A Modified Method Combined with a Support Vector Machine and Bayesian Algorithms in Biological Information

Wen-Gang Zhou, Yong-Feng Cui*, Ya Li

College of Computer Science and Technology
Zhoukou Normal University
Zhoukou 466001, China
E-mails: zhouwg@zknu.edu.cn, cuiyf@zknu.edu.cn, liya@zknu.edu.cn

**Corresponding author*

Received: January 07, 2015 **Accepted: June 18, 2015**

Published: June 30, 2015

Abstract: *With the deep research of genomics and proteomics, the number of new protein sequences has expanded rapidly. With the obvious shortcomings of high cost and low efficiency of the traditional experimental method, the calculation method for protein localization prediction has attracted a lot of attention due to its convenience and low cost. In the machine learning techniques, neural network and support vector machine (SVM) are often used as learning tools. Due to its complete theoretical framework, SVM has been widely applied. In this paper, we make an improvement on the existing machine learning algorithm of the support vector machine algorithm, and a new improved algorithm has been developed, combined with Bayesian algorithms. The proposed algorithm can improve calculation efficiency, and defects of the original algorithm are eliminated. According to the verification, the method has proved to be valid. At the same time, it can reduce calculation time and improve prediction efficiency.*

Keywords: *Modified method, Support vector machine, Bayesian method, Biological information.*

Introduction

With the deep research of genomics and proteomics [4, 12], the number of new protein sequences [24, 26] has expanded rapidly. With the obvious shortcomings of high cost and low efficiency of the traditional experimental method, the calculation method for protein localization prediction [1, 7, 16] has attracted more and more attention due to its convenience and low cost. It has become an important topic in bioinformatics.

Many scholars are devoted to the study of protein structure prediction [2, 15, 21]. Because of the great difficulty in the direct prediction of the three stage structures of protein, many methods begin the prediction work at two level structures [11], and the results of the prediction are used to predict the three level structure. Prediction of the protein structure also provides a powerful means for understanding the relation between the protein structure and its function. How to accurately and rapidly predict the protein structure is an unsolved problem.

With the machine learning techniques [13, 18] of predicting protein structure, the neural network [25] and the support vector machine (SVM) [6, 14, 20-23] are often used as learning tools. In the use of neural network for structure prediction, the PHD method uses three-layer feed forward neural network, and evolutionary information is included through a multiple sequence alignment. Besides PHD, there are also many other methods of using a neural

network system. For example, the high structure of a neural network which is composed of many small neural networks is used to predict the structure of the protein. With this method, the problem of excessive learning can be avoided. Moreover, the use of a neural network will combine each single series neural network and the multiple sequence alignment information will be included.

Since the classical SVM, proposed at the beginning of the 1990s, many good results in the practical application have been obtained due to its complete theoretical framework. In the field of machine learning, it has received wide attention and both the theory and the application have been developed. The method of the support vector machine is a novel classification technique. Application of the support vector machine has been successfully applied to various fields, such as protein far source detection in bioinformatics, micro array gene expression analysis, identification of the starting point of the transformation, protein structure prediction, intersection prediction between proteins and identification of the amino acid between the shrinkage, etc.

By using the support vector machine to predict the protein structure, some great results have been achieved [3, 6, 8, 10, 17]. Based on the solid theoretical foundation and better characteristics, the support vector machine is adopted in the prediction of the protein structure combined with Bayesian decision method [9, 19]. Based on the support vector machine algorithm, the Bayesian method is used to improve the prediction accuracy. The main contribution is the establishment of a modified method combining the support vector machine and the Bayesian method, and the remainder of the paper includes the following: A description of the support vector machine is introduced in Section 2. The modified support vector machine algorithm is summarized in Section 3. The results and analysis are shown in Section 4 and the conclusion is presented in Section 5.

