. Li, and J. J. Li, "System Reliability Analysis of Slopes Using Least Squares Support Vector Machines with Particle Swarm Optimization," Neurocomputing, vol.209, no.15, pp.46–56, 2016.
11. K. Li, Y. Wu, Y. Nan, P. Li, and Y. Li, "Hierarchical Multi-Class Classification in Multimodal Spacecraft Data Using DNN and Weighted Support Vector Machine," Neurocomputing, vol.259, no.15, pp.55-65, 2017.
12. S. Li, L. W. Li, D. F. Zhuang, and Y. Wang, "Research on Mixed Kernel Function and Its Application in the Field of Data Modeling," Computer Simulations, vol.32, no.7, pp.1-6, 2015.
13. T. W. Liao, "Feature Extraction and Selection from Acoustic Emission Signals with an Application in Grinding Wheel Condition Monitoring," Engineering Applications of Artificial Intelligence, vol. 23, no. 1, pp. 74–84, 2010.
14. K.C. Lin, S. Y. Chen, and J. Hung, "Feature Selection and Parameter Optimization of Support Vector Machines Based on Modified Artificial Fish Swarm Algorithms," Mathematical Problems in Engineering, vol.2015, Article ID 604108, 9 pages,2015.
15. C. Liu, W. Y. Wang, Q. Zhao, X. M. Shen, and M. Konan, "A New Feature Selection Method Based on a Validity Index of Feature Subset," Pattern Recognition Letters, vol.92, no.6, pp.1-8,2017.
16. X. L. Liu, D. X. Jia, and H. Li, "Research on Kernel Parameter Optimization of Support Vector Machine in Speaker Recognition," Science Technology and Engineering, vol. 10, no.7, pp. 1669-1673, 2010.
17. Y. Liu, J. W. Bi, and Z. P. Fan, "A Method for Multi-Class Sentiment Classification Based on an Improved One-Vs-One (OVO) Strategy and the Support Vector Machine (SVM) Algorithm," Information Sciences, vol.394-395, no.20, pp.38-52,2017.
18. H. J. Lu, J. Y. Chen, K. Yan, Q. Jin, Y. Xue, and Z. G. Gao, "A Hybrid Feature Selection Algorithm for Gene Expression Data Classification," Neurocomputing, vol.256, no.14, pp.56-62,2017.
19. J. Qu, H. Chen, W. Z. Liu, Z. Li, and B. Zhang, "Application of Support Vector Machine Based on Improved Grid Search in Quantitative Analysis of Gas," Chinese Journal of Sensors and Actuators, vol.28, no.5, pp.774-778,2015
20. M. R. G. Raman, N. Somu, K. Kirthivasan, R. Liscano, and V.S.S. Sriram, "An Efficient Intrusion Detection System Based on Hypergraph-Genetic Algorithm for Parameter Optimization and Feature Selection in Support Vector Machine," Knowledge-Based Systems, In Press, Corrected Proof, Available online 6 July 2017, pp.1-12, 2017.
21. B. K. Singh, K. Verma, A. S. Thoke, and J. S. Suri, "Risk Stratification of 2D Ultrasound-Based Breast Lesions Using Hybrid Feature Selection in Machine Learning Paradigm," Measurement, vol.105, no.4, pp.146-157,2017.
22. Q. J. Song, H.Y. Jiang, and J. Liu, "Feature Selection Based on FDA and F-Score for Multi-Class Classification," Expert Systems with Applications, vol. 81, no. C, pp. 22–27, 2017.
23. S. Szedmak, J. Shawe-Taylor, C. Saunders, and D. Hardoon, "Multiclass Classification by L1 Norm Support Vector Machine," in Proceedings of the Pattern Recognition and Machine Learning in Computer Vision Workshop, pp. 1-19, Grenoble, France, 2004.
24. A. Tharwat, A. E. Hassanien, and B. E. Elnaghi, "A BA-Based Algorithm for Parameter Optimization of Support Vector Machine," Pattern Recognition Letters, vol.93, no.7, pp.13-22, 2017.
25. Uncu and L. B. Türken, "A Novel Feature Selection Approach: Combining Features Wrappers and Filters," Information Sciences, vol. 177, no.2, pp. 449-466, 2007.
26. V. N. Vapnik. Statistical Learning Theory, New York, USA: Wiley Interscience, 1998.
27. Y. L. Wu, Q. He, and T. W. Xu, "Application of Improved Adaptive Particle Swarm Optimization Algorithm in WSN Coverage Optimization," Chinese Journal of Sensors and Actuators, vol. 29, no. 4, pp. 559-565, 2016.
28. X. Yang, H. Peng, and M. Shi, "SVM with Multiple Kernels Based on Manifold Learning for Breast Cancer Diagnosis," In: Proceeding of 2013 IEEE International Conference on Information and Automation (ICIA), IEEE Press, Yinchuang, China, pp. 396–399,2013.
29. X. Zhang, Z. H. Deng, S. T. Wang, and J. S. Cai, "Maximum Entropy Relief Feature Weighting," Journal of Computer Research and Development, vol. 48, no.6, pp. 1038-1048, 2011.
30. X. L. Zhang, W. Chen, B. J. Wang, and X.F. Chen, "Intelligent Fault Diagnosis of Rotating Machinery Using Support Vector Machine with Ant Colony Algorithm for Synchronous Feature Selection and Parameter Optimization," Neurocomputing, vol.167, no.16, pp.260-279,2015.
31. X. L. Zhang, X. F. Chen, and Z. J. He, "An ACO-based Algorithm for Parameter Optimization of Support Vector Machines," Expert Systems with Applications, vol.37, no.9, pp.6618-6628,2010.