A Modified Machine Learning Method Used in Protein Prediction in Bioinformatics

Chengduan Wang

School of Computer Engineering
Shandong Science and Technology University
Weifang University, 261061, Weifang, China
E-mail: wangchengduan@163.com

Received: November 02, 2014 **Accepted: March 26, 2015**

Published: April 01, 2015

Abstract: *Biological information resources have been growing rapidly with the development of biological science and technology, and the development of computer technology and the Internet has made large scale data storage, processing and transmission possible. Machine learning is often used to learn from experience and get useful information. Protein prediction is a main part of biological information, and many prediction methods have been proposed. However, improving the prediction success rate is always a research goal. In this paper, machine learning techniques are used in bioinformatics for protein prediction, and the support vector machine algorithm is used to develop a new prediction algorithm. This method is combinatorial. Two data sets are used to verify the success rate of the modified algorithm, and the results show that the algorithm has a higher success rate. The modified algorithm can be effectively used in protein prediction.*

Keywords: *Bioinformatics, Machine Learning, Support Vector Machine, Protein Prediction.*

Introduction

The high popularity of machine learning techniques in bioinformatics is due to their being task-oriented. People can understand their theoretical foundation and the formation of reliable rules. People can describe the pair of input/output in detail, but cannot know the relationship between them, such as the protein folding mechanism [7, 13]. However, machine learning [12, 17] can give approximate solutions for a specific problem by automatic adjustment of the internal structure.

The amount of information is expanding fast, and the number of DNA bases in the nucleic acid sequence database Genbank is growing exponentially. Besides, data in protein sequence databases are rapidly increasing as well. Therefore, the development of new bioinformatics tools [4, 8, 10, 16] to mine the useful knowledge and information in the data source is becoming more and more important.

Protein type prediction [2, 11, 19] is a kind of important basic research, and many prediction methods have been proposed. The protein type has a significant relationship with the amino acid component [18]. In essence, protein types are determined by their amino acid sequences, and the prediction effect just by the component content of the amino acids is limited. With the development of proteomics, a number of different types of protein will be identified and added to the database, and the enhancement of the training set can easily be found.

The support vector machine (SVM) algorithm [23-25] is a hot point research in machine learning and is widely used in many fields in bioinformatics.

Aiming at these problems, this paper puts forward a kind of protein classification algorithm based on the discriminant model. The algorithm is the optimization of the SVM itself, and this method can be applied to other classification algorithms based on protein identification models to achieve higher classification accuracy. The main contribution of this paper is the establishment of a modified prediction method for protein classification, and the remainder of the paper is organized as follows: Methods for protein classification are introduced in Section 2. A SVM model is summarized in Section 3. Results and analysis are described in Section 4. The conclusion is presented in Section 5.

Methods for protein classification

There are mainly three kinds of protein classification methods:

A) Double sequence alignment algorithm. It is the most basic algorithm for remote homology detection. It determines the correlation between any two sequences by comparing their difference. The homology is judged by the correlation. The typical global sequence alignment algorithm is the Needleman Wunsch algorithm, while the typical local sequence alignment algorithms are the Smith Waterman algorithm and the heuristic local sequence alignment algorithm. Some methods increase the comparison speed at the expense of some sensitivity to obtain a wide range of applications.

B) Statistical model based method. This kind of algorithm can be used to establish the statistical model for the protein families according to a multiple sequence alignment to estimate the probability. Then, an unlabeled protein sequence is compared with the statistical models. This method can predict the relative homology three times more effectively than the pairwise sequence alignment algorithm. The PSI-BLAST algorithm [3, 14, 20] develops a profile based method. It uses a site specific alignment to establish a multiple sequences alignment from large databases and to establish a dedicated search matrix for each alignment. This can reflect the probability of occurrence of amino acid replacement at different positions more accurately and improve the performance of the model.

C) The discriminant model based classifier. In the algorithm, the protein sequence is mapped into a high dimension space vector with a fixed length, and vectorized training samples are used to establish the training model with a SVM [1]. After the determination of the maximum hyper plane, the unlabeled samples are placed into the vector space classification. Compared to the previous method, the discriminant model based classifier provides higher accuracy of classification due to the decision making with positive and negative samples information. At present, a large number of studies focus on the kernel function of the sequence provided for SVM training. The SVM-pairwise algorithm [25] uses an empirical kernel and a double sequence alignment algorithm to transform the sequence into comparison value.