Description of the support vector machine

A sketch map of the SVM is shown in Fig. 1. There are various kinds of SVM in different occasions:

(1) Fuzzy SVM;
(2) Least squares SVM;
(3) Weighted SVM;
(4) Active learning SVM;
(5) The combination of rough sets and SVM;
(6) Decision tree based SVM;
(7) Hierarchical clustering SVM.

The main class steps of the SVM are listed as follows:

(1) A known training set

$$T = \{(x_1, y_i), ..., (x_i, y_i)\} \in \{X \times Y\}^l \quad (1)$$

where $x_i \in X = R^n$, $y_i \in Y = \{-1, 1\}$, $i = 1, ..., l$

(2) Constructing and solving optimization problems

$$\max_{\alpha} \left\{ \sum_{i=1}^{l} \alpha_j - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \right\} \quad (2)$$

subject to $\sum_{i=1}^{l} y_i \alpha_i = 0$, $\alpha_i \geq 0$, $i = 1, ..., l$

and we can get the optimal solution $\alpha^* = (\alpha_1^*, ..., \alpha_l^*)^T$

(3) Selecting a positive component α_i^* after the calculation of $\omega^* = \sum_{i=1}^{l} \alpha_i^* y_i x_i$ and calculating the $b^* = y_i - \sum_{i=1}^{l} \alpha_i^* y_i (x_i \cdot x_j)$

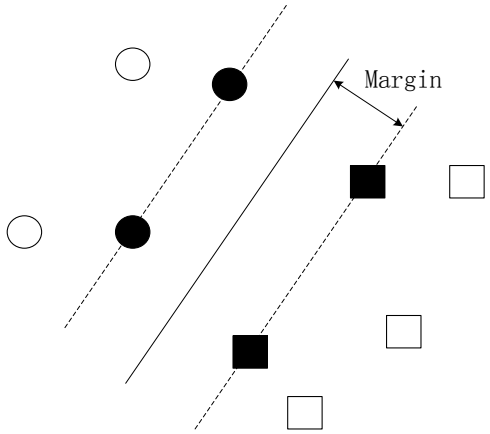

Fig. 1 Support vector machines

The main class processes of the SVM are listed as follows (Fig. 2):

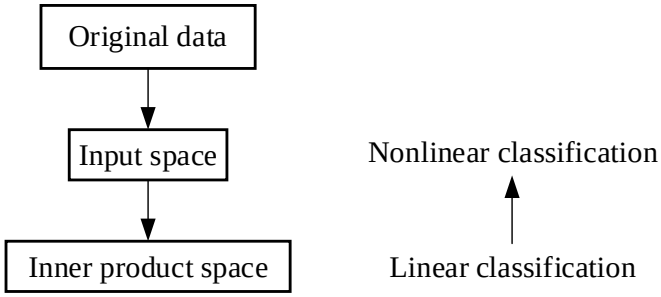

Fig. 2 The processing of SVM class process

(4) Constructing the classification hyper plane $((\omega^* \cdot x) + b) = 0$, and then the decision function of $f(x) = \text{sgn}\left\{\sum_{i=1}^{l} y_i \alpha_i^* (x_i \cdot x) + b^*\right\}$.

For nonlinear problems, we can transform the nonlinear function $X : x(x_1, ..., x_n)$ into a linear problem of $F : \Phi_1(x), ..., \Phi_1(x)$ in a high dimensional space, and then calculate the optimal classification face according to the transformation of space. According to the relevant functional theories, if a kernel function $K(x_i, x_j)$ meets the demand of the Mercer condition, it corresponds to the inner product in a transformation space. Therefore, the appropriate inner product function $K(x_i, x_j)$ can realize linear classification for nonlinear transformation. Computational complexity is not increased, and the corresponding classification function also becomes:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^{l} y_i \alpha_i^* \text{K}(x_i \cdot x) + b^*\right\} \quad (3)$$

In the prediction of protein two level structures, the training set should be first selected, and the feature sequence values for each sample point will be obtained by using the sliding window method with a length of 5-17. Middle amino acid residues which are corresponding to the two level structures are the category value of the sample. Then, according to the above method, the discriminant functions are obtained. Taking the prediction sample into a discriminant function, the two level structure types of the protein can be determined.

