International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

Predicting the Secondary Structure of Protein using the Back-Propagation with MADALINE

Ketan Priyam Khare
IMS Engineering College, Ghaziabad
Vidhi Rastogi
IMS Engineering College, Ghaziabad
Shivani Agarwal
IMS Engineering College, Ghaziabad

ABSTRACT

Protein, being a basic unit of a biological structure is an organic molecule made up of the amino acid sequences. Basically, protein can be distinguished on the basis of its structure and the structures are primary, secondary, tertiary and quaternary. Secondary Structure, being a very complex biological structure of protein consists of the long protein chains; there are regions in which the chains are organized into alpha-helices, beta-pleated sheets and coils. The prediction of secondary structure of the protein will help in the recognition of the actual symptoms of any disease and thus, in return predicting secondary structure can be marked as beneficial in drug designing. Predicting secondary structure of proteins can be done using various soft computing techniques like the Artificial Neural Network, Hidden Markov Model, and Support Vector Machine etc. In this paper, we proposed the model using the feed-forward- network with the back propagation method and also, by making use of the sigmoid function give much better results along with the sliding window concept, in which size of window can be varied as per the process. Prediction can be done by training the network well so that the error in prediction can be reduced and high accuracy can be achieved.

Keywords

Artificial Neural Network, Backpropagation, Drug Designing, Sliding window

INTRODUCTION

For the computational molecular biology the prediction of secondary structure of protein have become a matter of great calculations as it has lead to a holy grail of molecular biology too.

The major and foremost task is the identification of the secondary structure of protein that would in turn help to assist in diagnosing diseases, deficiencies. While coming on to the secondary structure it is necessary to understand the Protein structure.

Protein Structures are generally made up of the amino acid sequences. These amino acid sequence leads to the sub-structures of α-helices, β-sheets and random coils as a result of hydrogen bonds. Proteins are also recognized by 3D arrangements of these amino acids. The determination of the protein structure will help in better understanding of the protein traits and functions.

There are generally 4 types of protein structures and they are primary, secondary, tertiary and quaternary. The primary structure of protein can be described as the linear sequences of 20 amino acids, the structural unit proteins, joined together by peptide bonds.

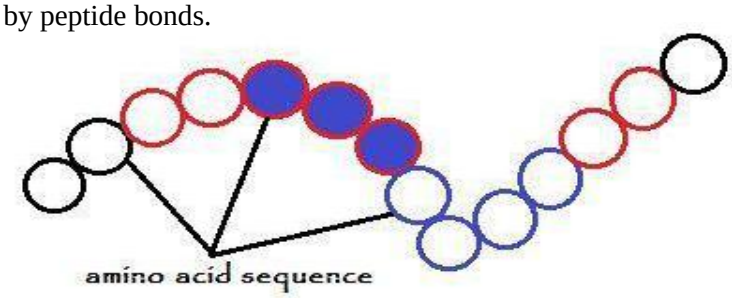

Figure1: Primary Structure of protein

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

The secondary structure is further formed as the result of folding and twisting of the primary structure sequence. There is also a super secondary structure which is the arrangement of alpha helices and/or beta strands into discrete folding units. There are generally three classes which are used to classify the secondary structure and they are, alpha helix (H), beta sheets (E) and coil (C).

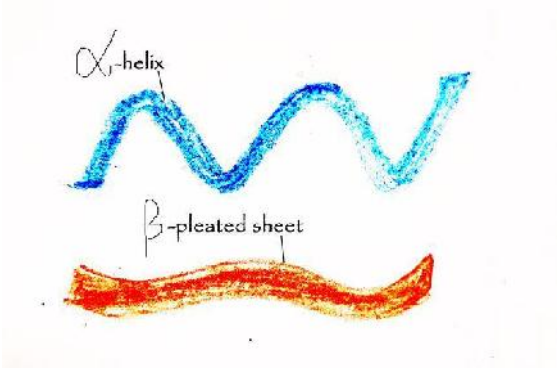

Figure 2: Secondary Structure of protein

The third structure is the three dimensional structure of the polypeptide chain known as the tertiary structure.

