2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Using Chou's Amphiphilic Pseudo-amino Acid Composition and Extreme Learning Machine for Prediction of Protein-protein Interactions

Qiao-Ying Huang
School of Computer Science and Technology
Harbin Institute of Technology
Shenzhen Graduate School
Shenzhen, China
Email: charwinghuang@gmail.com

Zhu-Hong You
College of Computer Science
and Software Engineering
Shenzhen University
Shenzhen, 518060, China
Email: zhuhongyou@gmail.com

Shuai Li
Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong

Zexuan Zhu
College of Computer Science
and Software Engineering
Shenzhen University
Shenzhen, China

***Abstract*—Protein-protein interactions (PPIs) play crucial roles in the execution of various cellular processes. Almost every cellular process relies on transient or permanent physical bindings of proteins. Unfortunately, the experimental methods for identifying PPIs are both time-consuming and expensive. Therefore, it is important to develop computational approaches for predicting PPIs. In this study, a novel approach is presented to predict PPIs using only the information of protein sequences. This method is developed based on learning algorithm-Extreme Learning Machine (ELM) combined with the concept of Chous Pseudo-Amino Acid Composition (PseAAC) composition. PseAAC is a combination of a set of discrete sequence correlation factors and the 20 components of the conventional amino acid composition, so this method can observe a remarkable improvement in prediction quality. ELM classifier is selected as prediction engine, which is a kind of accurate and fast-learning innovative classification method based on the random generation of the input-to-hidden-units weights followed by the resolution of the linear equations to obtain the hidden-to-output weights. When performed on the PPIs data of Saccharomyces cerevisiae, the proposed method achieved 79.66% prediction accuracy with 79.16% sensitivity at the precision of 79.96%. Extensive experiments are performed to compare our method with state-of-the-art techniques Support Vector Machine (SVM). Achieved results show that the proposed approach is very promising for predicting PPIs, and it can be a helpful supplement for PPIs prediction.**

***Index Terms*—Extreme Learning Machine(ELM); Pseudo-amino Acid Composition; Protein-protein Interactions; Support Vector Machine (SVM)**

I. Introduction

ProteinCprotein interactions (PPIs) are crucial for almost all of functions in the cell because they regulate a variety of cellular processes, including metabolic cycles, DNA transcription and replication, different signaling cascades, and many additional processes. In the past decades, many innovative techniques for detecting PPIs have been developed. Due to the progress in large-scale experimental technologies such as yeast two-hybrid (Y2H) screens [1], tandem affinity purification (TAP) [2], mass spectrometric protein complex identification (MS-PCI) [3] and other high-throughput biological techniques for proteinCprotein interaction detection, a large amount of PPIs data for different species has been accumulated [4]. This provides a rich data source for further investigations. However, the experimental methods are costly and time-consuming. Therefore, current PPIs pairs obtained from experiments only cover a small fraction of the complete PPIs networks [5][6]. Hence, it is of great practical significance to develop the reliable computational methods to facilitate the identification of PPIs [7][8][9].

A number of computational methods have been proposed for the prediction of PPIs based on different data types, including protein domain, phylogenetic profiles, gene neighborhood, gene fusion and literature mining knowledge etc.. There are also methods that combine interaction information from several different data sources [10]. However, these methods cannot be implemented if such pre-knowledge about the proteins is not available. Recently, many researchers have engaged in the development of sequences-based method for predicting new PPIs, and the experiment results showed that the information of protein sequences alone is sufficient to predict PPIs [11].

Among them, one of the excellent works is a SVM-based method developed by Shen et al. [12]. In the study, the 20 amino acids were clustered into seven classes according to their dipoles and volumes of the side chains, and then the conjoint triad method abstracts the features of protein pairs based on the classification of amino acids. When applied to predict human PPIs, this method yields a high prediction accuracy. Because the conjoint triad method cannot takes neighboring effect into account and the interactions usually occur in the discontinuous amino acids segments in the sequence, on the other work Guo et al. developed a method based on SVM and auto covariance to extract the interactions information in the discontinuous amino acids segments in the sequence [5]. Their method yielded a prediction accuracy of 86.55%, when applied to predicting Saccharomyces cerevisiae PPIs. In our previous works, we also obtained good prediction performance by using auto correlation descriptors and correlation coefficient, respectively [13].