Combined SVM

Linear separable SVM is to obtain the best generalization performance by maximizing the classification margin, and in common conditions, it usually cannot get the ideal state. Taking the positive relaxation factor into the optimization problem, we can describe the optimization problem as follows:

$$\min_{\omega}\left(\frac{\|\omega\|^2}{2}+C\sum_{i=1}^{n}\xi_i\right) \tag{4}$$

subject to $y_i[(\omega\cdot x_i)+b]\geq 1-\xi_i$

$i=1, ..., n$

$\xi_i\geq 0, i=1, ..., n$

Due to various parameters as ω, b and ξ_i, it is difficult to get the solution of the optimization problem. Converting the optimal hyper plane according to the Lagrange method, we can get the following description:

$$\min_{\lambda}=\left(\frac{1}{2}\sum_{i}\sum_{j}\lambda_i\lambda_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{n}\lambda_i\right) \tag{5}$$

subject to $\sum_{i=1}^{n}\lambda_i y_i=0,\ 0\leq\lambda_i\leq C$,

where λ_i is the Lagrange multiplier. According to KKT conditions, training samples which are not included in the classification of face should meet the demands of $\lambda_i=0$; then, training samples on the surface with the condition of $\lambda_i>0$ are called the support vectors. With the condition of $0<\lambda_i<C$ and $\lambda_i=0$, the support vector machine is called non boundary support vector and boundary support vector, respectively.

The aim of the optimization method is to calculate the λ_i and then calculate the b. Here, we can classify the testing samples with the equation of $f(x)=\text{sgn}(\omega\cdot\Phi(z)+b)$.

If we avoid the dimensionality problem, the kernel function $K(x_i,z)=\Phi(x_i)\cdot\Phi(z)$ is adopted in the computation, according to the Mercer theory. With different strategies and parameters optimization of the support vector machine, the performance of the discriminant model based algorithm for protein classification is further improved.

The input sample data set of the SVM is n, and then the training sample will be:

$$s_n=((x_1,y_1),(x_2,y_2),...,(x_n,y_n)),\ y_i\in\{-1,1\} \tag{6}$$

For linearly separable training samples, SVM could find the hyper plane with the maximum Euclidean distance and the nearest training sample. And for the non-separable training

samples, the total error rate can be expressed with slack variables N_i. The calculation of the hyper plane is equal to the solution of the basic optimization problems as follows:

$$\min V(\omega, b, \xi) = \frac{1}{2}\omega^{\mathrm{T}}\omega + c\sum_{i=1}^{n}\xi_i \tag{7}$$

subject to $\quad \forall_{i=1}^{n} : y_i[\omega^{\mathrm{T}}x_i + b] \geq 1 - \xi_i$

$$\forall_{i=1}^{n} : \xi_i > 0$$

The restricted conditions of the formula described above will provide the correct classification of the training samples. If the sample is located on the wrong side of the hyper plane, the corresponding ξ_i will increase at a fast speed or equal 1. Therefore, $\sum_{i=1}^{n}\xi_i$ is the upper bound of the entire training error rate. The constant *C* is the punishment degree of the misclassified samples. By using the Lagrange multiplier method,

$$\omega_0 = \sum_{i=1}^{n} a_i y_i x_i \tag{8}$$

if the limit of $b = 0$ is set, it will make the hyper plane cross the edge. In the case of $a_i > 0$, x_i is called a support vector.

Consider two typical kernel functions as follows:

$$K(x_i, x_j) = (x_i{}^{\mathrm{T}}x_i + 1)^d \text{ and } K(x_i, x_j) = \exp(-r\left\|x_i - x_j\right\|^2) \tag{9}$$

Then, the decision function is obtained as $f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} a_i y_i K(x, x_i) + b\right)$.

After the prediction of the SVM, the Bayesian method is used to classify the protein structure.

