J. Appl. Environ. Biol. Sci., 7(6S)79-85, 2017

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Optimization for Deep Learning Algorithms: A Review

Amirah Baharin, Siti Noorain Mohmad Yousoff, Afnizanfaizal Abdullah

Synthetic Biology Research Group,Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

Received: May 19, 2017
Accepted: July 18, 2017

ABSTRACT

In past few years, deep learning has received attention in the field of artificial intelligence. This paper reviews three focus areas of learning methods in deep learning namely supervised, unsupervised and reinforcement learning. These learning methods are used in implementing deep and convolutional neural networks. They offered unified computational approach, flexibility and scalability capabilities. The computational model implemented by deep learning is used in understanding data representation with multiple levels of abstractions. Furthermore, deep learning enhanced the state-of-the-art methods in terms of domains like genomics. This can be applied in pathway analysis for modelling biological network. Thus, the extraction of biochemical production can be improved by using deep learning. On the other hand, this review covers the implementation of optimization in terms of meta-heuristics methods. This optimization is used in machine learning as a part of modelling methods.

KEYWORDS: Deep Learning, Artificial Neural Network, Optimization Algorithm, Chemical Reaction Optimization

INTRODUCTION

Standard machine learning techniques has facing some restraints due to the limitation in processing raw data. The construction of machine-learning system needs proper engineering and domain knowledge. The process involved the extraction of raw data into internal representation, from classifier to detect the input patterns [1]. Thus, representation learning or deep learning is needed for the machine to process raw data for classification and detection.

For decades, there has been great interest towards the representation of deep learning methods. These methods comprise multiple levels of representation that acquired from the composition of simple but non-linear modules. The transformation of each representation begins with raw input into a higher level of representation. Thus, the complex functions can be identified through the transformation of the composition [1].

Deep learning methods offer great usefulness in terms of artificial intelligence society in solving problems. The utilization of these methods is applicable in vast category especially in science, government and business. Several successful implementations has been made in examined particle accelerator data, identify the effects of mutation non-coding DNA on gene expression and disease [2-3], brain circuit reconstruction [4], identify activity of potential drug molecules [5], image recognition [6] and speech recognition [7]. These implementations have beats other machine learning techniques.

There are some issues in using deep learning. Firstly, it is computationally expensive as it train several parameters of the data that include learning rate, number of layers and weight initialization [8]. It is also facing difficulties due to high tendency in getting stuck at local optima [9]. The contributions of evolutionary optimization methods are helpful to assist deep learning in achieving the best fit data. Optimization itself help in modelling process in a given data. The learning process from the data can be known as finding the model that gives best fit for the data. In searching process, heuristics and meta-heuristics can be used. There are two types of methods in finding solutions for optimization methods which are exact and approximate methods. Most of meta-heuristics can be group as approximate method. The output are vary in different run of computation due to randomization in calculation [10]. This paper focus on chemical reaction optimization (CRO) as the optimization method. It helps in optimizing the weights in the implementation of artificial neural network and achieve desirable performance by evolving the structure of the network [11].

In bioinformatics, both approaches contributes in genomics representations. Deep learning provides multiple learning layers. Thus, it helps in understanding abstraction and processing of the cells due to multiple levels of information provided. The biological structure in each layers can be inferred at multiple levels of resolution [12]. The extraction of biochemical production in a cell can be done effectively due to the high level of accuracy of prediction.

Corresponding Author: Amirah Baharin, Synthetic Biology Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor, Malaysia, E-mail: ami.b2525@gmail.com

The rest of this paper is organized as follows. In section 1 discussed about deep leaning methods that include several types of algorithms. Section 2 provide brief explanation about metaheuristics methods. In section 3 and 4 describes about the applications of deep learning and optimization algorithm in bioinformatics. While, in last section concludes the paper.