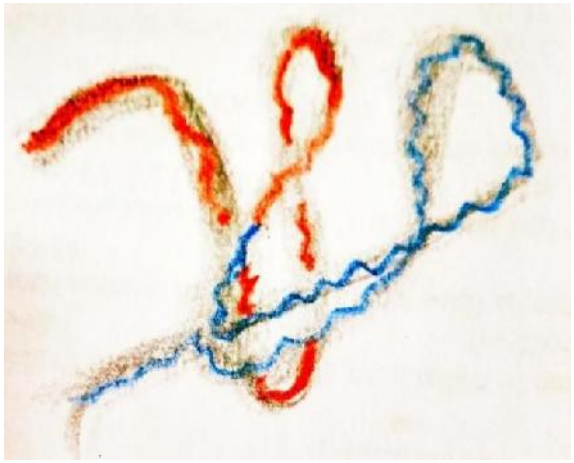

Figure 3: Tertiary Structure of protein

The next and the last structure formed of protein is the quaternary structure that is formed by more folding and the twisting of polypeptide chain. This is the arrangement of separate molecules, such as in protein – protein pr protein –nucleic acid interactions.

Figure 4: Quaternary Structure of protein

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

BIOLOGICAL NEURAL NETWORK

As the term Biological neural network suggests, the network is a series of interconnected neurons that helps in passing the information to the brain. Here, brain cells play a major role as they help in the prediction and to remember. The biological neural network works on the electrochemical process. Basically, the biological neural network consists of the receptors, neural network and the effectors. The receptor receives the information either implicitly or explicitly. The neural network formed by the interconnected neurons tasks to pass signals and accordingly detect the targeted output. All the neurons are connected to each other through dendrites. The information is first passed through synapses and further by using Axon. At the end the response is made to the external world with the help of effectors.

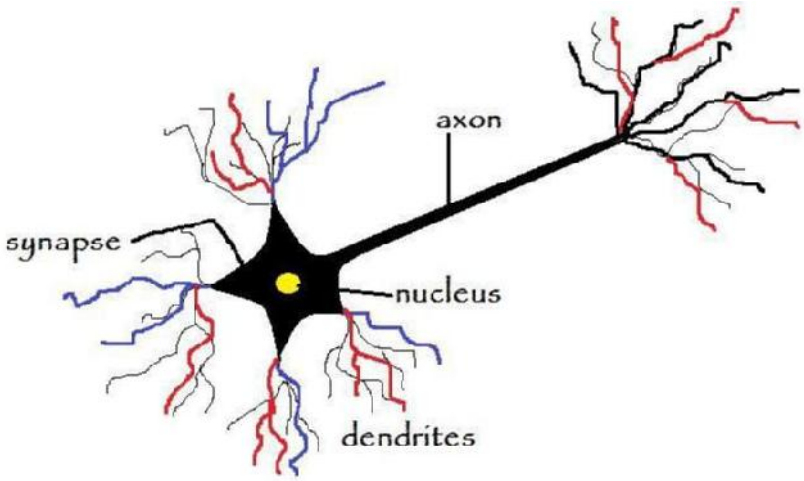

Figure 5: Biological neuron

ARTIFICIAL NEURAL NETWORK

An Artificial neural network can be termed as an information-processing model that is designed in such a manner so that it can replicate the most basic functions of the brain. Basically, in the artificial neural network there are 3 different layers - Input layer, hidden layer and output layer. As shown in the figure 6 the input layer acts as the dendrites. The input layers here are associated with certain weights. The values of the input layer are propagated through one or more hidden layers to an output layer with the help of an activation function.

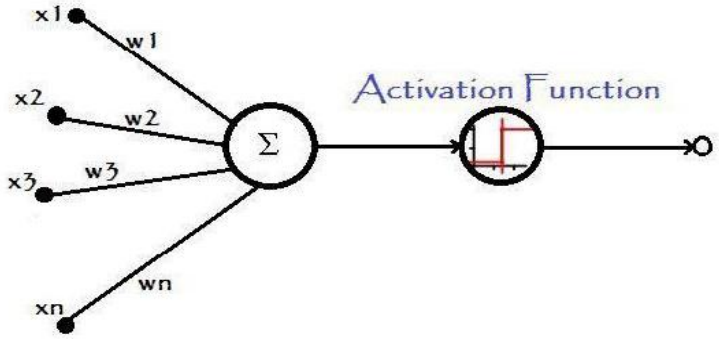

Figure 6: Artificial neural network

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

DRUG DESIGNING

When it comes to drug designing, we should be able to know about the symptoms of the disease particularly and correspondingly we should be able to target the protein sequence so that the drug can cure the disease as much faster as can.