Here the Bias method is considered as a two-class classification problem, the feature vector element is of two values and they are independent. Set $X = (x_1, ..., x_d)$, and each feature vector will give an answer of yes or no for the mode, where x_i is the predicted results by the support vector machine.

$$P_i = \Pr(x_i = 1|\theta_1) \text{ and } q_1 = \Pr(x_i = 1|\theta_2) \tag{10}$$

By assuming conditional independence, the probability of the element X can be written as $P(X|\theta_i)$. With this hypothesis, the class conditional probability can be expressed as follows:

$$P(x|\theta_1) = \prod_{i=1}^{d} P_i^{x_i}(1 - P_i^{1-x_i}) \tag{11}$$

and

$$P(x|\theta_1) = \prod_{i=1}^{d} P_i^{x_i}(1 - P_i)^{1-x_i} \tag{12}$$

Then, the Likelihood ratio is:

$$\frac{P(x|\theta_1)}{P(x|\theta_1)} = \prod_{i=1}^{d}\left(\frac{P_i}{q_i}\right)^{x_i}\left(\frac{1-P_i}{1-q_i}\right)^{1-x_i} \tag{13}$$

With the discriminant function

$$g(X) = \ln\frac{P(X|\theta_1)}{P(X|\theta_2)} + \ln\frac{P(\theta_1)}{P(\theta_2)} \tag{14}$$

$$g(X) = \sum_{i=1}^{d}\left[x_i \ln\frac{p_i}{q_i} + (1-x_i)\ln\frac{1-p_i}{1-q_i}\right] + \ln\frac{P(\theta_1)}{P(\theta_2)} \tag{15}$$

We can get the conclusion of the linear relation between the function and x_i. Then it can be expressed as:

$$g(X) = \sum_{i=1}^{d}\omega_i x_i + \omega_0 \tag{16}$$

where

$$\omega_i = \ln\frac{P_i(1-q_i)}{q_i(1-p_i)} \tag{17}$$

and

$$\omega_i = \sum_{i=1}^{d}\ln\frac{(1-p_i)}{(1-q_i)} + \ln\frac{P(\theta_1)}{P(\theta_2)} \tag{18}$$

If $G(x) > 0$, it is judged as θ_1. If $G(x) \le 0$, it is judged as θ_2. $G(x)$ is a weighted combination of each element of X.

Results and analysis

In order to verify the validity of the modified algorithm, two data sets have been selected for the prediction. In the simulation, the prediction precision and the prediction efficiency of the modified method have been taken to compare it with other methods. Besides, prediction precision for various dimer categories has also been given.

Data set I

With the characteristics of the combined algorithm proposed in the paper, a data set has been thoroughly selected. The data set is from reference [5], and it includes 914 homodimers, and 725 other multimers in the tertiary structure of a protein sequence. In order to exclude the membrane protein and other specific protein, the data set is restricted to prokaryotes and cytoplasm, which can be chosen from the SWISS-PROT.

From Fig. 3 we can see that the combined prediction success rate is much higher than that by PseAA. During the Jackknife test, the overall prediction success rate of the combined SVM is about 11% higher than the results calculated by PseAA. In the independent dataset test (Fig. 4), the overall prediction success rate of the combined SVM is about 17% higher than the results calculated by PseAA.