DEEP LEARNING

Deep learning is a representation learning methods which consist of multiple processing layers. Feature representation in deep learning consists of different levels of abstraction [13], as it owes hierarchical learning in representing data in each layer. It is a powerful approach to learn complex patterns from a lower level into a higher levels structures. Furthermore, the data can easily being monitored based on what has been learnt. The level of abstraction can be group in the form of edges, shape, contour and others. The implementation of deep learning comprised with the three learning methods known as supervised learning, unsupervised learning and reinforcement learning. These methods are relevant to be implemented in artificial neural network. Furthermore, it is usually creating hierarchies of more abstract representation in stationary data.

Conventional neural network made up of neurons that released real-valued activations. The process of activation started from input neurons through sensors that recognized the environment. Meanwhile, activation of other neurons through weighted connections from previously active neurons. The main idea is finding related weight that caused neural network desired behavior. However, the resultant behavior depends on the connection of the neurons. Sometimes, it requires long causal chains of computational stages; each stage transforms the aggregate activation of the network. The concept of learning by assigning credits towards these stages [14]. The main difference between deep and conventional neural network consists of more than one hidden layer. Figure 1 (a) and (b) shows conventional and deep neural network respectively.

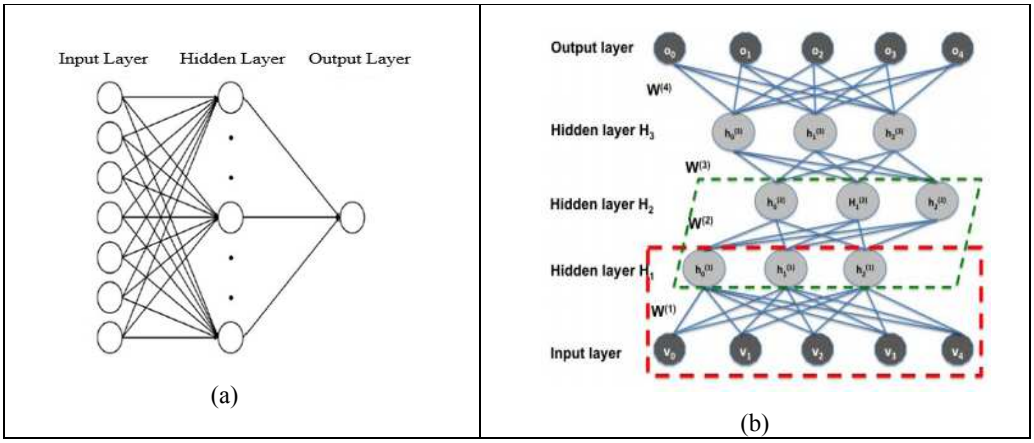

Figure 1: The architecture between (a) conventional neural network [15] and (b) deep neural network [16]

Backpropagation

Backpropagation is the most commonly used in artificial neural network which apply supervised learning method. The simplicity and computational efficient of this learning algorithm make it widely used in the neural network field. In backpropagation there are number of choices in order to train the networks. There are layers, nodes, training and testing node and others. These choices however are guided by heuristics and related theory in order in choosing the right choices.

This algorithm always seeks for better weights and varies in the gradient search. The lest-square optimality criterion often use in this algorithm. The concept of backpropagation, in every weight of the input, it always calculate the gradient error. It can be done by straight fully propagate the error in backward manner. The propagation is started at the top to the bottom throughout all modules. Figure 2 shows the implementation of backpropagation in every layer of the network.

However, there are several problems may arise in using backpropagation which are scaling problem and differentiability for calculating the gradient [17]. For the scaling problem may arise when using higher dimensionality of data. It may increase the complexity of data and lower performance. The training data might become slower due to gradient that stuck at local minima. Actually, the gradient can escape from the trap, however it might affect the next search whether it is better or worse. Global minima is hidden among the local minima. Thus backpropagation end up bouncing between local minima with no improvement. Backpropagation

was largely forsaken in late 1990's based on the thought that simple gradient descent may trapped at poor local minima [1].

