A Literature Review and Comparative Analysis for Restricted Boltzmann Machine and Deep Belief Neural Network

Technical Report · May 2022

DOI: 10.13140/RG.2.2.19344.69127

Ahmed Tammaa
The British University in Egypt

A Literature Review and Comparative Analysis for Restricted Boltzmann Machine and Deep Belief Neural Network

Ahmed Tammaa
Artificial Intellegince major. Senior Student

The British university in Egypt.

Cairo, Egypt
ahmed189292@bue.edu.eg

***Abstract*—This paper covers the architecture of the Restricted Boltzmann Machine and the Deep Belief Network and a comparison between both. Besides a literature review of the model that started to take place in 2007 and its deeper counterpart in 2009. The Restricted Boltzmann can be used in weights initialization and feature selection to be an incorporate utility with other learning models. The Restricted Boltzmann Machine uses the statistical measure to calculate the performance and improvement rate unlike other learning models that rely on the loss function or in the RBM context it is the (energy function). The statistical measure is the Kullback-Leibler divergence also known as relative entropy. The Deep Belief Network can extend its functionality from its widely used speech recognition and image recognition to recognize network intrusion and stop the attacker.**

Keywords—Restricted Boltzmann machine, RBM, Deep belief networks, survey, literature review, energy function, machine learning, artificial neural networks

I. Introduction

Restricted Boltzmann-Machine (RBM) is a generative stochastic model that consists of two layers. Besides, it is an unsupervised learning model that uses input $x^{(t)}$ exclusively to extract a meaningful feature out of the data given in the input. Furthermore, it takes advantage of the unlabeled data. However, it can also work in semi-supervised learning if there exist some labelled data and other unlabeled data [1]. Restricted Boltzmann-Machine is an undirected graph model that maps distribution of a visible layer $x^{(t)}$ (input) to a hidden random binary layer h. Thus, the RBMs can be represented as a fully connected bipartite graph. It's parameterized by the weight of the connections for each node, besides the bias from the visible and the hidden layer. It's an energy-based model, since, it uses an energy function for training [2]. The energy function reflects the configuration of the latent variables. In which it can be interpreted as the minimization of the loss function in the typical machine learning models. Thus, it tries to find the weights and biases to minimize $E(x, h)$. To define the state for input $x^{(t)}$ the probability distribution is applied to the energy function. That uses maximum distribution. However, the RBMs were constructed to solve the problem of vanishing gradients. The feature extraction and/or prioritizing does not solve the vanishing gradient problem. Hence, the deep believe network is introduced to provide a complete solution for the problem [3]. The deep belief is similar to the multi-layer perceptron in the structure; however, both are distinct in operation [4]. This paper will cover the aspects of motivation of this approach, the architecture and activation layers, the statistics techniques of optimization, the evaluation metrics, the unique property of the RBM, examples of evaluation, and applications in greater details.

II. Literature Review and Applications

A. *The breif history of the model*

The Restricted Boltzmann model was first introduced by Geoffrey Hinton in 2007 after modeling the idea by Paul Smolensky in 1986 [5]. The model was successful at its time for feature detection, classification which corresponds to its application in dimensionality reduction based on the feature prioritizing ability and topic modeling. Recently, this approach regained popularity after Geoffrey Hinton developed the Deep Belief Neural Network (DBN) which uses the RBM as its building block in 2009 [6]. The DBN outperformed its shallow counterpart as the Multilayer perceptron outperformed the single layer perceptron.

B. *The use of Restricted Boltzmann machine to determine the weights*

The (RBMs) excels at the feature sorting thus Pacheco et al 2018 proposed to determine the weights and the bias of the Extreme Learning Machine (ELM) to increase the efficiency and effectives of the model. The (ELM) is a feedforward model with a random initiated weight and bias. It does not update them. Thus, the complete randomness of the weight and bias decreases the effectiveness of this model. With knowing the input significance, the weights can be better assigned to improve the model [7]. The use of the RBMs in training is not exclusive for the ELMs; Wang et al. 2017 proposed a non-iterative model to improve the layer training in Deep Learning (DL) models. Thus, in cooperating DL methods with the RBM approach.