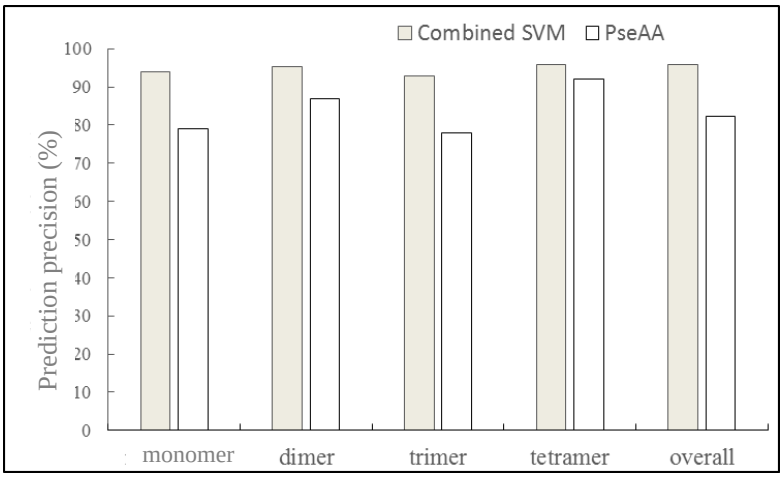

Fig. 3 Jackknife test

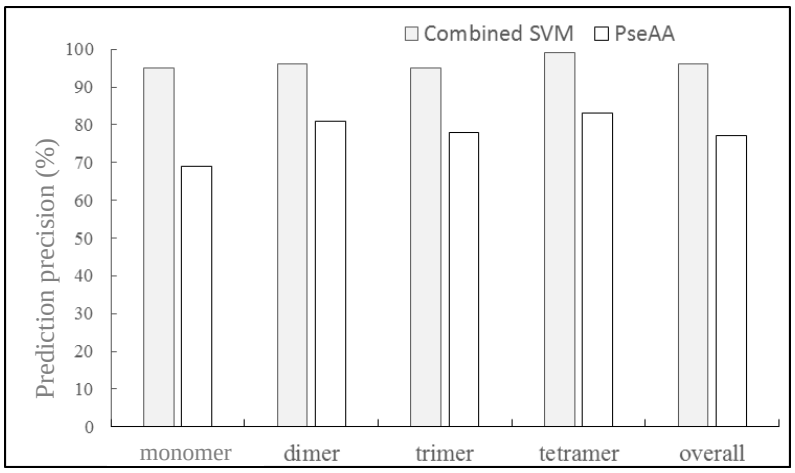

Fig. 4 Independent dataset test

Table 1 shows the comparison of the results with different feature extraction methods. The classification precision of the combined method has been improved to a certain extent. Compared with the traditional method, the total classification precision has increased to a certain extent. Correctness rate of classification experiments with the combined method is presented in Fig. 5.

Table 1. Comparison of the results using different feature extraction methods

	Traditional method		**Combined SVM**	
	Amino acid	**Dipeptide**	**Amino acid**	**Dipeptide**
(C, g)	(4, 2)	(8, 0.0625)	(5, 3)	(3, 0.02125)
Q	78.10	82.06	83.19	85.77
TPR	77.45	80.51	80.52	84.04
FPR	20.86	15.41	19.04	18.08
MCC	55.39	63.63	65.02	69.27

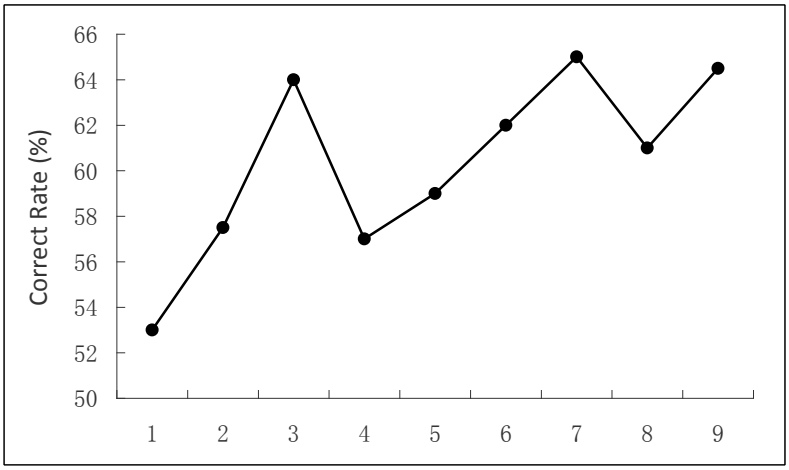

Fig. 5 Correctness rate of classification experiments with the combined method

From Table 2 we can see the validity of the combined method. The access indicators have increased slightly.