Feedforward neural network is often used by deep neural network as an architecture and usually implementing by backpropagation algorithm. These applications learned through the mapping of fixed-size input (such as; image) to a fixed-size output (such as; probability for each of several categories). Units that are not in the input or output layer are conventionally called hidden units. The hidden layers can be seen as distorting the input in a non-linear way so that categories become linearly separable by the last layer.

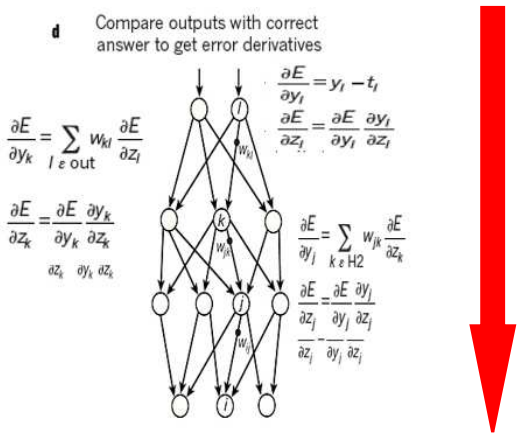

Figure 2: The implementation of backpropagation [1]

Convolutional Neural Network

The implementation of convolutional neural network is the alternative method apart from using deep neural network. It can be said that CNN provide better efficiency compared to deep neural network. CNN able to use less number of parameters to capture translational invariance. It can be implemented by weights replication over frequency and time. Thus, CNN improve the limitation occur in deep neural network by where it requires large network size and large number of training samples if the size is sufficient. Furthermore, deep neural network caused the input topology to be ignored [18]. This situation occur due to the representation of the input is in fixed order thus, the performance of the network are not affected. CNN local correlation for modelling provide advantageous towards other fields and spectral representation of speech shows high correlation [19]. The layers of CNN are fully connected at top and it may contains one or more convolutional layers as shown in Figure 3.

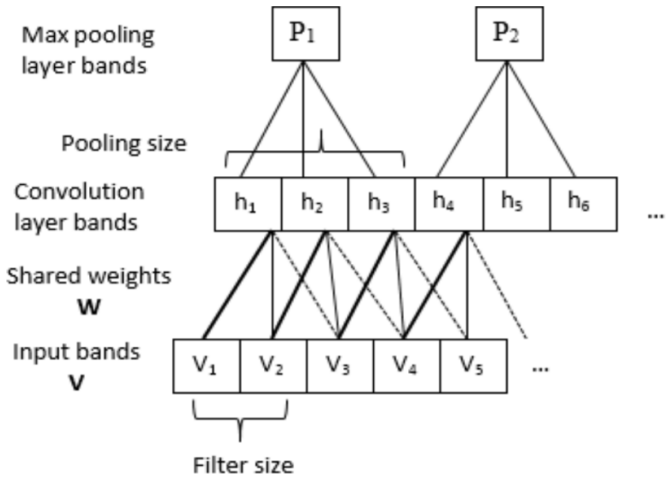

Figure 3: Convolutional neural network architecture shows one or more fully connected convolutional layers [20]

DeepBindAlgorithm

In [21] use deep learning techniques; state-of-the-art to be incorporated with DeepBind development. DeepBind algorithm is implemented based on the concept of CNN. It can identify pattern generated without

recognizing the location of patterns in the sequence. DeepBind algorithm is implemented to increase predictive power. The prediction is based on RNA editing and alternative interpretation of genetic variants, discovery of regulatory motifs, and splicing prediction.

Basically, genomics are well suited with deep learning that capture multiple levels of abstraction and information processing through multiple learning layers. The representation of biological structures in each layer inferred at multiple levels of resolution. Deep learning can be used on calculating protein–nucleic acid interactions through vast data sets. This method is applied for protein binding affinity prediction towards RNA or DNA sequence within two steps. These steps involved presentation learning convolution module and feature combination prediction module. The whole process is shown in Figure 3.

