Protein Protein Interactions Techniques, Challenges, and Its Applications: Review

B. L. Pal[1], Akshay Deepak[2], Arvind Kumar Tiwari[3]
[1]*Dept. of CSE, KNIT Sultanpur U.P., India*
[2]*Dept of CSE, NIT Patna, Bihar, India*
[3]*Dept. of CSE, KNIT Sultanpur U.P., India*

Keywords: Protein-protein interactions, Deep learning, challenges

Abstract: Protein-protein interactions provide a vital role in both the biological and cellular functional processes of all organisms. PPIs aim to establish parasitic pathogens, and bacterial viral in the harbors to minimize the causes of disease. PPIs are utilized to of the human host that have great potential for medicinal development identify specific diseases with associated interfaces for human interaction. Protein-protein interactions provide a vital role in both the biological and cellular functional processes of all organisms. PPIs aim to establish parasitic pathogens and bacterial viral in the harbors of the human host that have great potential for medicinal development to minimize the causes of disease. PPIs are utilized to identify specific diseases with associated interfaces for human interaction.

1 INTRODUCTION

Proteins are the main content used in framework of all living species .(Shatnawi, *et.al*, 2015) In the primary structure of the protein, twenty different kinds of amino acids are merged. An appropriated figuring engineering docking site, the Map The decrease strategy (Sun, T. *et.al*, 2017) is appeared. An auto encoder a counterfeit neural framework executing an unaided learning framework a calculation that reasons the capacity to extend to create cover structures from unlabeled information (Zaki, N. *et.al*, 2009).. In spite of the same structure, all amino acids have different R groups. Every R groups is connected by the carbon atom, i.e. alpha carbon. The secondary structure of the protein follows a 3-dimensional structure. The alpha (α) helices, and beta (β) sheets are the most commonly used structures of the protein. The α-helix follows right-handed spiral array whereas, β consist of crosswise using more than one hydrogen bonds.

To do a specific task, a protein interacts with other proteins. The physical interactions between at least two protein molecules are protein-protein interactions (PPIs). The accurate PPI for an organism is very useful because one or more protein-protein interactions are involved in most biological processes (Xenarios, I. *et.al*, 2001)– (Li, H. *et.al*, 2012). Furthermore, defects in PPIs will affect the actions and function of cells that lead to many diseases, such as neurodegenerative, cancers, and etc. The study of solving key biological issues in proteomics by various computational techniques is called computational proteomics. These biological issues can be protein identification, protein-protein structure prediction, functional classification of protein-protein structure, protein interactions, quantitative analysis, drug design, and etc. In literature, there are various prediction techniques exist for PPIs. Therefore, in many protein research areas, the formation of accurate and effective methodology for the identification of PPIs has very important implications.

In section 2, we elaborate few related works on PPI. Section 3 illustrates various techniques of PPI prediction and section 4 provides the limitations of techniques used in PPI prediction. Applications of PPI predictions are discussed in section 5, and the conclusion of this survey is presented in section 6.

Pal, B., Deepak, A. and Tiwari, A.
Protein Protein Interactions Techniques, Challenges, and Its Applications: Review.
DOI: 10.5220/0010563900003161
In *Proceedings of the 3rd International Conference on Advanced Computing and Software Engineering (ICACSE 2021)*, pages 123-128
ISBN: 978-989-758-544-9

2 RELATED WORK

The various modeling methods of PPIs have been explored by Biologists on different platform, including In order to enhance protein prediction, An appropriated figuring engineering docking site, the Map The decrease strategy (Sun, T. *et.al,* 2017) is appeared. An auto encoder a counterfeit neural framework executing an unaided learning framework a calculation that reasons the capacity to extend to create cover structures from unlabeled information (Huang, Y. A . *et.al*, 2016). The remainder of the PPIs Information is collected via test strategies, along with yeast, Two hybrid (Y2H) screens, the purification of tandem affinity (TAP) and Complex ID (MS-PCI) of mass spectrometric proteins and other Large Elevation Throughput procedures to collect data from PPIs (Szilagyi, Andras, *et.al*, 2014). The most accurate structure X-ray crystallography and crystallography of protein complexes are given by NMR spectroscopy, but these are labor-intensive and time-consuming techniques [Huang, Y. A. *et.al*, 2016). The pVLASPD algorithm is designed to increase performance and Efficiency to deal with the problem on a wide PPI scale. In paper, (Hu, L., Yuan, *et.al*, 2017). prediction is performed using Deep Learning techniques and also highlight some previous work.