The SVM-Fisher algorithm [9] uses Fisher kernel and a profile hidden Markov model to express the sequence into Fisher scores. The Mismatch algorithm with String Kernel has higher computational efficiency [15, 21]. Profile SVM [9, 22], which is based on the Mismatch generated by the PSI-BLAST sequence to express the sequence, is the SVM based classifier with the best performance. But the entire protein classification algorithm based on SVM faces the problem of far fewer positive training samples than negative class samples. This means that there is a training imbalance and it will affect the ability of the discriminant model to further improve the classification performance of the classifier.

Support vector machine model

The most accurate protein homology detection method is based on the discriminant model. These methods are derived from a simple idea: the protein sequence is mapped into a high dimension space vector with a fixed length. For each protein structure class, positive and negative samples are used to establish the SVM model, where positive samples mean protein sequences belonging to the structure of class and negative samples mean protein sequence not belonging to the class.

The SVM only performs a simple task: judging the positive and the negative on a vector of a fixed length with a maximum margin hyper plane. Then, the protein sequence vector will be placed into the trained support vector machine, and we can draw the determination of whether it belongs to the structure of the class. Based on the SVM method and due to the modeling process for decision making with both positive and negative samples information, the prediction performance is better than other methods. The current study uses mainly a certain method to define the kernel function of the SVM to complete the vectorization of protein sequences. The protein sequences with indefinite length will be transformed into a high dimensional space vector and placed into the SVM to train and discriminate.

Modified SVM model

Linear separable SVM has to obtain the best generalization performance by maximizing the classification margin on the condition of correctly classified training samples. Usually, it is unable to reach the ideal state normally. In order to allow the SVM to construct the linear decision boundary at a linear inseparable case, it should be eclectic between a number of errors allowed in the linear decision boundary and the width of the interval. A positive relaxation factor should be introduced in the optimization problem with constraints. Then, the optimization problem can be described as:

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

subject to

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

$$i=1,...,n$$

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

Because of the large number of parameters involved as ω, b and ξ_i, the solution of the optimization problem is a thorny issue. Lagrange method can be used to convert the optimal hyperplane problem into its dual problem.

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

subject to

$$\sum_{i=1}^{n} \lambda_i y_i = 0, \quad 0 \le \lambda_i \le C$$

where λ_i is the Lagrange multiplier. According to Karush-Kuhn-Tucher (KKT) conditions, many training samples which are not in the classification of face must meet the condition of $\lambda_i = 0$, and training samples on the surface and with the condition of $\lambda_i > 0$ are called a support vector. If $0 < \lambda_i < C$, the corresponding SVM is called a non boundary support vector, and if $\lambda_i = 0$, the corresponding SVM is called a boundary support vector. In fact, the latter consists of the misclassified training sample points.

The optimization method is used to get the value of λ_i and further calculate the parameter *b*. Then, we can classify the testing samples with the classification function listed as follows:

$$f(x) = \text{sgn}(\omega \cdot \Phi(z) + b) \tag{3}$$

In order to avoid the curse of the dimensionality problem and according to the Mercer theory, the kernel function $K(x_i, z) = \Phi(x_i) \cdot \Phi(z)$ is adopted to replace the vector dot product in the computation. At present, a lot of research through the design of different kernel functions for protein classification has gained great achievements. With different strategies and optimization of parameters 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:

$$s_n = ((x_1, y_1), (x_2, y_2), ..., (x_n, y_n)), \quad y_i \in \{-1, 1\} \tag{4}$$

For the linearly separable training samples, SVM could find the hyperplane with the maximum Euclidean distance and the nearest training sample. This distance is called classification space *D*, as shown in Fig. 1.