PROPOSED WORK

Dataset: Dataset of proteins with both primary as well as respective secondary structure is used which is provided by the RCSB Protein Data Bank.

ALGORITHM

Step -1

Input and the output sequence is loaded from the database.

Input Sequence

A	Z	R	I	S	T	B	E	Y	Z	H	X

Output Sequence

H	H	H	H	E	E	C	C	T	T	C	C

Weight matrices (v and w) and bias (vo and wo) are defined. 'α' is defined as learning rate.

Input Sequences are encoded. As there are 21 possible characters in the input so each character of the sequence is represented using 21 nodes and the corresponding node position is encoded as 1 and else nodes are kept 0.

A	M	K	C	F	....	G	W	S	X	Y

A:

1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0									

M:

0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0									

K:

0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0									

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

C:

1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0									

Y:

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1									

Rest of the characters are encoded respectively to their positions.

Step-2

Input Layer receives the input in X(i=1 to n) as per the value of sliding window (n). Neurons at hidden layer, p.

Step-3

At hidden layer, Z_{in_j} (j=1 to p) is calculated as:

$$Z_{in_j} = v_{0j} + \sum_{i=1}^{n} x_i \cdot v_{ij}$$

Output is:

$$Z_j = f(Z_{in_j})$$

Where $f(x) = 1/(1 + e^{-x})$

Step-4

Output of hidden layer is treated as input for input for the output layer. Y_{in_k} (k=1) is calculated as:

$$Y_{in_k} = w_{0k} + \sum_{j=1}^{p} Z_j \cdot w_{jk}$$

Output is:

$$Y_k = f(Y_{in_k})$$

Step-5

The output value is compared with the target value if both are same then we go for next inputs by shifting the sliding window forward. Otherwise we use back propagation algorithm to train the network by updating the weights and the bias.

Step-6

Error correction factor is computed with the help of target value (t_k) (k=1):

$\delta_k = (t_k - Y_k) \cdot f'(Y_{ink})$

Where $f'(Y_{ink}) = Y_k(1 - Y_k)$

Step-7

Change in weights and bias is:

$\Delta w_{jk} = \alpha . \delta_k . Z_j$

$\Delta w_{0k} = \alpha . \delta_k$

Also, δ_k is remit to hidden layer.

Step-8

For each hidden unit (j=1 to p):

$\delta inj = \delta k . wjk$

Error correction factor is calculated as:

$\delta_j = \delta_{inj} . f'(Z_{inj})$

Step-9

Change in weights and bias is:

$\Delta v_{ij} = \alpha . \delta_j . x_i$

$\Delta v_{0j} = \alpha . \delta_j$

Step-10

All the weights and bias are updated as:

At output unit (Y_k , $k = 1$) :

$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$

$w_{0k}(new) = w_{0k}(old) + \Delta w_{0k}$

At hidden unit (Z_j , $j = 1$ *to* p):

$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$

$v_{0j}(new) = v_{0j}(old) + \Delta v_{0j}$

Step-11

Stopping condition is being checked i.e. if the majority of weights does not gets changed then training is stopped.

RESULTS

Predicting the secondary structure using the artificial neural network results into the accuracy of 70% while using the back propagation learning algorithm.

REFRENCES

[1] Prediction of Protein Secondary Structure by S.N. Vel Arjunan, Safaai Deris, Rosli Md Illias.
[2] Design and Implementation of an Algorithm to Predict Structure of proteins using Artificial Neural Network by Shivani Agarwal, Arushi Baboota, Deepali Mendiratta.
[3] Reviewing the Methods of Predicting Protein Secondary Structure by Shivani Agarwal, Rishabh Kaushik, Atul Kumar
[4] Stephen R. Holbrook, Steven M. Muskal and Sung-Hou Kim,"Predicting Protein Structural Features With Artificial Neural Networks". ARTIFICIAL INTELLIGENCE & MOLECULAR BIOLOGY.
[5] L. HOWARD HOLLEY AND MARTIN KARPLUS, "Protein secondary structure prediction with a neural network" Department of Chemistry, Harvard University, Cambridge, MA 02138, Contributed by Martin Karplus, October 5, 1988.
[6] To Improve the Performance Of Secondary Structure Prediction by Soft Computing Technique by Shivani Agarwal, Sakshi Gupta, Shivangi Gupta.
[7] A Comparative Study of the Protein Secondary Structure Prediction methods by Shivani Agarwal,Arushi Baboota, Atul Kumar