All the above research use the machine learning method to learn the rules from PPIs and furthermore to predict novel interactions. One key issue in machine learning is to extract features from protein sequence. Feature extraction methods are essential because it helps to build the prediction model and

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

improve the prediction quality. But many feature extraction methods are only based on the amino acid composition in which no sequence order effect was taken into account. To improve the prediction performance, it is necessary to incorporate such an effect. The Chous pseudo amino acid composition (PseAAC) is one of the most widely used feature extractors for proteins. The PseAAC is adopted to represent the protein samples. The advantage of using PseAAC composition approach is that it allows us to deal with such a complicated problem with more than 20 discrete factors without completely losing the sequence-order and sequence-length effects like the case treated by the conventional amino acid composition. The first 20 factors reflect the effect of amino acid composition, whereas the additional factors reflect the effect of sequence order. Here, we will predict PPIs with a standard feature extraction method that is based on the Chous pseudo amino acid composition.

Among all the machine learning techniques, neural network is very useful and popular in solving the PPIs problems. We have used neural network to explore many issues [14][15][16]. Meanwhile, the extreme learning machine (ELM), firstly proposed by Guang-Bin Huang [17], is an effective learning algorithm for single-hidden-layer feed-forward neural networks (SLFNs). Some classical learning algorithm in neural network, e.g. Back Propagation, requires setting several user-defined parameters and easily goes into local minimum. However, ELM randomly chooses the input weights and bias, it only requires setting the number of hidden neurons and the activation function. In theory, ELM tends to provide the best generalization performance at extreme fast learning speed [18][19].

In this study, a new predictor combining the concept of PseAAC and ELM system is proposed for predicting PPIs. To evaluate the performance, the proposed method was applied to Saccharomyces cerevisiae and Helicobacter pylori datasets. The results obtained by ELM prediction model with PseAAC are quite promising. The experiment results show that our method achieved 79.66% prediction accuracy with 79.16% sensitivity at the precision of 79.96%. It demonstrates that the ELMs can be a powerful tool to predict PPIs.

II. MATERIALS AND METHODS

A. Data Set

The PPIs dataset was collected from publicly available S.cerevisiaecore subset of interacting proteins (DIP) database [5]. First, we removed the protein pairs which length are less than 50 residues or have more than 40% sequence identity, in order to reduce the data scale. Finally, we obtained 5594 protein pairs as positive data set. The selection of negative data set (the non-interacting protein pairs) is essential to the final prediction results. There is an assumption that proteins occupying different subcellular localizations do not interact. According to this principle, we constructed 5594 protein pairs as negative set. The final data set combines the positive data set and the negative data set, which contains 11188 protein pairs.

B. The Pseudo-Amino Acid Composition

As illustrated above, one of the difficulties for discovering new PPIs is to find a way fully encode the information of proteins. In this paper, each protein sequence is represented by pseudo-amino acid composition (PseAAC)[20][21]. The essence of PseAAC is, on the one hand, to include the main feature of amino acid composition, but on the other, to include information beyond amino acid composition. Type 1 PseAAC composition is also called the parallel-correlation type and generates $20+\lambda$ discrete numbers to represent a protein [22].The basic idea of pseudo-amino acid composition is as following:

Let $H_1^0(i)$, $H_2^0(i)$, $M^0(i)\,(i=1\cdots 20)$ be the original hydrophobicity values [23], the original hydrophilicity values [10] and the original side-chain masses of the 20 natural amino acids, respectively. They are converted to following qualities by a standard conversion:

$$H_1(i)=\frac{H_1^0(i)-\sum_{i=1}^{20}\frac{H_1^0(i)}{20}}{\sqrt{\frac{\sum_{i=1}^{20}\left[H_1^0(i)-\sum_{i=1}^{20}\frac{H_1^0(i)}{20}\right]^2}{20}}}$$

$$H_2(i)=\frac{H_2^0(i)-\sum_{i=1}^{20}\frac{H_2^0(i)}{20}}{\sqrt{\frac{\sum_{i=1}^{20}\left[H_2^0(i)-\sum_{i=1}^{20}\frac{H_2^0(i)}{20}\right]^2}{20}}}$$

$$M(i)=\frac{M^0(i)-\sum_{i=1}^{20}\frac{M^0(i)}{20}}{\sqrt{\frac{\sum_{i=1}^{20}\left[M^0(i)-\sum_{i=1}^{20}\frac{M^0(i)}{20}\right]^2}{20}}} \quad (1)$$

Then, a correlation function can be defines as:

$$\Theta(R_i,R_j)=\tfrac{1}{3}\left\{[H_1(R_j)-H_1(R_i)]^2+[H_1(R_j)-H_1(R_i)]^2+[M(R_j)-M(R_i)]^2\right\} \quad (2)$$

$H_1(R_i)$,$H_2(R_i)$,$M(R_i)$ is the standard conversion of the amino acid R_i calculated by equation 1.This correlation

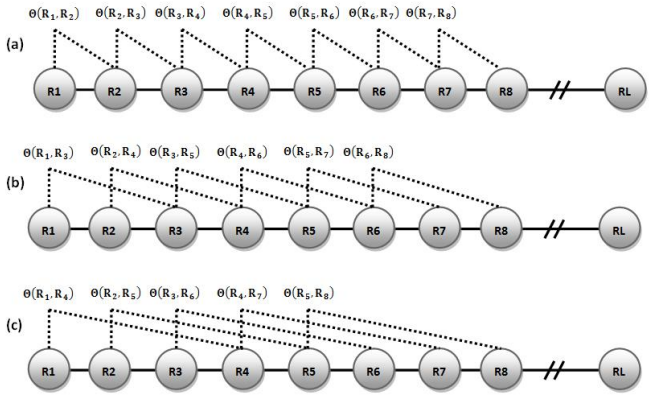

Fig. 1. A schematic drawing to show (a) the first-tier, (b) the second-tier, and (3) the third-tier sequence order correlation mode along a protein sequence. Panel (a) reflects the correlation mode between all the most contiguous residues, panel (b) that between all the second-most contiguous residues, and panel (c) that between all the third-most contiguous residues.

function is actually an averaged value for the three amino acid properties: hydrophobicity value, hydrophilicity value and side-chain mass.

As we can see from Figure 1, the sequence order effect of a protein can be, to some extent, reflected through a set of sequence-correlation factors $\theta_1, \theta_2, \cdots, \theta_\lambda$ as defined below:

$$\theta_1 = \frac{1}{L-1}\sum_{i=1}^{L-1}\Theta\left(R_i, R_{i+1}\right)$$

$$\theta_2 = \frac{1}{L-2}\sum_{i=1}^{L-2}\Theta\left(R_i, R_{i+2}\right)\cdots$$

$$\theta_\lambda = \frac{1}{L-\lambda}\sum_{i=1}^{L-\lambda}\Theta\left(R_i, R_{i+\lambda}\right), \lambda < L \tag{3}$$

where θ_1 is called the first-tier correlation factor that reflects the sequence order correlation between all the most contiguous residues along a protein chain (Fig. 1a), θ_2 is the second-tier correlation factor, and so on. The correlation function Θ is given by equation 2.