The P-P docking speed has two big sampling challenges (Umbrin, H. *et.al*, 2018). Flexibility and conformationalness. Docking methods must be used to the ability to filter through billions of possible items Configurations. Thus, various methods use FFT-based review as it is faster than Monte Carlo and based on geometrical fitting.

The conformational variations are large. For instance, those observed in computationally illusive, some influenced fit interactions remain. Fast and accurate scoring to identify a binding opportunity (an exhaustive testing algorithm) is required to complement the function. The scoring method should ideally measure the free energy system. About restricting. These figurings are hard to accomplish and none of the new estimations are the general stoichiometry of restricting accomplices is another part of discovering that necessities are improving, taking into account the developing number of multi-section edifices being demonstrated.

We need to choose, as such, the number of simultaneous restricting accomplices a given protein is probably going to have. The scoring plan ought to have the option to decide whether the perplexing's free energy with/without extra integrators will diminish. All in all, the environment of the cooperation could influence the force of restricting. In the docking protocol, a latest method allows the utilization of explicit water.

It is likewise an illustration of the advancements we have completed throughout the years in PPI docking. Generally, expectation problems are associated with interfaces that are made of more than 1 surface patch, or with adaptable interfaces. The overall precision with which the 3D structures of Formulated protein complexes have gradually improved over the years.

3 TECHNIQUES OF PROTEIN-PROTEIN PREDICTION

Deep Learning (Wei, L., Yang, *et.al*, 2005), a sub-field of Machine Learning , is focused on artificial neural networks, stressing the utlizing of multiple neural networks related layers to convert inputs into features suitable for equivalent outputs are predicted. Considering a sufficiently large dataset A training algorithm can be used to automate input-output pairs— Identify the mapping of outputs from inputs by considering a set of outputs parameters on each network layer. Although FFNN or similar elementary cells are the basic frameworks of a Deep Learning system in many instances, these are combined using different connectivity patterns into deep stacks. This architectural versatility enables the customization of Deep Learning models for any specific form of data. In general, deep learning models are trained by back-propagation on examples (Khotanzad, A., *et.al*, 1990), leading to successful internal data gained for a mission. This automated learning function effectively eliminates the need for manual feature engineering, potentially and laborious error-prone process that requires expert domain knowledge and is needed in more approaches to machine learning.

3.1 Convolutional Neural Networks

The structure of the Convolutionary Neural Network (CNN) (He, D. C., *et.al*, 1991) is built to process information that is structured with daily spatial information Dependency (like the series tokens or the pixels in a sequence Picture). By utilizing the same set of local convolutionary filters from various data, a CNN layer provides advantage of this regularity, thus bringing two gains: it escapes the over fitting problem by providing a very limited number of weights for tuning with respect to the various input layer and the dimensionality of the next layer, and it is translation invariant. Typically, a CNN module is collected from many consequential CNN layers because the nodes have used wider receptive fields at later layers. Furthermore, it can also be encoded in more complex features. It can be considered that the above-mentioned " windowed 'FFNN is used as a basic, shallow, version of CNN, although we will maintain . In this report, FFNN to suit the historical The practice of naming in literature.