C. *The Deep Belief Network optimization*

Since the Deep Belief outperform the RBM it is being used in more sophisticated applications. Thus, Wei et al 2019 proposed an Optimization for Deep Belief Neural Network for Intrusion detection (DBN-IDS) which can potentially be generalized. The approach is based on Particle Swarm Optimization and Fish Swarm Optimization. Both models are stochastic optimization model which incorporating in finding the global optimum solution. The Fish swarm and particle swarm algorithms optimized the training time and limited the number of layers which is an efficient way to extent the applications of the DBN since it reduced the hyperparameter of number of the needed layer The optimization used has increased the average detection time for network intrusion by at least 24.69% So combining the Deep Belief Network with other optimization algorithms can form a better function for the DBN. The Deep Belief Network has applications in image recognition and speech recognition thus it makes it suitable for detection of the intrusions of network in cyber security [9].

III. Properties and Comparative Analysis

A. *Architecture of Restricted Boltzmann Machine*

The Restricted Boltzmann Machine has a very basic structure consisting of two-layers. The visible layer input $x^{(t)}$ and a hidden layer h in a fully undirected connected bipartite graph as shown in diagram 1 [4]. In graph theory, the bipartite graph is a graph that can be divided into a disjoint set (U, V). In this context, the disjoint set is the visible layer and the hidden layer. A disjoint set is a pair of sets that are mutually exclusive.

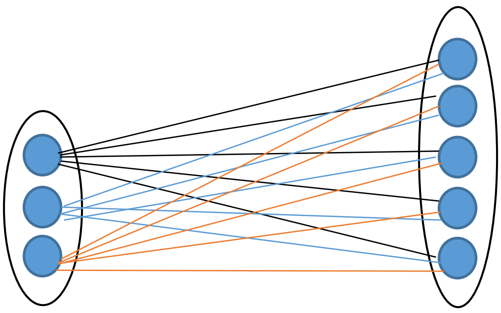

Diagram 1

$x^t(visible\ layer)$ $h\ (hidden\ layer)$

The Restricted Boltzmann Machine could be interpreted as the mathematical two-ways translator; moreover, it is divided into two steps the forward path and the backwards path [3].

1- Forward Path: takes the input from the input layer and encodes the input into a set of numbers representing the input.

2- Backwards path: reconstructs the results from the hidden layer to be as close as possible to the forward path

3- Measuring the quality of the iteration: After the forward and backward path is executed this step gives the metric of quality which is the Kullback-Leibler divergence will be described in the metrics section (IV).

The model trains until the backward path can reconstruct as close as possible to the forward path as shown in diagram 2. This way the model can infer the interrelationship between the inputs and knows the significance of each feature by re-adjusting weights and bias.

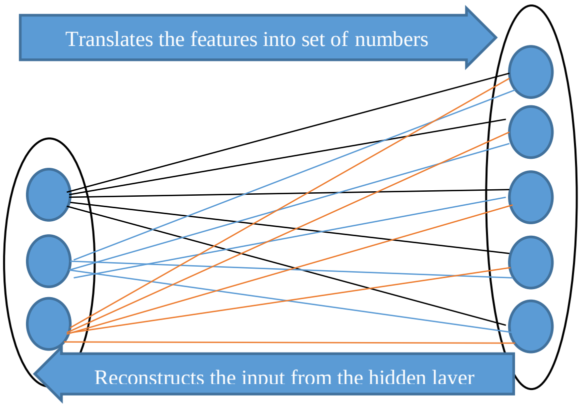

Diagram 2

B. *Architecture of Deep Belief Neural Network*

Like its shallow counterpart the DBN works in the same architecture of the Restricted Boltzmann Machine (RBM) the only difference is that the hidden layer works as visible to its next layer. More formally, for every $i\ where\ i$ $0 < i < n - 1$ such that n is the number of layers, the i^{th} layer is hidden for the $(i-1)^{th}$ layer and visible for the $(i+1)^{th}$. Moreover, layer 0 is always visible and layer n-1 is always hidden. As illustrated in diagram 3. Each layer is an RBM model.