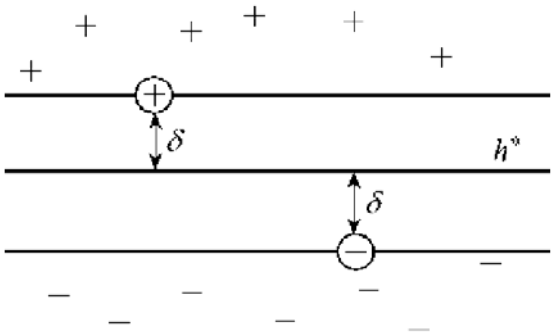

Fig. 1 Schematic drawing of the classification of two classes in 2D space [6]

For the non-separable training samples, the total error rate can be expressed with slack variables N_i. Computing the hyperplane can be equal to the solution of the basic optimization problems as in the following equation:

$$\min V(\omega, b, \xi) = \frac{1}{2}\omega^{\mathrm{T}}\omega + c\sum_{i=1}^{n}\xi_i \tag{5}$$

subject to

$$\forall_{i=1}^{n} : y_i[\omega^{\mathrm{T}} x_i + b] \geq 1 - \xi_i$$
$$\forall_{i=1}^{n} : \xi_i > 0$$

The restricted conditions of the formula described above have the requirements for a slack variable N_i to keep all the training samples correctly classified. If the sample falls on the wrong side of the hyper plane, the corresponding ξ_i will increase greatly or equal to 1. Therefore, $\sum_{i=1}^{n} \xi_i$ is the upper bound of the entire training error rate. The constant C plays the role of a control punishment degree of the misclassified samples. By using the Lagrange multiplier method,

$$\omega_0 = \sum_{i=1}^{n} a_i y_i x_i \tag{6}$$

and the classification gap $\delta = \frac{1}{\|\omega_0\|}$. If we set the limit of $b = 0$, it will make the hyperplane cross the edge. In the case of $a_i > 0$, x_i is called a support vector. The decision function will be:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} a_i y_i x^{\mathrm{T}} x_i + b \right) \tag{7}$$

The input vector is mapped into a high dimensional feature space by a kernel function, and two typical kernel functions are listed as follows:

$$K(x_i, x_j) = (x_i^{\mathrm{T}} x_i + 1)^d \tag{8}$$

$$K(x_i, x_j) = \exp(-r\|x_i - x_j\|^2) \tag{9}$$

Eq. (8) is called the polynomial kernel function with dimension of d. When d = 1, Eq. (8) is converted into a linear kernel function. Eq. (9) is called the radial basis kernel function, where r is the initialization parameter of the kernel function. The decision function is obtained as:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} a_i y_i K(x, x_i) + b \right) \tag{10}$$

By choosing a kernel function and an adjusted parameter C, the corresponding SVM model will be obtained.

Parameter optimization

In the standard linear inseparable SVM, for the boundary support vector, we have:

$$\lambda_i = C \tag{11}$$

If we set N_{b+} to be the support vector number of the positive kind boundary and N_+ to be the number of the positive samples, we can get:

$$N_{b+}C \le \sum_{y_i=+1} \lambda_i = M \tag{12}$$

and

$$N_{b+} \le \frac{M}{C} \tag{13}$$

$$\sum_{i=1}^{n} \lambda_i y_i = \sum_{y_i=+1} \lambda_i - \sum_{y_i=-1} \lambda_i = 0 \tag{14}$$

Then, we can get $\sum_{y_i=+1} \lambda_i = \sum_{y_i=-1} \lambda_i$ and $N_{b-} \le \frac{M}{C}$.

With the comparison of the classification error rate of the positive and negative classes and the same penalty parameter C, the classification rate of the samples with a larger number is low, while it is high with a small samples number. It means that in the standard linear SVM, classes with more samples have a greater weight. In protein classification, the number of the negative class samples is bigger than that of the positive class samples. However, positive class samples are more significant than negative samples. To solve this problem in the standard SVM, different penalty parameters are needed for positive samples and negative samples to make positive and negative classes get the same weight in order to balance the SVM classification.

Setting a penalty parameter $C_+ = \frac{N_+}{N_+ + N_-} C$ for the positive kind and a penalty parameter $C_- = \frac{N_+}{N_+ + N_-} C$ for the negative kind, both classification error rates of the positive and negative kinds will have the same upper bound. This will improve the positive attention degree when the number of samples is smaller, and improve classification performance when there is a serious imbalance between the number of positive samples and the number of negative ones. With the new penalty parameter, the original optimization problem can be described as:

$$\min_{\lambda} = \left(\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{n} \lambda_i \right) \tag{15}$$

subject to

$$\sum_{i=1}^{n} \lambda_i y_i = 0$$

$$0 \le \lambda_i \le C_+ \qquad (y_i = +1)$$

$$0 \le \lambda_i \le C_- \qquad (y_i = -1)$$

The SVM is trained with optimized parameters to get the optimal classification plane, and then protein sequences can be classified.