Table 3 shows the results of the independent testing test with different algorithms. All the indexes have improved slightly and the efficiency has been improved.

Table 2. Comparison of the results using methods with different parameters

	3 physicochemical parameters	**4 physicochemical parameters**	**5 physicochemical parameters**	**12 physicochemical parameters**
(C, g)	(8, 0.5)	(8, 0.5)	(8, 0.5)	(8, 0.0625)
TP	800	816	827	841
FN	114	98	87	92
TN	523	513	501	547
FP	202	212	224	221
Q	80.72	81.09	81.03	82.19
TPR	79.84	79.38	78.69	84.47
FPR	17.90	16.04	14.80	19.28
MCC	60.79	61.67	61.70	68.57

Table 3. Results of the independent testing test with different methods

	Pseudo amino acid composition	**Dipeptide composition**	**Linear combinatorial forecast**	**Combined SVM**
(C, g)	(8, 2)	(2, 0.02985)	(2, 0.0313)	(6, 0.0625)
TP	160	163	165	176
FN	16	12	9	8
TN	122	119	123	124
FP	23	21	22	25
Q	88.11	89.93	90.55	90.32
TPR	91.26	92.24	95.08	96.92
FPR	15.86	15.76	15.17	15.67
MCC	75.85	78.92	80.93	80.73

Data set II

The data set is acquired from the public protein sequence database SWISS-PROT, as shown in Table 4. Table 5 shows the higher overall accuracy of the modified SVM.

Table 4. Training and test data set

	Sample	
	Training set	**Test set**
Type I	435	492
Type II	152	181
Multipass	1311	1832
Lipid chain anchored	51	22
GP I anchored	110	92
Total	2059	2619

Table 5. Prediction results for different test methods using a combined SVM

	Successful prediction rate (%)					
	Type I	**Type II**	**Multipass**	**Lipid chain anchored**	**GPI anchored**	**Total**
Self-consistency	82.2	70.4	95.2	55.2	67.7	88.2
Independent dataset	79.1	67.5	93.8	35.1	59.8	87.4
Jackknife	68.3	66.9	95.4	15.33	63.8	83.5

From Table 6, we can conclude that the prediction results calculated by the combined SVM have a higher accuracy. It shows that the combined SVM has an obvious advantage in the protein prediction process.

Table 6. The prediction results for the five types of membrane proteins by different algorithms and test methods based on amino acid compositions

	Successful prediction rate (%)		
	Self-consistency	**Independent dataset**	**Jackknife**
The minimum Hamming distance	62.7	66.9	62.3
The minimum Euclidean distance	63.6	69.4	62.7
ProtLock	66.8	63.3	65.1
Covariance discriminant	82.1	78.9	76.8
Combined SVM	88.6	88.1	83.7

Conclusion

With the development of the genomics and proteomics, lots of new protein sequences have been studied. Traditional experimental methods have the obvious shortcomings of high cost and low efficiency, which is why the calculation method for protein localization prediction has attracted a lot of attention. In the machine learning techniques, the neural network and the SVM are often used as learning tools. Due to its complete theoretical framework, SVM has been widely applied.

In this paper, we make an improvement on the existing machine learning algorithm of the SVM algorithm, and a new improved algorithm has been developed, combined with Bayesian

algorithms. The improved algorithm can increase calculation efficiency, and the defects of the original algorithm are eliminated.

In the paper, the SVM and the Bayesian methods are used in bioinformatics to achieve better protein prediction. We use two data sets to verify the prediction success ratio of the combined algorithm, and the results show that the algorithm has a higher success ratio.

According to the verification, the methods are proved to be valid. The modified algorithm can be used in protein prediction effectively. At the same time, it can reduce the calculation time and improve the prediction efficiency.

Acknowledgements

This work was supported by the Natural Science Foundation of China under Grant 61103143, the Basic and frontier technology research projects of Henan province under Grant 132300410276.