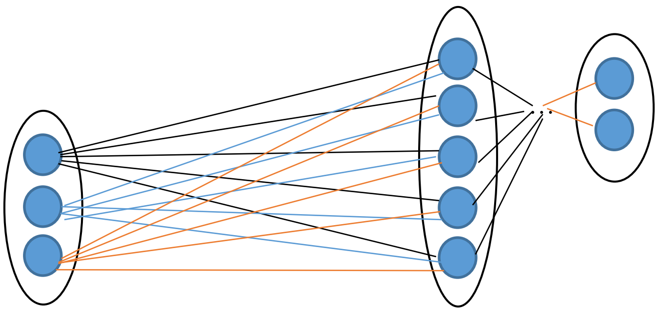

Diagram 3

C. *Restricted Boltzmann Machine Equations*

The forward path starts with random seed for the weights and bias. The visible layer is getting its value from this formula

$$\varphi(XW + b)\ (1)$$

Where φ denotes the activation function X denotes the visible layer W denotes the Weights and b denotes the bias. The backward path where the variables are adjusted so it can be trained. To train the model it needs to be distributed; thus, distributing this information uses this formula

$$p(x,h) = \frac{e^{E(x,h)}}{z}\ (2)$$

Where z is a normalization constant. However, the z is not interactable in this case, which implies, an inferred formula to be tractable. To solve this problem, the Gibbs sampling is introduced to the problem as solution. Gibb's sampling denoted by this equation

$$p(h_j = 1|x) = \frac{1}{1 + e^{-(b_j + W_j x_i)}} = \sigma\left(b_j + \sum_i (x_i + \omega_{ij})\right)$$
(3)

Then, the second step is the prediction of h given x denoted by this equation

$$p(x_i = 1|h) = \frac{1}{1 + e^{-(b_j + W_j h_i)}} = \sigma(a_i + \sum_j (h_j + \omega_{ij}))$$
(4)

The Gibbs step is done. The next step is doing the contrastive divergence

$$\Delta W = v_0 \oplus p(h_0|x_0) - x_0 \oplus p(h_k|x_k) \ (5)$$

Similar to gradient decent the new Weight matrix is denoted by this formula

$$W_{new} = W_{old} + \Delta W \ (6)$$

That concludes that its training is based on stochastic gradient decent. Moreover, the training can be illustrated in Diagram 4

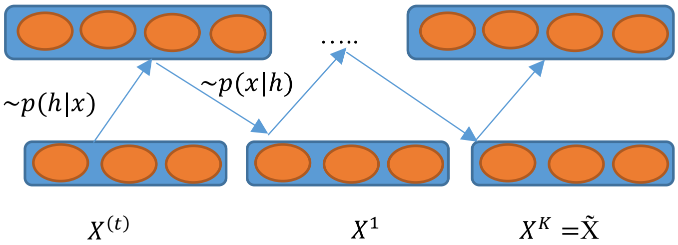

Diagram 4

The process starts by taking a sample by using Gibb's sampling as the visible layer. The hidden layer is sampled by the conditional probability of observing the hidden layer. After that the visible layer is reconstructed from the hidden layer by resampling it on P(x) given the values from the hidden layer P(x |h). This operation is repeated K times. Such that the last iteration $\tilde{X}$ is the negative sampling which estimates the negative phase of the gradient. Thus, the $\tilde{X}$ will be used to calculate the point estimates of the expectation. Furthermore, these steps are trying to minimize the energy of the samples from the dataset because minimizing energy results in high probability of observing a digit and maximizing the energy of other random images that looks different from the dataset and noise. Hence, makes more accurate generation.

The loss function is denoted by this energy function formula

$$E(x,h) = -\sum_i \sum_j w_{ij} v_j h_i - \sum_i b_i v_i - \sum_j c_j h_j \ (6)$$

$$such\ that\ w, x, h, b\ and\ c\ are\ weight\ vector,\ visible\ layer, hidden\ layer\ , the\ bias\ for\ hidden\ and\ visible\ layer\ respctively$$

Furthermore, the formula can be intuitively described in example of $b_i v_i$ term. If the bias vector b_i is negative, then v_i is preferable to be one than zero to increase the energy and the vice versa. The same applies for if v_i is negative since both variables are related to each other. This idea applies to all the three terms with products since they have a directly proportional relationship [10].