Classification performance test

The classification results are assessed with two kinds of objective and strict testing methods: one is the *K* cross method, and the other one is the independent testing method. The following four parameters are adopted to evaluate the effectiveness of the classification method: total classification accuracy (*Q*), True Positive Rate (*TPR*), False Positive Rate (*FPR*) and Matthews Correlation Coefficient (*MCC*) respectively. The definitions are as follows:

$$Q = \frac{TP + TN}{TP + FN + TN + FP} \tag{16}$$

$$TPR = \frac{TP}{TP + FN} \tag{17}$$

$$FPR = \frac{RP}{RP + TN} \tag{18}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \tag{19}$$

where *TP* is the number of the correctly classified positive samples, *TN* is the number of the correctly classified negative samples, *FP* is the number of the incorrectly classified positive samples, and *FN* is the number of the incorrectly classified negative samples.

Result and analysis

Data set I

This data set includes 914 homodimers, and 725 cognate multimers [5]. From Fig. 2 we can see that the modified prediction accuracy is higher than PseAA.

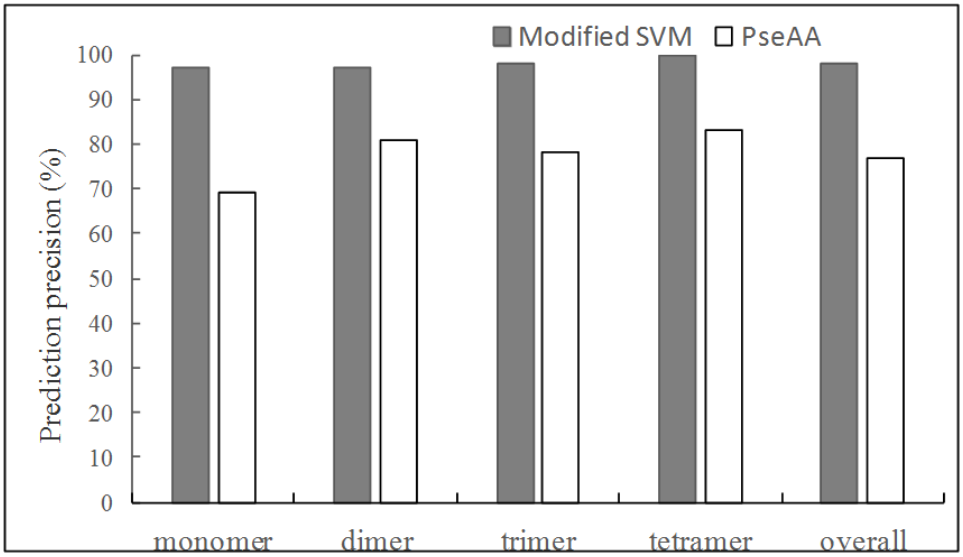

Fig. 2 Independent dataset test

During a Jackknife test, the overall prediction accuracy of the modified SVM is about 10% higher than that of PseAA (Fig. 3). In the independent dataset test, the overall prediction accuracy of the modified SVM is about 16% higher than that of PseAA. The correctness rate of classification experiments is shown in Fig. 4.

