



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 9 • Number 45 • 2016

Protein Secondary Structure Prediction Using Feed Forward Artificial Neural Network and Perceptron

Sarneet Kaur^a and Babita Pandey^b

^aResearch Scholar, Department of Computer Sc and Engg, Lovely Professional University, Jalandhar, Punjab, India

E-mail: sarneetkaur@live.com

^bDepartment of Computer Applications, Lovely Professional University, Jalandhar, Punjab, India

E-mail: shukla_babita@yahoo.co.in

Abstract : Protein secondary structure prediction is a very important and challenging task. In this paper, two neural networks *i.e.* Feed Forward and Perceptron Neural Network had been used to predict secondary structure of protein and their performances had been compared based on experiment data. We had used Q3 (alpha-helix, beta-strands and irregular turns/loops) to evaluate the accuracy of secondary structure. Feed forward ANNs achieved a Q3 accuracy of 50% and perceptron achieved Q3 accuracy of 83.33%.

Keywords: Protein Secondary Structure, Feed Forward, Perceptron.

1. INTRODUCTION

Protein structure prediction is necessary to know the condition of any disease and for drug design. There are four types of protein structures: Primary, Secondary, Tertiary and Quaternary. The tertiary structure helps to predict the functions of protein but it is difficult to predict the tertiary structure. With the help of secondary structure the prediction of tertiary structure becomes easy. Secondary structure of protein can be predicted by its primary sequence. Artificial neural network (ANN) is very useful in prediction of secondary structures and used by many researchers^{1,2,3,4,5,6,7,8,9,10,11,12}. But the major problem for prediction with ANN is the large size of the input to ANN which increases the likelihood of convergence problems for the training algorithm adopted, the memory requirement and the processing time also¹³. Therefore we need an encoding scheme which produces small size inputs in support of ANN used for secondary structure prediction.

In this paper, we have proposed a new method to encode the amino acid sequence. This method reduces the size of input and output data set of ANN for secondary structure prediction.

2. ARTIFICIAL NEURAL NETWORK

Artificial neural networks are usually set of interconnected neurons which work together to solve a specific problem as shown in Figure 1. As human beings learn by example, similarly artificial neural network based on the experience adjust the numeric weights on the connections and are capable of learning.

2.1. Feed Forward Network

It consists of number of layers connected with each other. Each connection is having a specific weight attach to it. Once the data enters in to layer 1 it moves forward to next layer and so on. There is no feedback between the layers.

2.2. Backpropagation Algorithm

Backpropagation algorithm works with layered feed forward ANNs. Backpropagation algorithm provides the network with sets of inputs and outputs, and then the errors are calculated. Until the neural network learns the training data, backpropagation algorithm reduces the error. The weights of the connection are chosen randomly at the start and then the goal is to change them so that the error will be negligible¹³.

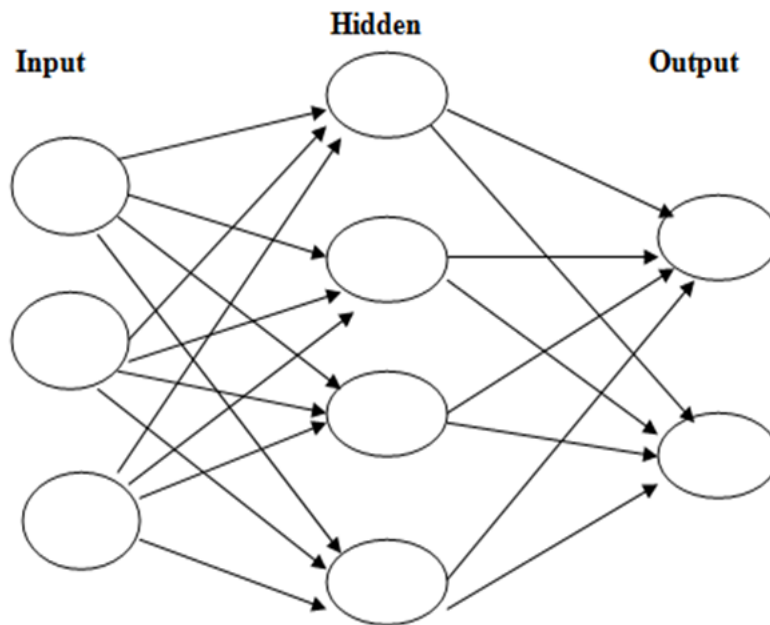


Figure 1: Artificial Neural Network

2.3. Perceptron

Perceptron has an ability to learn and differentiate data sets. Perceptron learn to classify any linear separable set of inputs. Perceptron is capable to categorize data into two classes. It contains a single neuron with a number of adjustable weights as shown in Figure 2.