3.2 Recurrent Neural Network

The continuous deepening of artificial neural network of research work various problems are hard to determine in many areas of pattern recognition, intelligent robots, automated control, and biology, have been successfully solved. Economics and medicine. The recurrent neuron network (Qian, S., *et.al*, 1993), is a sequence data modelling neural network. RNN is achieved exceptional success in natural language processing, recognition of speech, and , image recognition in recent years. The structures, i.e. in between layers is highly linked with conventional neural network model. And, within the layer, the neurons are not linked. For some problems, this form of neural network is effectiveless. A sequence's current output in RNNs is dependent by the outputs of previous steps. Specifically, the network learns the information of previous steps and applies it to the current performance measurement, i.e. linking nodes between hidden layers. The hidden layer input not only comes from the input layer output, but also contains the hidden layer output on the input layer. The preceding moment. Whereas, neurons are sequential in the secret layer of the RNN. In the biological information field, the potential of this technology has not been published yet, but its unique capacity provides the attention of biologists. Since in biological sequence data [Zhou Z. Learnware *et.al* 2016- Gregor, K., Danihelka *et.al* (2015), this clear front-to-back positional relationship also exists.

3.3 Long Short-Term Memory

If the gap between the relevant information and the expected location is less, the RNNs will l utilize the previous information, but with the time interval increases, the long-distance information can not be learned by ordinary RNNs. Long-term short-term memory (LSTM) neural network (Sainath, T. N., *et.al,* 2015- Dyer, C., Ballesteros, *et.al* 2015), To resolve this problem, it is suggested that long-term dependency can be taught. The primary distinction between the LSTM and other networks is the use of complex memory blocks rather than general neurons. The memory block, along with some memory cells, comprises three multiplicative "gate" units (input, forget, and output gates) (one or more). To control the information flow, the gate unit is used and the memory cell is utilized to control the information flow. Historical information should be preserved. The gate, i.e., removes or restores data to the state of the cell by regulating the flow of information. The input and output of the information, more precise, The input and output gates are handled by flow, respectively. The forgotten gate decides how much information is stored from the previous unit to the present unit (Sak, *et.al,* 2014. – Lazib , *et.al,* 2020).

3.4 Feed Forward Neural Networks

An ANN (Khotanzad, et.al, 1990) having no cycles, is a Feed Forward Neural Network (FFNN). In particular, layered FFNN is NN, the nodes of which can be categorized into various groups (layers) where the outputs of layer I are work as inputs to and only to layer i + 1. Then, the layer I is referred to as the input layer, and output layer referred as last, and every layer in between is a hidden layer whose nits make up an instance's intermediate representation. Layered FFNN, which can be trained using the back propagation algorithm from examples and which has been shown to have universal approximation properties (Zhang, F., et.al ,2009), In their alleged " windowed 'form, these organizations have for the most part been utilized, in which each portion of amino acids in a succession is utilized as the contribution for a different model, the objective for the section being the PSA of interest for one of the amino acids in the fragment (typically the focal one).

3.5 Support Vector Machine

Support Vector Machines (SVM) the one of the most advanced Algorithms, which actually have the benefits of good classification, Quality and solid potential for generalization. The basic principle of the support vector machine is the non-linearly that map into train data set. The aim of this non-linear mapping is to form the data set in the original spalinearly inseparable. An optimal hyperplane of separation with the greatest isolation distance Then, it is formed in the space of features, meaning In the input space, an optimal nonlinear decision boundary is created. The optimal SVM hyperplane of separation not only mitigates the empirical risk, but also minimises the error of generalization. (Shen et al., 2007) have proposed an SVM algorithm-based PPI prediction model, But before prediction is completed, this approach must consider the homology of proteins. In resolving this constraint, the Conjoint was suggested by them to Triad feature for amino acid description and choosed the SVM with a kernel function as a predictor for the prediction of protein interaction. Guo et al. (Guo, et al., 2008), have proposed a combined Auto Covariance code of PPI prediction method with SVM and radial base function.