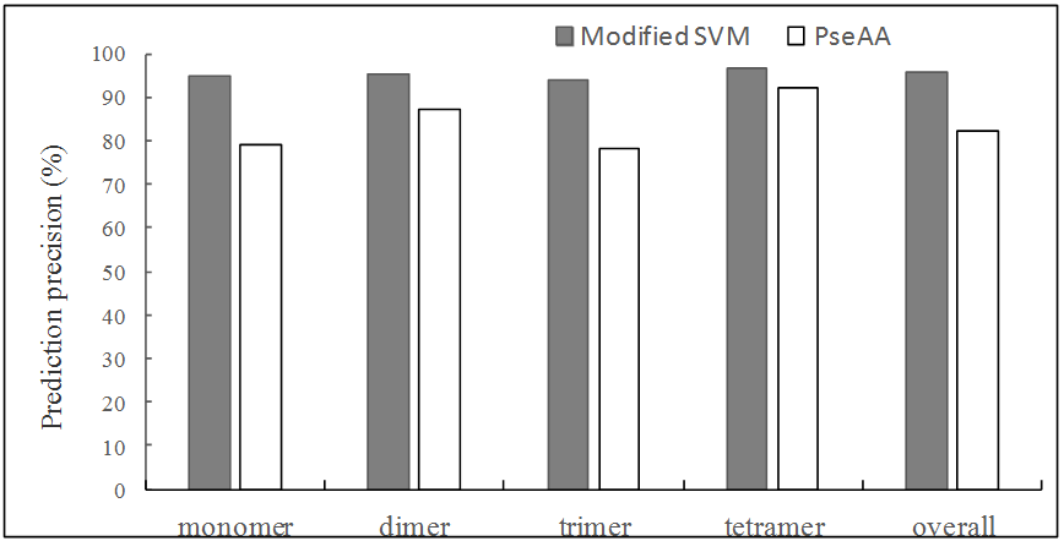

Fig. 3 Jackknife test

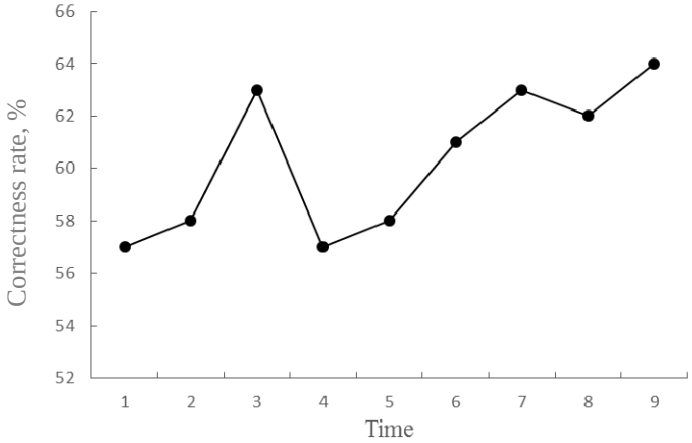

Fig. 4 Correctness rate of classification experiments

Table 1 shows the results comparison. The classification accuracy of the modified algorithm has been improved to a certain extent. Compared with the traditional method, the total classification precision has increased slightly. From Table 2 we can see the effectiveness of the modified algorithm. The access indicators have increased to some extent. Table 3 shows the results of the independent testing test with different methods. The indexes of *Q*, *TPR*, *FPR*, and *MCC* have improved to a different extent.

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 success rate of the modified SVM. From Table 6, we can conclude that the prediction of the modified SVM has a higher success rate. It shows that the modified SVM provides obviously superior protein prediction.

Table 1. Results comparison featuring extraction methods

	Traditional method		Modified SVM	
	Amino acid	Dipeptide	Amino acid	Dipeptide
(*C*, g)	(4,2)	(8,0.0625)	(4,3)	(3,0.02378)
Q	78.10	82.06	82.21	85.26
TPR	77.45	80.51	80.13	83.78
FPR	20.86	15.41	18.92	17.92
MCC	55.39	63.63	64.17	69.34

Table 2. Results comparison of methods with different parameters

	3 physicochemical parameters	4 physicochemical parameters	9 physicochemical parameters
(*C*, g)	(8,0.5)	(8,0.5)	(8,0.0625)
TP	800	816	836
FN	114	98	89
TN	523	513	542
FP	202	212	215
Q	80.72	81.09	82.3
TPR	79.84	79.38	83.14
FPR	17.90	16.04	18.05
MCC	60.79	61.67	67.74

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

	Pseudo amino acid composition	Linear combinatorial forecast	Modified SVM
(*C*, g)	(8,2)	(2,0.0313)	(6,0.0625)
TP	160	165	174
FN	16	9	8
TN	122	123	121
FP	23	22	26
Q	88.11	90.55	90.26
TPR	91.26	95.08	96.51
FPR	15.86	15.17	15.53
MCC	75.85	80.93	80.55

Table 4. Training and test data set

	Sample	
	Training set	Test set
Type I	435	478
Type II	152	180
Multipass	1311	1867
Lipid chain anchored	51	14
GPI anchored	110	86
Total	2059	2625