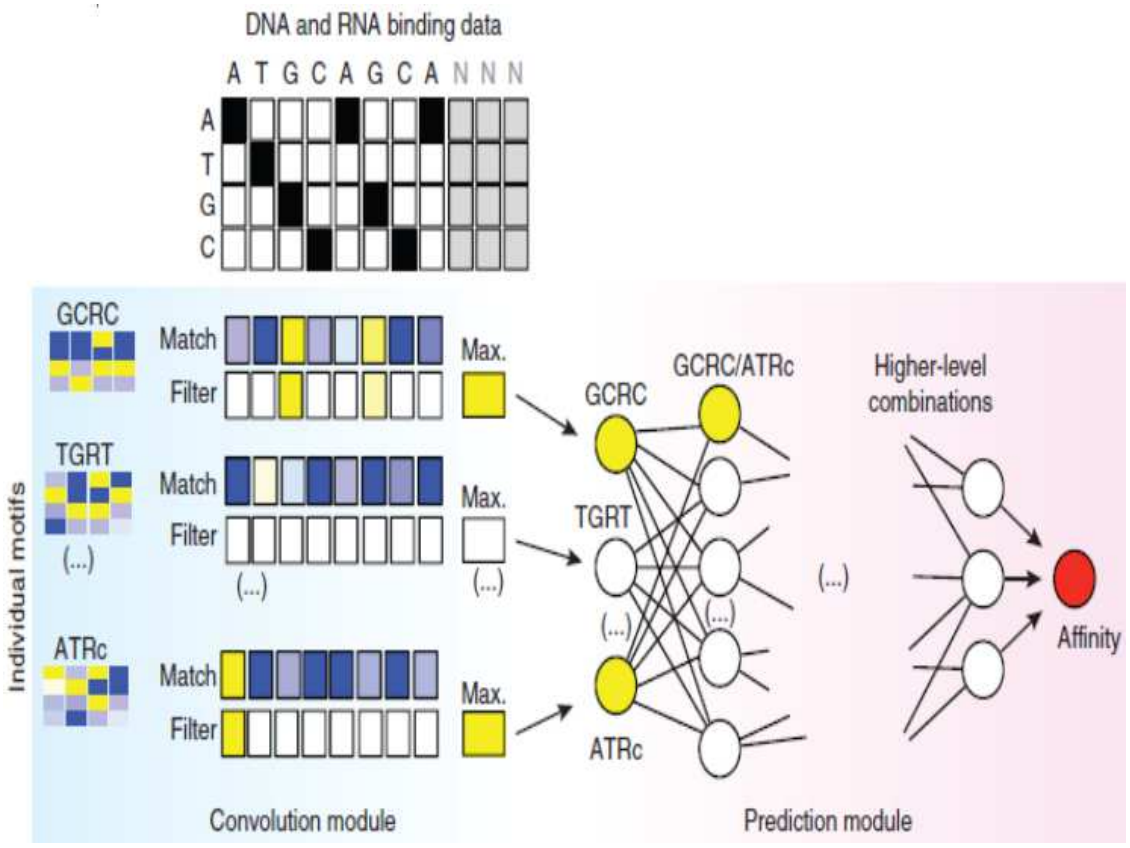

Figure 4: The application of CNN in deepbind algorithm in identifying protein motif [21]

OPTIMIZATION METHOD

Optimization is a process to identify the best characteristic or mathematical process of the input in order to find the optimal output values that solve a particular problem. Optimization algorithm has been widely used in many applications to solve variety of problems. The algorithms usually aid in finding the best solution depending on the formulation of the problems. Evolutionary algorithms inspired by the natural evolution by the process of natural selection. They are several implementations to train conventional artificial neural network by employing evolutionary optimization methods. These methods show the effectiveness and robustness in finding high quality of solutions even in the large search space [22]. Furthermore, this quality are qualified in training in infinite search space, multimodal and high dimensional. In this section, the listed meta-heuristic optimization methods are elaborated. They are Chemical Reaction Optimization (CRO) [10].