The sample of a protein X should be defined in a $(20+\lambda)-D$ vector, as formulated below:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_{20+\lambda} \end{bmatrix} \tag{4}$$

$$x_u = \begin{cases} \frac{f_u}{\sum_{i=1}^{20} f_i + w\sum_{j=1}^{\lambda}\theta_j}, (1 \le u \le 20) \\ \frac{w\theta_{u-20}}{\sum_{i=1}^{20} f_i + w\sum_{j=1}^{\lambda}\theta_j}, (20+1 \le u \le 20+\lambda) \end{cases} \tag{5}$$

where f_1 is the normalized occurrence frequency of the 20 amino acids in the protein X. θ_j is called the j-tier correlation factor computed according to equations 1-3 for the protein X. And w is the weight factor for the sequence order effect and it is set as 0.05 according to Chou [22]. $(\lambda < L)$ is a parameter to be chosen. In the study, it is found by preliminary tests that the optimal value for λ is 30. Given a protein, 20+30=50 PseAAC components are generated. The two interacted proteins should combine together, thus each protein pair is coded by a vector with 50+50 dimensions.

C. ELM Optimization and Evaluation of Performance

In the study, the classification model for predicting PPIs was based on Extreme Learning Machine (ELM). ELM is a new learning method for single-hidden layer feed-forward networks (SLFNs). Compared with traditional learning algorithms, ELM consumes little training time while achieves higher accuracy [17][24]. If an SLFN with L hidden nodes can approximate these N samples $(x_i, t_i) \in R^n \times R^m$ with zero error, then we have

$$\sum_{i=1}^{L}\beta_i G\left(a_i, x_j, b_i\right) = t_j, j = 1, \cdots, N \tag{6}$$

where (a_i, x_j) is the parameter of the ith hidden node and β_i is the output weight linking the ith hidden node to the output node. ELM works for wide spread of hidden nodes. Equation 6 can be compactly as

$$H\beta = T \tag{7}$$

where

$$H = \begin{bmatrix} G\left(a_1, x_1, b_1\right) & \cdots & G\left(a_L, x_1, b_L\right) \\ \vdots & \ddots & \vdots \\ G\left(a_1, x_N, b_1\right) & \cdots & G\left(a_L, x_N, b_L\right) \end{bmatrix}_{N\times L} \tag{8}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L\times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N\times m} \tag{9}$$

While computing, $\hat{\beta} = \mathbf{H}^+ T$ is used as the estimated value of β, where $\mathbf{H}^+$ is the Moore-Penrose generalized inverse [25] of the hidden layer output matrix H. The original algorithm of ELM proposed by Huang et al.[24] contains three steps:

ELM Algorithm Given a training set $P = \left\{(x_i, t_i)_{i=1}^N\right\}$ and hidden node number L,

1) Assign random hidden nodes by randomly generating hidden node parameters $(a_i, b_i), i = 1, \cdots, L$.

2) Calculate the hidden layer output matrix H.

3) Calculate the output weight $\hat{\beta} = \mathbf{H}^+ T$.

As analyzed above, compared to traditional SLFNs, ELM does not need to adjust the value of a and b in the training process, and a global optimal solution will be obtained. It significantly improves the training speed. For parameter selection, we only need to select the number of hidden neurons. Different number of hidden neurons will lead to obvious difference in prediction performance. The prediction performance was evaluated by the overall prediction accuracy (ACC), sensitivity (SN), precision (PE) and Mathews correlation coefficient (MCC):

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \tag{10}$$

$$SN = \frac{TP}{TP+FN} \tag{11}$$

$$PE = \frac{TP}{TP+FP} \tag{12}$$

$$MCC = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FN)\times(TN+FP)\times(TP+FP)\times(TN+FN)}} \tag{13}$$

where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively. We also used the receiver operating characteristic (ROC) curve to assess the prediction performance. An ROC curve is a graphical plot of the true positive rate (TPR) versus the false positive rate (FPR) for a binary classifier system as its discrimination threshold is varied. The area under an ROC curve is called AUC, which ranges from 0 to 1.The larger AUC is, the better predictor is.