Table 5. Prediction results for different test methods using modified SVM

	Successful prediction rate (%)					
	Type I	**Type II**	**Multipass**	**Lipid chain anchored**	**GPI anchored**	**Total**
Self-consistency	81.5	70.2	95.3	55.1	67.4	88.3
Independent dataset	79.2	67.6	93.2	35.4	59.3	87.1
Jackknife	67.9	66.6	95.1	15.2	63.3	83.4

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	62.8	66.7	62.1
The minimum Euclidean distance	63.5	69.2	62.8
ProtLock	66.6	63.8	65.5
Covariance discriminant	81.1	79.4	76.4
Modified SVM	88.4	87.8	83.5

Conclusion

Biological information resources have been growing rapidly with the development of biological science and technology, and the development of computer technology and Internet has made large scale data storage, processing and transmission possible. Machine learning is often used to learn from experience and get useful information. It can be applied to biological information to get fruitful results.

In this paper, machine learning techniques are used in bioinformatics to achieve better protein prediction. Protein type prediction is a kind of important basic research. The SVM algorithm is used in the research to develop a new prediction algorithm. This method is combinatorial, and it is applied in protein prediction. We use two data sets to verify the efficiency of the modified algorithm, and the results show that the algorithm has a higher efficiency. The modified algorithm can be effectively used in protein prediction.

Acknowledgements

This work is supported by Shandong Science and Technology Development Program (2012GGX10106).

References

1. Bagchi S., P. G. Poonacha (2014). An Object Recognition Algorithm Using Maximum Margin Correlation Filter and Support Vector Machine, Proceedings of 20th National Conference on Communications, 1-6, doi: 10.1109/NCC.2014.6811272.
2. Chen K. E., Y. Jiang, L. I. Du, L. Kurgan (2009). Prediction of Integral Membrane Protein Type by Collocated Hydrophobic Amino Acid Pairs, Journal of Computational Chemistry, 30(1), 163-172.
3. Chen K., L. Kurgan, J. Ruan (2007). Prediction of Protein Structural Class Using PSI-BLAST Profile Based Collocation of Amino Acid Pairs, Proceedings of 1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2007, 17-20.
4. Choi D., S. M. Sharma, S. Pasadhika, Z. Kang, C. A. Harrington, J. R. Smith, S. R. Planck, J. T. Rosenbaum (2009). Application of Biostatistics and Bioinformatics