Chemical Reaction Optimization (CRO)

Chemical reaction optimization (CRO) implemented based on the concept of chemical reactions in the nature [10]. It is based on metaheuristic optimization to solve general computational problems. The terms metaheuristics is a type of framework that teach the computers in finding solutions of search space. The advantage in implementing metaheuristics in CRO is high degree of flexibility. Due to the random calculation,

the outputs obtain are varies due to different computational run. This method known as optimization problem solver by imitating the behavioral nature of reactions involved in the system.

```
1. Initialize the population.
2. If molecules satisfy inter molecular collision criteria
3. Then Select two or more molecules from population.
4. If molecules satisfy the synthesis criteria
5. Then perform synthesis operation.
6. Else perform inter molecular collision operation.
7. Else Select one molecule.
8. If it satisfies decomposition criteria
9. Then Perform decomposition operation.
10. Else Perform on wall effective collision operation.
11. Check for minimum point
12. If stopping criteria satisfies
13. Then obtain the best minimum point, gotostep15.
14. Else goto step 2.
15. End
```

Figure 5: Psuedo code for CRO algorithm [11]

There are three main stages involved in the CRO flowcharts which are initialization, iteration and final. In the initialization process, configuration of parameters and molecules. The molecule size has been set up equal as PopSize. The solutions are randomly generated in search space. The initial kinetic energy has been set up to InitialKE, while initial potential energy is identified by corresponding objective function. For the iterations, only one elementary reaction takes place. There are several criteria involved in the iterations. They are to identify whether the collision of the reaction is uni- or inter-molecular. Then, proceed to the next criteria to for decomposition reaction, followed by synthesis reaction and lastly stopping criteria. In the stopping criteria after the pre-defined total number of function evaluations is reached, the algorithm ends and outputs the so far best value as the global optimum. Figure 5 shows the overall pseudo code for CRO algorithm.

The efficiency in solving particles problems become provide great advantages towards CRO. Furthermore, training artificial neural network using CRO give a better performance compared to other evolutionary methods [11]. This is due to global minimum searching in representing configuration of the network with better performance. The employment of CRO in artificial neural network causes prevention in suffering structural hill climbing problem. It is caused by no restriction towards evolution of weight adaptation and network structure [23].

APPLICATION IN BIOINFORMATICS

The study of biological network in context of genomics has been implemented by the researchers. Graph partitioning method is implemented to study the genome-wide protein interactions. The functional modules are derive from clustering genome-wide protein interactions into several groups of protein interaction which shares similar function. However, biologist has facing obstacles in detecting the modules for cellular mechanisms. Thus, this methods is upgraded by including pre-processing, protein-protein interaction and graph partitioning algorithms based on nearest neighbor approach [24]. The advantage of this method includes the whole protein interaction network and the prediction of modules are closely related to known protein complexes. Whereas, deep learning introduced deepbind algorithm for regulatory genomics [21]. Biophysical models are incomplete and restricted towards specific types of structures for the interaction of regulatory protein and its target DNA or RNA. Thus, over-representation in target sequence occurred to identify regulatory motifs. Deepbind provides high accuracy in prediction compared to single-domain methods. The predictions are then implemented in identifying regulatory motifs. The prediction may include in genetic variants interpretations, RNA editing and alternative splicing [25].

On the other hand, computational systems biology facing great attention towards the researchers. It will leads to biochemical production in metabolic pathway of cell. Metabolic pathway can be known as nonlinear equation system as it can be describes as a mathematical model. Thus, in silico optimization method is implemented to overcome this situations [26]. The combinatorial optimization approach is applied in several studies in finding optimal solutions from finite set of objects. Several approaches has been proposed that contains several sets of hybridizations of optimization algorithms. There are Newton Cooperative Genetic Algorithm (NCGA; newton methods and cooperative genetic algorithm) [23], LEPSO (particle swarm optimization and differential evaluations) [27]and S-CRO (swarm-based chemical reaction optimization; firefly algorithm and CRO) [28]. These methods required fine-tuning of parameters in order to perform robust searching capability. The NCGA enhance metabolic pathway prediction, minimize the concentration of total

components involved [26]. Meanwhile LEPSO improved the resultant accuracy, the exploitation of neighbourhood solution is increases and effectively escape from local optima in high dimensionality [27]. Whereas, in S-CRO parameter estimation is improved, better fitness values are consistently found, able to select appropriate model and estimates reliable parameters in experimental data [28].