4 CHALLENGES IN PROTEIN-PROTEIN PREDICTION

The PPI prediction computational method poses many challenges:

- Proteins are the combination of chemical and physical properties and various structural characteristics. The common problems of PPIs are the effective and precise extraction of features.
- Normally, the main features are rough. Effectively reducing the size and noise of the function, removing similar information will lead to improving model accuracy, reducing the model's computational complexity, and improving model interpretability. The noise reduction technology used for processing biological data, however, it has been not officially opened.
- Proteins have a number of physical and chemical properties and structural components characteristics. A common problem of these components are faced by PPIs, is the accurate and suitable extraction of features.
    - How to find or extract an accurate and appropriate prediction algorithm that can make full use of present knowledge and construct an efficient model to decrease the PPI prediction error.
    - Most of the previous models of PPI prediction are focused on balanced sets of data. But practical datasets of PPIs are always unbalanced, which contributes to "preference" training for a predictor.
    - Some Deep Learning algorithms, when implemented, are easy to overfit or trap in local optimization.

5 APPLICATIONS

PPIs are important for the creation of enzymatic complexes and macromolecular design. Due to their high specificity, PPIs have emerged in recent years as promising targets for appropriate drug design, which may enable researchers to target specific disease-related diseases. Two sorts of exploratory strategies that uncover the components of organic macromolecules of a few kinds of techniques of PPIs are utilized for enormous scope screening, and numerous others are utilized to specific circumstance PPIs, for example, high throughput techniques, for example, the two mixture arrangement of yeast. Individual techniques like X - ray crystallography, spectroscopy of nuclear magnetic resonance (NMR) and cry electron microscopy are utilized. There are certain drawbacks of these experimental procedure. Post translational modification, due to different physicochemical problems such as transient dynamics (PTM). Necessities to precisely perceive PPIs and PPI locales in silico ways to deal with broaden PPI inclusion and channel out bogus positives dependent on certainty scores of protein connections. Categorize the prediction approach based on the sequence of features below, In normal medication plan, area of interest expectation and docking, structure, homology, areas, useful comparability, quality co articulation and organization geography and their potential applications. PPIs have discovered application as medication targets, including single buildups, and the very much described PPI intuitive comprises of various underlying gatherings with different estimations of medication capacity. There is the pharmaceutical industry PPIs are resistant to be used for drug detection. It is hard to measure enzymatic activity with PPIs. The specifics of molecular level interaction of the PPI interfaces are important. Essential for small molecule modulator detection. With the availability of silico structures corresponding to different states, the structural information of the interfaces can be identified.

A good application for deep learning needs a very large amount of data (thousands of images) to train the model to process the data easily, as well as GPUs or graphics processing units. By performing transfer learning or highlight extraction, pre-prepared profound neural organization models can be utilized to rapidly apply profound figuring out how to your issues. AlexNet, VGG-16, VGG-19, and Caffe models imported using import Caffe Network, are some of the models available.

In fields, for example, computer vision, machine vision, prediction, preparation of common language, noise recognization, interpersonal organization separating, machine interpretation, bioinformatics, drug plan, clinical image recognization, content review and assessment, deep learning structures, for example, deep neural networks, profound conviction organizations, intermittent neural organizations and convolutionary neural organizations have been actualized.

In biological systems, artificial neural networks (ANNs) have been motivated by information processing and distributed communication nodes. ANNs are different from biological brains with different variations. In particular, neural networks tend to be static and symbolic, whereas most living organisms have dynamic (plastic) and similar biological brains.

6 CONCLUSION

The various predefined methods show how the protein structure and PPIs are coordinated by a number of levels. These strategies not just permit us to build up how a pathogenic protein interfaces on an atomic scale with its host, yet in addition how such collaborations work in a bigger cell organization. Machine (AI) and deep learning strategies are utilized to anticipate high confirmation associations by joining proper arrangements of negative and positive preparing sets. Here, we have reviewed all purposed applications, issues, and techniques of protein protein interactions and we will solve the challenge by utilizing the machine learning and deep learning technique to predict combination of protein protein interactions of based on learning data.

REFERECES

Guo, Y., Yu, L., Wen, Z., & Li, M. (2008). Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences. Nucleic acids research, 36(9), 3025-3030.