Tools to Identify Putative Transcription Factor-gene Regulatory Network of Ankylosing Spondylitis and Sarcoidosis, Communications in Statistics – Theory and Methods, 38(18), 3326-3338.
5. Garian R. (2001). Prediction of Quaternary Structure from Primary Structure, Bioinformatics, 17(6), 551-556.
6. Guo Z., Z. Zhang, Y. Pan, Z. Huang, G. Feng, L. He (2004). Prediction of Membrane Protein Types by Using Support Vector Machine, Journal of Shanghai Jiao Tong University, 38(5), 806-809.
7. Haq S. R., M. C. Jürgens, C. N. Chi, C.-S. Koh, L. Elfström, M. Selmer, S. Gianni, P. Jemth (2010). The Plastic Energy Landscape of Protein Folding: A Triangular Folding Mechanism with an Equilibrium Intermediate for a Small Protein Domain, Journal of Biological Chemistry, 285(23), 18051-18059.
8. Imer O., L. Çavas (2009). The Bioinformatics Tools for the Estimation of Disordered Regions in Proteins, Proceedings of 14th National Biomedical Engineering Meeting, BIYOMUT 2009, 1-4, doi: 10.1109/BIYOMUT.2009.5130377.
9. Kahsay R. Y., G. R. Gao, L. Liao (2005). Discriminating Transmembrane Proteins from Signal Peptides using SVM-Fisher Approach, Proceedings of 4th International Conference on Machine Learning and Applications, ICMLA 2005, 151-155.
10. Khan H. N., M. H. Khan, H. Rashid (2009). Structure Prediction and Analysis of Mouse Amiloride-sensitive Cation Channel 2, Neuronal Using Bioinformatics Tools, International Conference on Emerging Technologies, ICET 2009, 168-172.
11. Kim S., J. Kim, B.-T. Zhang (2009). Ensembled Support Vector Machines for Human Papillomavirus Risk Type Prediction from Protein Secondary Structures, Computers in Biology and Medicine, 39(2), 187-193.
12. Mostafa M. M., N. Billor (2009). Recognition of Western Style Musical Genres Using Machine Learning Techniques, Expert Systems with Applications, 36(8), 11378-11389.
13. Oliveira L. C., A. Schug, J. N. Onuchic (2008). Geometrical Features of the Protein Folding Mechanism are a Robust Property of the Energy Landscape: A Detailed Investigation of Several Reduced Models, Journal of Physical Chemistry B, 112(19), 6131-6136.
14. Rekepalli B., A. Vose, P. Giblock (2012). HSPp-BLAST: Highly Scalable Parallel PSI-BLAST for Very Large-scale Sequence Searches, Proceedings of 4th International Conference on Bioinformatics and Computational Biology 2012, BICoB 2012, 37-42.
15. Ruan H., R. C. De Lamare (2014). Robust Adaptive Beamforming Using a Low-complexity Shrinkage-based Mismatch Estimation Algorithm, IEEE Signal Processing Letters, 21(1), 60-64.
16. Shachak A., S. Fine (2008). The Effect of Training on Biologists Acceptance of Bioinformatics Tools: A Field Experiment, Journal of the American Society for Information Science and Technology, 59(5), 719-730.
17. Short E., D. Feil-Seifer, M. Mataric (2011). A Comparison of Machine Learning Techniques for Modeling Human-robot Interaction with Children with Autism, Proceedings of 6th ACM/IEEE International Conference on Human-Robot Interaction, 251-252.
18. Spears J. L., X. Xiao, C. K. Hall, P. F. Agris (2014). Amino Acid Signature Enables Proteins to Recognize Modified tRNA, Biochemistry, 53(7), 1125-1133.
19. Štambuk N., P. Konjevoda (2001). New Computational Algorithm for the Prediction of Protein Folding Types, International Journal of Quantum Chemistry, 84(1), 13-22.
20. Startek M., S. Lasota, M. Sykulski, A. Bulak, L. Noe, G. Kucherov, A. Gambin (2012). Efficient Alternatives to PSI-BLAST, Bulletin of the Polish Academy of Sciences: Technical Sciences, 60(3), 495-505.

21. Wan G., J. Wang, H. Cao, S. Wang, L. Wang (2013). Efficient Algorithm for Mismatching Point Elimination with Interzone Limited Pre-testing, Journal of Xidian University, 40(5), 107-112+199, doi: 10.3969/j.issn.1001-2400.2013.05.017.
22. Wang M.-H., C.-H. Li, W.-Z. Chen, C.-X. Wang (2007). PSPhos: PK-specific Phosphorylation Site Prediction Using Profile SVM, Journal of Computational and Theoretical Nanoscience, 4(7-8), 1324-1329.
23. Yang X. L., H. Y. He (2014). Variable Selection in Proteomic Profile Classification by Interval Support Vector Machines (iSVM), Applied Mechanics and Materials, 556-562, 347-350.
24. Yepes-Calderon F., F. Pedregosa, B. Thirion, Y. Wang, N. Lepore (2014). Automatic Pathology Classification Using a Single Feature Machine Learning – Support Vector Machines, Proceedings SPIE 9035, Medical Imaging 2014: Computer-aided Diagnosis, 903524 (24 March 2014), doi: 10.1117/12.2043943.
25. Zhou S., G.-B. Li, L.-Y. Huang, H.-Z. Xie, Y.-L. Zhao, Y.-Z. Chen, L.-L. Li, S.-Y. Yang (2014). A Prediction Model of Drug-induced Ototoxicity Developed by an Optimal Support Vector Machine (SVM) Method, Computers in Biology and Medicine, 51, 122-127.

Prof. Chengduan Wang
E-mail: wangchengduan@163.com

Chengduan Wang received a M.Sc. degree in Computer Applications from Shandong Science and Technology University (2007). He is currently a Professor and Dean of the School of Computer Engineering at Weifang University, China. He has published 5 papers and over 5 books in professional fields. His scientific interests are in the field of intelligent computing and software engineering.