III. RESULTS AND DISCUSSION

A. Assessment of Prediction Ability

The number of hidden neurons is very important to ELM for effective learning, we should find out the most suitable one in the experiment. The number of hidden neurons is initialized as 500 and it is increased by 50. Different activation functions (Sin, Sigmoidal, Hardlim, Triangular basis and Radial basis) were chosen to evaluate the performance of ELM. During the experiment, we found out that the prediction error of Hardlim function and Triangular basis function is very large, so these two activation functions were excluded. The results of the experiments are shown in Fig. 2. As we can see, with the number of hidden neurons increases, the test accuracy increases as well. The test accuracy all increase significantly in the beginning, then they become almost stable. When the number of hidden neurons is 1250, the test accuracy reaches highest of Sin function. We finally choose the Sin function as activation function and set the number of hidden neurons as 1250.

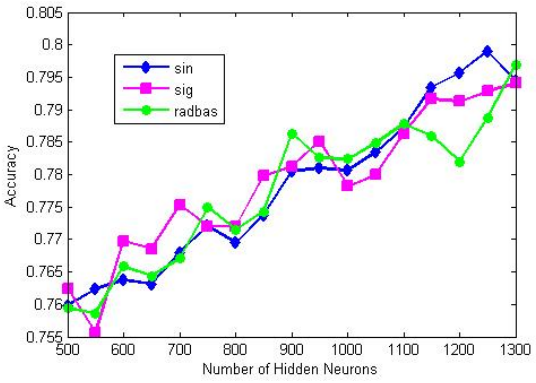

Fig. 2. Corresponding test accuracy of Sin, Sigmoidal and Radial basis

After setting the best number of neurons, we used 5-fold cross-validation to investigate the data set, five models were constructed. The prediction results are shown in Table 1. As we can see, the precisions are $\geq$79.22%, the sensitivities are $\geq$77.81%, and the prediction accuracies are $\geq$78.68%. On average, our method yields a PPI prediction model with an accuracy of 79.66$\pm$0.84%. We also calculated the MCC and AUC values to evaluate the ability of our model. From Table 1, the average value of MCC and AUC is 67.60% and 87.42%. The standard deviation of sensitivity, precision, accuracy, MCC and AUC are 0.90, 1.08, 0.84, 1.00 and 0.50% respectively. In order to determine the extreme learning speed of ELM, we also take train time in account. The average of test time is 19.88 seconds. It is very fast compares to other methods.

B. Performance of Independent Data Set

Our method gained a good performance from the PPIs data of S.cerevisia. We next tested our algorithm with an independent dataset. In H.pylori, the data set contained 1365 interaction pairs, out of which that which contained a protein with $<$50 amino acids and those with noises were all excluded. We constructed our final model using the whole data set (1365*2=2730 protein pairs) with the optimal parameters (λ=30, the number of hidden neurons=250).The prediction results are shown in Table 2. From Table 2, our model obtains the best accuracy is 69.63%.

TABLE I
PREDICTION RESULTS OF THE TEST SETS

Test Set	SN(%)	PE(%)	ACC(%)	MCC(%)	AUC(%)	TIME(s)
1	78.68	80.31	79.39	67.27	87.60	20.47
2	77.81	79.29	78.68	66.44	87.29	19.47
3	79.87	79.22	79.93	67.90	87.30	19.44
4	79.55	79.27	79.35	67.44	86.78	20.47
5	79.89	81.71	80.94	69.14	88.14	19.56
Average	79.16$\pm$0.90	79.96$\pm$1.08	79.66$\pm$0.84	67.60$\pm$1.00	87.42$\pm$0.50	19.88

TABLE II
PREDICTION RESULTS ON H.PYLORI DATA SET

Feature Extraction	SN(%)	PE(%)	ACC(%)	TIME(s)
Auto Covariance	73.47	74.64	74.24	21.88
Moran Autocorrelation	73.30	74.63	74.19	21.86
Geary Autocorrelation	72.68	74.57	73.95	21.96
Conjoint Triad	65.49	67.03	66.63	23.09
Pseudo-Amino Acid Composition	79.16	79.96	79.66	19.88