Gregor, K., Danihelka, I., Graves, A., Rezende, D., & Wierstra, D. (2015, June). Draw: A recurrent neural network for image generation. In International Conference on Machine Learning (pp. 1462-1471). PMLR.

He, D. C., & Wang, L. (1991). Texture features based on texture spectrum. Pattern recognition, 24(5), 391-399.

Hu, L., Yuan, X., Hu, P., & Chan, K. C. (2017). Efficiently predicting large-scale protein-protein interactions using MapReduce. Computational biology and chemistry, 69, 202-206.

Huang, Y. A., You, Z. H., Chen, X., Chan, K., & Luo, X. (2016). Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding. BMC bioinformatics, 17(1), 1-11.

Khotanzad, A., & Hong, Y. H. (1990). Invariant image recognition by Zernike moments. IEEE Transactions on pattern analysis and machine intelligence, 12(5), 489-497.

Lazib, L., Qin, B., Zhao, Y., Zhang, W., & Liu, T. (2020). A syntactic path-based hybrid neural network for negation scope detection. Frontiers of computer science, 14(1), 84-94.

Li, H., Tounkara, J. C., & Liu, C. (2012). Prediction of Protein-Protein Docking Sites Based on a Cloud-Computing Pipeline. International Journal of Machine Learning and Computing, 2(6), 798.

Li Z, Wang Y, Zhi T, Chen T. A survey of neural network accelerators. Frontiers of Computer Science, 2017, 11(5): 746–761

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In Eleventh annual conference of the international speech communication association.

Qian, S., & Chen, D. (1993). Discrete gabor transform. IEEE transactions on signal processing, 41(7), 2429-2438.

Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015, April). Convolutional, long short-term memory, fully connected deep neural networks. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4580-4584). IEEE.

Shatnawi, M. (2015). Review of recent protein-protein interaction techniques. Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology, 12(5), 99-121.

Sun, T., Zhou, B., Lai, L., & Pei, J. (2017). Sequence-based prediction of protein protein interaction using a deep-learning algorithm. BMC bioinformatics, 18(1), 1-8.

Sun, T., Zhou, B., Lai, L., & Pei, J. (2017). Sequence-based prediction of protein protein interaction using a deep-learning algorithm. BMC bioinformatics, 18(1), 1-8.

Szilagyi, A., & Zhang, Y. (2014). Template-based structure modeling of protein–protein interactions. *Current opinion in structural biology*, *24*, 10-23.

Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.

Shen, J., Zhang, J., Luo, X., Zhu, W., Yu, K., Chen, K., ... & Jiang, H. (2007). Predicting protein–protein interactions based only on sequences information. Proceedings of the National Academy of Sciences, 104(11), 4337-4341.

Umbrin, H., & Latif, S. (2018, March). A survey on Protein Protein Interactions (PPI) methods, databases, challenges and future directions. In 2018 International

Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.
Wan, K. K., Park, J., & Suh, J. K. (2002). Large scale statistical prediction of protein-protein interaction by potentially interacting domain (PID) pair. Genome Informatics, 13, 42-50.
Wei, L., Yang, Y., Nishikawa, R. M., Wernick, M. N., & Edwards, A. (2005). Relevance vector machine for automatic detection of clustered microcalcifications. IEEE transactions on medical imaging, 24(10), 1278-1285.
Xenarios, I., & Eisenberg, D. (2001). Protein interaction databases. Current Opinion in Biotechnology, 12(4), 334-339.
Zaki, N., Lazarova-Molnar, S., El-Hajj, W., & Campbell, P. (2009). Protein-protein interaction based on pairwise similarity. BMC bioinformatics, 10(1), 1-12.
Zhang, Y. (2014). Template-based structure modeling of protein–protein interactions. Current opinion in structural biology, 24, 10-23.
Zhang, F., Liu, S. Q., Wang, D. B., & Guan, W. (2009). Aircraft recognition in infrared image using wavelet moment invariants. Image and Vision Computing, 27(4), 313-318.
Zhou Z. Learnware: on the future of machine learning. Frontiers of Computer Science, 2016, 10(4): 589–590