It has three essential elements:

1. A number of connecting links called synapses; every link carry a weight w_0, w_1, w_2 .
2. Once they are multiplied by their particular weights, summation (or adder) sums the input signals.
3. Output of the neuron is limited due to the activation function $f(x)$. Usually the output is restricted to the interval $[0,1]$ or otherwise $[-1,1]$.

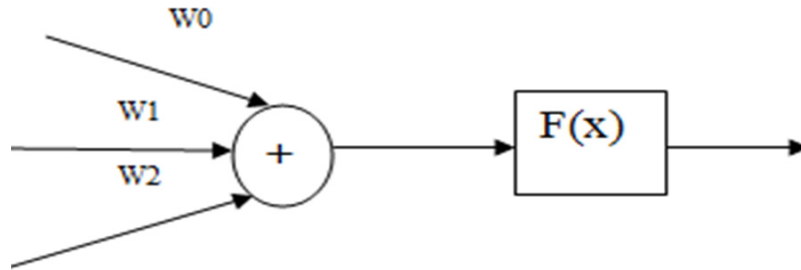


Figure 2: Perceptron

3. RELATED WORK

¹ has deployed the Feed forward neural network to predict protein secondary structure¹. ² has deployed the Multi-Modal Neural Networks and achieved an average accuracy of 66%, which is about 6.9% higher than single neural network. ³ has achieved a sustained performance of about 78% correct prediction. Multimodal neural network (MNN) had been deployed by ⁴ and 66% of prediction accuracy was achieved. Segmented-memory recurrent neural network (BSMRNN) has been deployed by ⁵. Two stage neural networks had been deployed by ⁶.

Method based on breaking down the single first level NN classifier, into three separate ones, had been deployed for each of the secondary structure elements (SSE) types. They also introduce the use of sparsely connected feed-forward NNs, instead of the classic fully interconnected one. This network architecture gains considerable speed improvements⁷. Neural network had been deployed and concluded that contact number can also be used as a rich source of information for improvement of secondary structure prediction⁸. Artificial neural network had been deployed and the prediction of secondary structure gives good results⁹. Pattern Recognition Neural Network had been deployed and achieved a Q8 accuracy of 72.3%¹⁰. Multiple Artificial Neural Network Classifier had been deployed¹¹. Multilayer Feed-forward Neural Networks had been deployed. It has better generation ability & higher accuracy rate¹².

4. IMPLEMENTATION

The work is implemented in a Matlab 7.10.0, using nn tool. The complete implementation of our method is shown in figure 3. Protein Data Bank (PDB) database sequences are used for training. Sequences in a database consist of 22 amino acids such as: M K R R I R R E R N ...K L. These sequences are taken as input in perceptron, the input layer of perceptron encodes a moving window in the amino acid sequence and prediction is made for the central residue in the window. A preliminary window size of 3 is used for each sequence.

An integer encoding scheme is used for network input. In this scheme each amino acid at each window position is encoded by an integer value, one for each possible amino acid type at that position. A integer value ranges from 1-26 is assigned to each amino acid based on their alphabetical increasing order as shown in Table 1. Each sequence is suffixed and prefixed by #, which is assigned an integer value 0. Therefore the first window consist of # and two amino acid (#MK) and the last window consist of two amino acid and # (KL#) as shown in second and last column of first row (Headed by Inputs) respectively as shown in Table 2. For each window one output is assigned which consists of H (Helix), E (Strand) & C (Coil) as shown in last row(Headed by Outputs) of Table 2.

Table 1
Assigning integer value to each sequence

<i>Amino acid</i>	<i>Integer value</i>	<i>Amino acid</i>	<i>Integer value</i>
A	1	F	6
R	19	P	16
N	14	S	20
D	4	T	21
C	3	W	23
Q	17	Y	25
E	5	V	22
G	7		
H	8		
I	9		
L	12		
K	11		
M	13		

Table 2
Training data using 3 window sizes

Inputs	#	M	K	R	R	I	R	R	E	...	K
	M	K	R	R	I	R	R	E	R	...	L
	K	R	R	I	R	R	E	R	N	...	#
Outputs	C	C	C	C	H	H	H	H	H	...	E

Secondary structure consist of 8 sub-structures according to DSSP code such as Turns (T), Alpha Helix (H), 3/10 Helix (G), Coil (C), pi Helix (I), Bridge (B), beta ladder and beta sheet (E) and Bend (S). Integer values are assigned to these secondary sub-structures as shown in Table 3.