C. Comparison with Other Methods

By comparing with different methods, we analyzed the prediction ability of the ELM prediction model using pseudo-amino acid composition. First of all, we constructed different ELM prediction model using other feature extraction, such as Auto Cavariance, Moran Autocorrelation, Geary Autocorrelation and Conjoint Triad. These four feature extraction methods are popular in PPIs prediction. Then we used 5-fold cross-validation to train and test the data. From Table 3, we can see that the model based on ELM with pseudo-amino acid composition gives good result of average sensitivity, precision and accuracy of 79.16, 79.96 and 79.66%, respectively. Moreover, we found that the training time of our prediction model is much less than others. Because the dimension of the protein extracted from PseAAC is much less than other feature extraction methods.

TABLE III
COMPARISON OF SVM AND ELM

Method	TIME(s)		SN(%)	PE(%)	ACC(%)	No of SVS /Neurons
	Traning	Testing				
SVM	129.73	4.32	76.37	76.05	76.16	5719
ELM	19.88	0.275	79.16	79.96	79.66	1250

We also constructed a SVM prediction model using pseudo-amino acid composition to compare with the prediction ability with ELM. SVM is a traditional machine learning method in PPIs prediction. Here, two parameters, C and g were set as 32 and 32. Results are shown as Table 4. ELM with much less number of hidden neurons has a better learning performance with SVM (ELM uses 1250 hidden neurons and SVM produces 5719 support vector), the accuracy difference

of these two algorithms is about 3.5%; the sensitivity, precision are about 2.79 and 3.91% difference.

As illustrated above, ELM performs much better than SVM. Furthermore, for the learning speed, ELM consumes 19.88s in the learning process because it does not need to adjust the input weights and the hidden neurons biases of the network. However, because SVM with Radial basis as its kernel function consumes lots of time for parameters adjustment, the training process totally consumes 129.73s which is about 6.5 times more than ELM model. It has shown the obvious advantages of ELM in training time. The learning speed of ELM is 16 times of SVM for testing samples. That means, after trained and deployed the ELM may react to new observations much faster than SVM in such real application. Therefore, ELM has practical values in predicting protein-protein interactions.

IV. Conclusion

In this paper, we developed a novel method for predicting PPIs only using the primary sequences of protein. The prediction model was constructed based on ELM and PseAAC. PseAAC was used to take the sequences order effects into account and results in the improvement of predictive accuracy, and then an ELM algorithm was employed to construct the predictor. ELM provides better generalization performance and faster speeds than other popular learning algorithm. Moreover, ELM model consumes much less time than SVM for unknown samples which shows the greatest advantage of ELM. When performed on the PPIs data of Saccharomyces cerevisiae, the proposed method achieved 79.66% prediction accuracy with 79.16% sensitivity at the precision of 79.96%. Given the complex nature of PPIs, the performance of our method is promising and it can be a helpful supplementary for PPIs prediction.

Acknowledgment

This work is supported in part by the National Science Foundation of China, under Grants 61102119, 61373086, 61272339, in part by the China Postdoctoral Science Foundation, under Grant 2012M520929.

References

[1] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proceedings of the National Academy of Sciences*, vol. 98, no. 8, pp. 4569–4574, 2001.

[2] A.-C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat *et al.*, "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, pp. 141–147, 2002.

[3] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier *et al.*, "Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry," *Nature*, vol. 415, no. 6868, pp. 180–183, 2002.

[4] L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. Hao, C. Ooi, B. Godwin, E. Vitols *et al.*, "A protein interaction map of drosophila melanogaster," *Science*, vol. 302, no. 5651, pp. 1727–1736, 2003.

[5] Y. Guo, L. Yu, Z. Wen, and M. Li, "Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences," *Nucleic acids research*, vol. 36, no. 9, pp. 3025–3030, 2008.

[6] Z.-H. You, Z. Yin, K. Han, D.-S. Huang, and X. Zhou, "A semi-supervised learning approach to predict synthetic genetic interactions by combining functional and topological properties of functional gene network," *Bmc Bioinformatics*, vol. 11, no. 1, p. 343, 2010.