CONCLUSION

In this review, discussed about deep learning techniques which implementing multiple level of abstraction in feature representation. Deep learning can be characterized as rebranding of artificial neural network. This learning methods gains a large interest among the researchers because of better representation and easier to learn tasks. Even though deep learning is implemented, however there are some issues has been arise. There are easily getting stuck at local optima and computationally expensive. DeepBind algorithm shows that deep learning can cooperate in genomics study. It is to ensure on achieving high level of prediction protein binding affinity.

On the other hand**,** the optimization method which has been discusses consists of several meta-heuristics methods which can be categorized under evolutionary algorithms. The application of the techniques involved-CRO shows the diversity of optimization algorithm to improve the analysis of modelling techniques. Furthermore, these methods are able to solve the problems arise in conventional neural network as it provides high quality in finding solution in a given search space. The application of optimization methods enable the extraction of biochemical production of metabolic pathway. Deep learning will gives a good advantage in the biochemical production as it allows high level abstraction in cellular biological network. Thus, the use of CRO will improve the problems arise in deep learning which are getting stuck at local optima and it is computationally expensive. As CRO use global search in the search space to identify global minimum point. Thus, it will improve the training process in the network on refining the weight in order to have minimum error.

ACKNOWLEDGEMENT

This project was supported by Malaysian Ministry of Higher Education (MOHE), managed by Universiti Teknologi Malaysia Research Management Centre (RMC) under Fundamental Research Grant Scheme (FRGS) vote number R.J130000.7828.4F481.

REFERENCES

1. LeCun, Y., Y. Bengio and G. Hinton, 2015. Deep Learning. Nature, 521 (7553):436-444.
2. Leung, M.K., H.Y. Xiong,L.J. Lee and B.J. Frey, 2014.Deep Learning of the Tissue-Regulated Splicing Code. Bioinformatics, 30 (12): 121-129.
3. Xiong, H.Y., B. Alipanahi, L.J. Lee, H. Bretschneider, D. Merico, R.K. Yuen, Y. Hua, S. Gueroussov, H.S. Najafabadi, T.R. Hughes and Q. Morris, 2015.The Human Splicing Code Reveals New Insights into the Genetic Determinants of Disease. Science 347 (6218): 1-8.
4. Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, 2015. Going Deeper with Convolutions. In the Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp: 1-9.
5. Ma, J., R.P. Sheridan, A. Liaw,G.E. Dahl andV. Svetnik, 2015.Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships.Journal of Chemical Information and Modeling, 55 (2): 263-274.
6. Krizhevsky, A., I. Sutskever andG. Hinton, 2012. Imagenet Classification with Deep Convolutional Neural Networks.In the Proceedings of the 2012Advances on Neural Information Processing Systems, pp: 1090-1098.
7. Mikolov, T., A. Deoras, D. Povey, L. Burget andJ. Černocký, 2011. Strategies for Training Large Scale Neural Network Language Models.In the Proceedings of the 2011 IEEE Workshopon Automatic Speech Recognition and Understanding, pp: 196-201.
8. Bengio, Y., N. B. Lewandowski andR. Pascanu, 2013. Advances in Optimizing Recurrent Networks. In the Proceedings of the 2013 IEEE International Conferenceon Acoustics, Speech and Signal Processing, pp: 8624-8628.
9. Tirumala, S.S.,2014. Implementation of Evolutionary Algorithms for Deep Architectures. In the Proceedings of the 2014AIC, pp: 164-171.
10. Lam, A. and V.O. Li, 2010.Chemical-Reaction-Inspired Metaheuristic for Optimization. IEEE Transactions onEvolutionary Computation, 14 (3): 381-399.
11. Yu, J.J., A. Lam and V.O. Li, 2011. Evolutionary Artificial Neural Network Based on Chemical Reaction Optimization.In the Proceedings of the 2011 IEEE Congresson Evolutionary Computation, pp: 2083-2090.
12. Kotsiantis, S.B., I. Zaharakis and P. Pintelas, 2007. Supervised machine learning: A review of classification techniques.In: Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI

Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies (ed I.G. Maglogiannis) pp. 3-24.IOS Press, Amsterdam.
13. Bengio, Y.,2013.Deep Learning of Representations: Looking Forward. In the Proceedings of the 2013 International Conference on Statistical Language and Speech Processing, pp: 1-37.
14. Schmidhuber, J., 2015. Deep Learning in Neural Networks: An Overview.Neural Networks, 61: 85-117.
15. Liu, W.C. and C.E. Chung, 2014. Enhancing the Predicting Accuracy of the Water Stage Using a Physical-Based Model and an Artificial Neural Network-Genetic Algorithm in a River System. Water, 6 (6): 1642-1661.
16. Chen, X.W. and X. Lin, 2014. Big Data Deep Learning: Challenges and Perspectives. IEEEAccess, 2: 514-525.
17. Montana, D.J. and L. Davis, 1989. Training Feedforward Neural Networks using Genetic Algorithms. In the Proceedings of the 1989International Joint Conference on Artificial Intelligence, pp: 762-767.
18. LeCun, Y. and Y. Bengio, 1995. Convolutional Networks for Images, Speech, and Time Series. The Handbook of Brain Theory and Neural Networks, 3361(10): 1-14.
19. LeCun, Y., F.J. Huang and L. Bottou, 2004. Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting.In the Proceedings of the 2004 IEEE Computer Society Conferenceon Computer Vision and Pattern Recognition, pp: 1-8.
20. Sainath, T.N., A.R. Mohamed, B. Kingsbury and B. Ramabhadran, 2013. Deep Convolutional Neural Networks for LVCSR. In the Proceedings of the 2013IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 8614-8618.
21. Alipanahi, B., A. Delong, M.T. Weirauch and B.J. Frey, 2015. Predicting the Sequence Specificities of DNA-and RNA-Binding Proteins by Deep Learning.Nature Biotechnology, 33(8):831-838.
22. Whitley, D., T. Starkweather and C. Bogart, 1990. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. Parallel Computing, 14 (3): 347-361.
23. Nayak, J., B. Naik and H.S. Behera, 2015. A Novel Chemical Reaction Optimization Based Higher Order Neural Network (CRO-HONN) for Nonlinear Classification. Ain Shams Engineering Journal, 6(3):1069-1091.
24. Abdullah, A., S. Deris, S.Z.M. Hashim and H.M. Jamil, 2009. Graph Partitioning Method for Functional Module Detections of Protein Interaction Network. In the Proceedings of the 2009 IEEEInternational Conference onComputer Technology and Development, pp: 230-234.
25. Park, Y. and M. Kellis, 2015. Deep Learning for Regulatory Genomics. Nature Biotechnology, 33 (8): 825-826.
26. Ismail, M.A., S. Deris, M.S. Mohamad and A. Abdullah, 2015. A Newton Cooperative Genetic Algorithm Method for In Silico Optimization of Metabolic Pathway Production.PloS One, 10(5): 1-19.
27. Abdullah, A., S. Deris, S.Z.M. Hashim, M.S. Mohamad and S.N.V. Arjunan, 2011. An Improved Local Best Searching in Particle Swarm Optimization using Differential Evolution. In the Proceedings of the 2011 11th IEEE International Conference on Hybrid Intelligent Systems, pp: 115-120.
28. Abdullah, A., S. Deris, M.S. Mohamad and S. Anwar, 2013. An Improved Swarm Optimization for Parameter Estimation and Biological Model Estimation Process. Plos One, 8(4): 1-16.