References

1. Couto B. R. G. M., M. M. Santoro, A. P. Ladeira, M. A. dos Santos (2013). A Predictive Alignment-free Method based on Logistic Regression for Feature Selection and Classification of Protein Sequences, Proc. of the Int. Conf. on Bioinformatics Models, Methods and Algorithms, (Fernandes P., J. Solé-Casals, A. L. N. Fred, H. Gamboa (Eds.)), Bioinformatics, SciTePress, 171-177.
2. Custódio F. L., H. J. C. Barbosa, L. E. Dardenne (2014). A Multiple Minima Genetic Algorithm for Protein Structure Prediction, Applied Soft Computing Journal, 15, 88-99.
3. Fernandez-Lozano C., J. A. Seoane, M. Gestal, D. Rivero, J. Dorado, A. Pazos (2013). A Texture-based Classification Method for Proteins in Two-dimensional Electrophoresis Gel Images: A Feature Selection Method Using Support Vector Machines and Genetic Algorithms, Proc. of the Int. Conf. on Computer Vision Theory and Applications – VISAPP 2013, 1, 401-404.
4. Gagna C. E., W. C. Lambert, P. C. Lambert (2012). Advanced Techniques for Genomics and Proteomics: Transitional Structural Chemogenomics, Chemoproteomics, Pharmacogenomics, and Pharmacoproteomics, American Laboratory, 44(6), 14-21.
5. Garian R. (2001). Prediction of Quaternary Structure from Primary Structure, Bioinformatics, 17(6), 551-556.
6. Georgara D., K. Kermanidis, L. Mariolis (2013). Investigating Motif Selection Methods for Support Vector Machine Classification of Protein Sequences to Functional Families, Engineering Intelligent Systems, 21(2-3), 101-109.
7. He X.-M., Z. Qin, J. Chen, B. Liao, W.-J. Li, D.-F. Zhang (2014). New Method Expression of Protein Sequence and its Application for Protein Sub-cellular Localization Prediction, Journal of Computational and Theoretical Nanoscience, 11(3), 873-877.
8. Huang H.-L., Y. S. Srinivasulu, P. Charoenkwan, H.-C. Lee, S.-Y. Ho (2013). Designing Predictors of Halophilic and Non-halophilic Proteins Using Support Vector Machines, Proc. of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, IEEE SSCI 2013, 230-237.
9. Jie Y. (2013). A Multi-language Speech Recognition Method Based on Confidence Bayesian Decision-making, Advanced Materials Research, 717, 475-480.
10. Koppisetty C. A. K., M. Frank, G. J. L. Kemp, P.-G. Nyholm (2013). Computation of Binding Energies Including Their Enthalpy and Entropy Components for Protein-ligand Complexes Using Support Vector Machines, Journal of Chemical Information and Modeling, 53(10), 2559-2570.