IV. Meterics Used to Evaluate Restiricted Boltzmann Machine

The Restricted Boltzmann Machine (RBM) performance is measured by Kullberg-Leibler Divergence (KL divergence) also called relative entropy. The relative entropy shows the difference in probability distribution from one distribution to another. Hence, it is very suitable to give a performance measure to the Restricted Boltzmann Machine as it is based on the probability distribution.

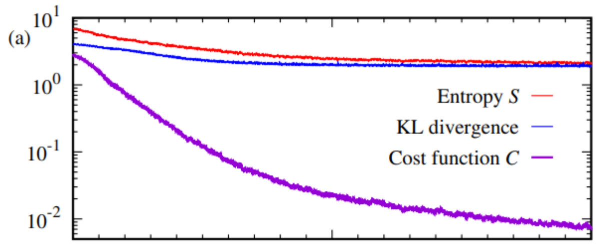

Figure 1 [10]

As shown in Figure 1, despite that the Cost function (Energy function) is approaching zero the KL-divergence is almost steady. Hence the cost function cannot be an indicator for the performance of the RBM since the cost function does not give accurate results for performance. Oh et al. 2020 introduced that the entropy can also be a good measure of performance for the Restricted Boltzmann Machine (RBM) which has similar results to KL divergence as shown in Figure 1.

V. Conclusion

In conclusion, the Restricted Boltzmann Machine (RBM) is a generative stochatsic model that is consisting of two-layered neural network (visible and hidden layer) The model was originally introduced in physics displine, however, it been formulated to be a good performing artificial neural network by taking advantages of its statistical properties and the idea of the energy function that functions as the loss function for other artificial neural networks. Its deeper counter part Deep Belief Network (DBN) uses the (RBM) as its building block to achieve a better results by taking the advantage of the multilayer structure that consists of Restricted Boltzmann Machine to enhance its performance based on the required application. The RBM is good at feature sorting thus it has an application of inializing the wieghts of the extreme learning machine. While the Deep Belief Network works in more advanced application like intrusion detection. The Restricted Boltzmann Machine performance is measured by the statisical method of relative entropy as it is more accurate method that the energy function curve as the RBMs are based on probability distribution.

REFERENCES

[1] Wong, Melvin, Bilal Farooq, and Guillaume-Alexandre Bilodeau. "Discriminative conditional restricted Boltzmann machine for discrete choice and latent variable modelling." Journal of choice modelling 29 (2018): 152-168.

[2] Fischer, Asja, and Christian Igel. "An introduction to restricted Boltzmann machines." Iberoamerican congress on pattern recognition. Springer, Berlin, Heidelberg, 2012.

[3] Smys, S., Joy Iong Zong Chen, and Subarna Shakya. "Survey on Neural Network Architectures with Deep Learning." Journal of Soft Computing Paradigm (JSCP) 2.03 (2020): 186-194.

[4] Ghojogh, Benyamin, et al. "Restricted Boltzmann Machine and Deep Belief Network: Tutorial and Survey." arXiv preprint arXiv:2107.12521 (2021).

[5] Hinton, Geoffrey E. "Boltzmann machine." Scholarpedia 2.5 (2007): 1668.

[6] Hinton, Geoffrey E. "Deep belief networks." Scholarpedia 4.5 (2009): 5947.

[7] Pacheco, Andre GC, Renato A. Krohling, and Carlos AS da Silva. "Restricted Boltzmann machine to determine the input weights for extreme learning machines." Expert Systems with Applications 96 (2018): 77-85.

[8] Wang, Xi-Zhao, Tianlun Zhang, and Ran Wang. "Noniterative deep learning: Incorporating restricted Boltzmann machine into multilayer random weight neural networks." IEEE Transactions on Systems, Man, and Cybernetics: Systems 49.7 (2017): 1299-1308.

[9] Wei, Peng, et al. "An optimization method for intrusion detection classification model based on deep belief network." IEEE Access 7 (2019): 87593-87605.

[10] Oh, Sangchul, Abdelkader Baggag, and Hyunchul Nha. "Entropy, Free Energy, and Work of Restricted Boltzmann Machines." Entropy 22.5 (2020): 538