Table 3
Secondary sub-structure and their integer value

<i>Sub-Structure</i>	<i>Integer Value</i>
G	1
H	2
I	3
E	4
B	5
T	6
S	7
C	8

Artificial neural network is trained with input and output set which is normalized by formula given in equation 1:

$$\text{Normalized } (e_i) = \frac{e_i - E_{\min}}{E_{\max} - E_{\min}} \quad (1)$$

The normalized set is shown in Table 4.

Table 4
Normalized training data

0	0.2857	0	1	1	0	1	1	0	0.9166
1	0	1	1	0	1	1	0	1	1
0.8462	1	1	0	1	1	0	1	0.6923	0
1	1	1	1	0	0	0	0	0 0.3333

The network used consists of an input layer and an output layer (Figure 3). An initial window size of 3 is used. We divided our input data into two set training & testing. For training 60% & for testing 40% of data had been taken. The parameters of ANN using feed forward neural network and perceptron neural network is shown in Table 5. There performance had been compared in Table 6.

Table 5
Some feed forward neural network, perceptron neural network parameters and their value

Parameters	Value	Value
Adaptive Learning Function	LEARNP	LEARNGDM
Performance Function	MAE	MSE
Transfer Function	HARDLIM	PURELIN
Layer1	32 neurons	32 neurons

Table 6
Performance comparison of two neural networks

Method	Q3
Feed Forward Neural Network	50%
Perceptron Neural Network	83.33%

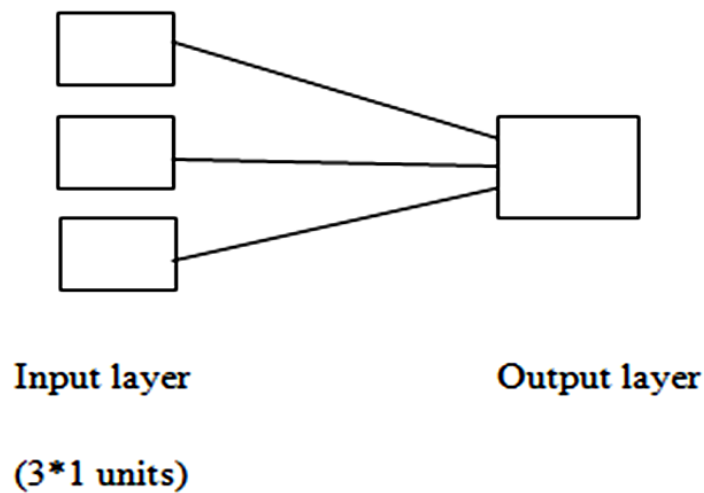


Figure 3: Neural network topology. Each of the 3 blocks shown in the input layer represents network inputs used to encode the amino acid at the corresponding window position

Each of the 3 blocks shown in the input layer represents network inputs used to encode the amino acid at the corresponding window position.

The accuracy of prediction of secondary structure using Q3 (alpha-helix, beta-strands and irregular turns/ loops) formula as given in equation 2.

$$Q_3 = \frac{\text{No. of correctly classified secondary structures}}{\text{Total no. of amino acid residues}} * 100\% \quad (2)$$

After implementing feed forward neural network we got Q3 (alpha-helix, beta-strands and irregular turns/ loops) accuracy, calculated as follows:

$$\begin{aligned} Q_3 &= 12/24 * 100 \\ &= 50\% \end{aligned}$$

After implementing perceptron neural network we got Q3 (alpha-helix, beta-strands and irregular turns/ loops) accuracy, calculated as follows:

$$\begin{aligned} Q_3 &= 20/24 * 100 \\ &= 83.33\% \end{aligned}$$

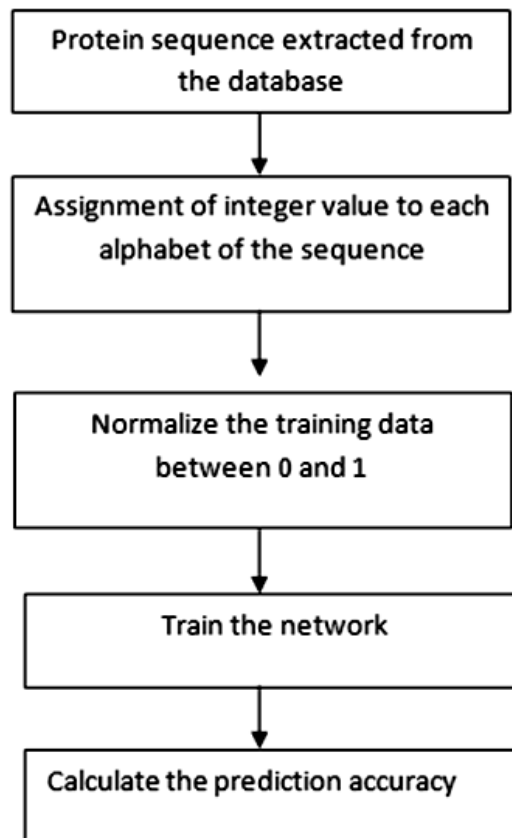


Figure 4: Implementation of our proposed method

5. CONCLUSION

We had extracted some sequences from the database, developed a way to encode the sequences, normalize the training data and trained the perceptron neural network, feed forward neural network with the training data using Matlab.

After training the results had been noted and the maximum accuracy we got is of perceptron neural network *i.e.* 83.33%. Whereas with feed forward neural network we got an accuracy of 50%. We had found that perceptron neural network is proficient in solving the classification problem. In future we will try to make use of the physico-chemical properties of the amino acids.

REFERENCES

- [1] Barlow T.W. Feed-forward neural networks for secondary structure prediction, *Journal of Molecular Graphics*, 1995; 3: pp.175-183.
- [2] Hanxi ZHU, Lkuo Y., Kunihito Y. Prediction of Protein Secondary Structure by Multi-Modal Neural Networks, *IEEE International joint conference on Neural Networks*, 2002.
- [3] Gianluca P., Darisz P., Burkhard R., Pierre B. Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles, *PROTEINS: Structure, Function, and Genetics*, 2002; 47: pp 228–235. WILEY-LISS, INC.
- [4] Nakayama K., Hirano A., Fukumura K.-i., On Generalization of Multilayer Neural Network Applied to Predicting Protein Secondary Structure, *IEEE International joint conference on Neural Networks*, 2004. .
- [5] Chen J., and Chaudhari N.S., Capturing Long-Term Dependencies for Protein Secondary Structure Prediction, *Advances in Neural Networks - ISNN 2004, International Symposium on Neural Networks, Part II*, volume 3174 of *Lecture Notes in Computer Science*
- [6] Kakumani R., Devabhaktuni V., Ahmad M. O. A Two-Stage Neural Network Based Technique for Protein Secondary Structure Prediction, *30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008*.
- [7] Dzikovska V., Oreskovic M., Kalajdziski M., Trivodaliev K., Dacev D. Protein Secondary Structure Prediction Method based on Neural Networks, *IEEE 2nd International Conference on Bioinformatics and Biomedical Engineering*, 2008
- [8] Lakizadeh A., Marashi S. A., Addition of contact number information can improve Protein secondary structure prediction by neural Networks, *EXCLI Journal* 2009; 8: pp.66-7.
- [9] Thalatham M.N.V., Rao P. V., Varma K.V.S.R.P., Murty N.V.R., Apparao A. Prediction of Protein Secondary Structure using Artificial Neural Network, *International Journal on Computer Science and Engineering*, 2010; 2(5): pp.1615-1621.
- [10] Rao P.V.N., Devi T.U., Kaladhar D.S.V.G.K., Sridhar G.R., Rao A.A., Protein Secondary Structure Prediction using Pattern Recognition Neural Network, *International Journal of Engineering Science and Technology*, 2010; 2(6): pp.1752-1757.
- [11] Bordoloi H., Sarma K.K. Protein Structure Prediction Using Multiple Artificial Neural Network Classifier, as a Chapter of a volume titled *Soft Computing Techniques in Vision Science, Studies in Computational Intelligence*, 2012; 395/2012: pp. 137-146.
- [12] Jian-wei L.I.U., Guang-hui C.H.I., Hai-en L.I., Yuan L.I.U., Xiong-lin L.U.O. Prediction of Protein Secondary Structure Using Multilayer Feed-forward Neural Networks, *25th IEEE conference on Chinese Control and Decision Conference (CCDC)*, 2013.
- [13] Muknahallipatna, S., Chowdhury, B.H. Input dimension reduction in neural network training-case study in transient stability assessment of large systems. *Intelligent Systems Applications to Power Systems*, 1996. Proceedings, ISAP '96.
- [14] Gershenson C. *Artificial Neural Networks for Beginners* (2003).