11. Lin K.-L., C.-Y. Lin, C.-D. Huangr, H.-M. Chang, C.-Y. Yang, C.-T. Lin, C. Y. Tang, D. F. Hsu (2007). Feature Selection and Combination Criteria for Improving Accuracy in Protein Structure Prediction, IEEE Transactions on Nanobioscience, 6(2), 186-196.
12. Mohamadi M. R., L. Mahmoudian, N. Kaji, M. Tokeshi, H. Chuman, Y. Baba (2006). Nanotechnology for Genomics & Proteomics, Nano Today, 1(1), 38-45.
13. Ozdemir A. T., B. Barshan (2014). Detecting Falls with Wearable Sensors Using Machine Learning Techniques, Sensors (Switzerland), 14(6), 10691-10708.
14. Payal S., P. Chatterjee, S. Basu, M. Kundu, M. Nasipuri (2013). Comparisons of Different Feature Sets for Predicting Carbohydrate-binding Proteins from Amino Acid Sequences Using Support Vector Machine, Advances in Intelligent Systems and Computing, 201 AISC, 1, 519-529.
15. Rokde C. N., M. Kshirsagar (2013). Bioinformatics: Protein Structure Prediction, 4th International Conference on Computing, Communications and Networking Technologies – ICCCNT 2013, doi: 10.1109/ICCCNT.2013.6726753.
16. Romero-Ximil J. M., A. Díaz-Pérez (2009). An FPGA Design for Evaluating Score Function in Protein Energy Calculation, 19th International Conference on Field Programmable Logic and Applications, 666-669.
17. Singh L., G. Chetty, D. Sharma (2012). A Hybrid Approach to Increase the Performance of Protein Folding Recognition Using Support Vector Machines, Lecture Notes in Computer Science, 7376, 660-668.
18. Singh N., N. Chandra (2014). Integrating Machine Learning Techniques to Constitute a Hybrid Security System, Proc. of the 4th Int. Conf. on Communication Systems and Network Technologies – CSNT 2014, 1082-1087.
19. Sun L.-N., B. Yang (2013). An Explosives Identification Method Based on the Bayesian Decision Theory, Journal of Northeastern University, 34(1), 123-126.
20. Wang C. (2015). A Modified Machine Learning Method Used in Protein Prediction in Bioinformatics, Int J Bioautomation, 2015, 19(1), 25-36.
21. Wang S., F. Wang, S. Jiang, H. Zhang (2012). Research on Particle Swarm Optimization for Protein Structure Prediction, Advanced Materials Research, 605-607, 2497-2500.
22. Wu J., W. Han, D. Hu, X. Xu, S. Yan, L. Tang (2013). Recognition of MicroRNA-binding Sites in Proteins from Sequences Using Laplacian Support Vector Machines with a Hybrid Feature, Proc. of the 6th Int. Conf. on Biomedical Engineering and Informatics – BMEI 2013, 477-483.
23. Wu J.-S., Z.-H. Zhou (2013). Sequence-based Prediction of Micro RNA-binding Residues in Proteins Using Cost-sensitive Laplacian Support Vector Machines, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 10(3), 752-759.
24. Yona G., M. Levitt (2000). Unified Sequence-structure Classification of Protein Sequences: Combining Sequence and Structure in a Map of the Protein Space, Proc. of the Annual Int. Conf. on Computational Molecular Biology – RECOMB 2000, 308-317.
25. Zhang G.-Z., D.-S. Huang (2004). Radial Basis Function Neural Network Optimized by a Genetic Algorithm for Soybean Protein Sequence Residue Spatial Distance Prediction, Proc. of the Congress on Evolutionary Computation, CEC 2004, 1, 1015-1019.
26. Zhang Z., L. Liu, J. Li, Z. Zhang (2011). Spectral Representation of Protein Sequences, Journal of Computational and Theoretical Nanoscience, 8(7), 1335-1339.

Wen-Gang Zhou, M.Eng.
E-mail: zhouwg@zknu.edu.cn

Zhou Wen-Gang received his B.Eng. degree in Computational Mathematics from the Henan Normal University and his M.Eng. degree in Computer Application Technology from the North China University of Technology in 1997 and 2007, respectively. He is currently doing research in the Analysis and Design of Intelligent Algorithms.

Yong-Feng Cui, M.Sc.
E-mail: cuiyf@zknu.edu.cn

Yong-Feng Cui received his B.Sc. degree in Computer Science and Technology from the Henan Normal University and his M.Sc. degree in Computer Application Technology from the Huazhong University of Science and Technology, China, in 2000 and 2007, respectively. He is currently doing research in the Computer Application Technology.

Ya Li, M.Sc.
E-mail: liya@zknu.edu.cn

Ya Li received his B.Sc. degree in Computer Science and Technology from the Northeast Normal University and his M.Sc. degree in Computer Technology from the Beijing Jiaotong University, China, in 1995 and 2005, respectively. He is currently doing research in the Computer Application Technology and Mobile Internet Technology.