[7] Y.-K. Lei, Z.-H. You, T. Dong, Y.-X. Jiang, and J.-A. Yang, "Increasing reliability of protein interactome by fast manifold embedding," *Pattern Recognition Letters*, vol. 34, no. 4, pp. 372–379, 2013.

[8] L. Zhu, Z.-H. You, D.-S. Huang, and B. Wang, "t-lse: A novel robust geometric approach for modeling protein-protein interaction networks," *PloS one*, vol. 8, no. 4, p. e58368, 2013.

[9] Y.-K. Lei, Z.-H. You, Z. Ji, L. Zhu, and D.-S. Huang, "Assessing and predicting protein interactions by combining manifold embedding with multiple information integration," *Bmc Bioinformatics*, vol. 13, no. Suppl 7, p. S3, 2012.

[10] Y. Qi, J. Klein-Seetharaman, and Z. Bar-Joseph, "Random forest similarity for protein-protein interaction prediction from multiple sources." in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 2004, pp. 531–542.

[11] L. Yang, J.-F. Xia, and J. Gui, "Prediction of protein-protein interactions from protein sequence using local descriptors," *Protein and peptide letters*, vol. 17, no. 9, pp. 1085–1090, 2010.

[12] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, Y. Li, and H. Jiang, "Predicting protein–protein interactions based only on sequences information," *Proceedings of the National Academy of Sciences*, vol. 104, no. 11, pp. 4337–4341, 2007.

[13] M.-G. Shi, J.-F. Xia, X.-L. Li, and D.-S. Huang, "Predicting protein–protein interactions from sequence using correlation coefficient and high-quality interaction dataset," *Amino Acids*, vol. 38, no. 3, pp. 891–899, 2010.

[14] S. Li, M.-H. Meng, W. Chen, Y. Li, Z. You, Y. Zhou, L. Sun, H. Liang, K. Jiang, and Q. Guo, "Sp-nn: A novel neural network approach for path planning," in *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*. IEEE, 2007, pp. 1355–1360.

[15] S. Li, H. Cui, Y. Li, B. Liu, and Y. Lou, "Decentralized control of collaborative redundant manipulators with partial command coverage via locally connected recurrent neural networks," *Neural Computing and Applications*, vol. 23, no. 3-4, pp. 1051–1060, 2013.

[16] S. Li and Y. Li, "Nonlinearly activated neural network for solving time-varying complex sylvester equation," 2013.

[17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. IEEE, 2004, pp. 985–990.

[18] Z.-H. You, Y.-K. Lei, L. Zhu, J. Xia, and B. Wang, "Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis," *BMC bioinformatics*, vol. 14, no. Suppl 8, p. S10, 2013.

[19] Z.-H. You, Y.-K. Lei, J. Gui, D.-S. Huang, and X. Zhou, "Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data," *Bioinformatics*, vol. 26, no. 21, pp. 2744–2751, 2010.

[20] Y.-D. Cai and G.-P. Zhou, "Predicting protein-protein interactions with pseudo amino acid composition," in *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, vol. 1. IEEE, 2008, pp. 158–163.

[21] C. Chen, Y.-X. Tian, X.-Y. Zou, P.-X. Cai, and J.-Y. Mo, "Using pseudo-amino acid composition and support vector machine to predict protein structural class," *Journal of Theoretical Biology*, vol. 243, no. 3, pp. 444–448, 2006.

[22] K.-C. Chou, "Prediction of protein cellular attributes using pseudo-amino acid composition," *Proteins: Structure, Function, and Bioinformatics*, vol. 43, no. 3, pp. 246–255, 2001.

[23] J. Damborskỳ, "Quantitative structure-function and structure-stability relationships of purposely modified proteins." *Protein Engineering*, vol. 11, no. 1, pp. 21–30, 1998.

[24] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[25] C. R. Johnson, *Matrix theory and applications*. American Mathematical Soc., 1990, vol. 40.