Deepan Sanghavi

Where do features come from?

- Geoffrey Hinton

Abstract

Usually, in machine learning problems, you don’t have large amount of data which are labelled. As we know, the more data we have, the better is the prediction. But data without labelled is difficult to use for producing some insights. Hence, there is a algorithm known as Restricted Boltzmann Machine (RBM) which helps in producing insights and features from unlabeled data. This feature further can be used as one of the hidden layers in neural networks which can improve the model further. This algorithm understands the higher-order statistical structure of input vectors to generate features from them. These features can be used further in the feedforward neural network to tune weights of the network properly.

Introduction:

The author is trying to explain how features are extracted from unlabeled data and mentions that they are not generated innately by specifying various analogies and drawbacks. Also, this paper tries to analyze how the objective function helps in computing gradient which optimizes the problem. Paper is trying to find procedures that run well for learning purpose. Moreover, how backpropagation uses error derivatives to learn and optimize the model is being mentioned.

History:

Paper mention about two approaches to learn non-linearity in data using multiple layers of hidden units. One is back propagation method which requires lot of labelled data to train and explain disadvantages of having the requirement of labelled data. Moreover, it explained how without weight sharing, backpropagation was not useful earlier. Basically, paper explain entire path of how feedforward neural networks became as one of the top solution by overcoming its drawbacks in past. Other approach that paper cover was using statistical features of data to cover the non-linearity and mentioned how initially it was a theoretical concept and later it was implemented practically. A more insightful way to say this is that it treated each training case as a vector of desired outputs of a stochastic generative model, so the training data consisted entirely of high-dimensional labels and what was missing was the inputs. It explains how RBM is better than traditional feed forward network due to the requirement of labelled data and explains how RBM performs its computation to be better.

Undirected vs directed graphical models:

Paper creates an analogy between directed and undirected graphical models where these models generally deal with probabilistic models whose parameters have local structure that can be depicted by using graph. It says that undirected are the models which determines binary values for all the variables and directed graphs have the ancestral partial ordering of variables. There is a parent child relationship in directed models. Also, they have explained examples of both types of models and they have used 4 different models as an example with some scenarios.

Learning with incorrect inferences:

Furthermore, the paper tries to explain how to learn from inferences which are classified incorrectly such that it does not happen again. It says that sample the data from less complicated distributions where there is less correlation complexity and learn from this sample. Though this doesn't guarantee that it will improve the learning of the model.

RBM's:

One model that does allow simple, correct inference of distributed non-linear representations is a "Restricted Boltzmann Machine" (RBM) in which there are no connections between hidden units and no connections between visible units. It explains how RBM is converted into deep belief network. Once RBM is trained, its weights and biases define a joint distribution. So, to create a deep belief net, paper explains the process of training multiple RBM. After training the second RBM, we can apply the same trick again to improve its model of the aggregated posterior of the first RBM. After training a stack of RBMs in this way, we end up with a peculiar kind of composite model called a deep belief net (DBN).

Once stack of RBM are developed, it can be fine-tuned so that he weights in earlier layers have a chance to adapt to the weights that were subsequently learned in later layers. Generative fine-tuning maximizes the probability that the DBN assigns to the training data and can be done using a contrastive version of the wake-sleep algorithm. Each connection that is not part of the top-level RBM is split into a bottom-up recognition connection and a top-down generative connection and the weights on these two connections are untied so that their values can become different. It explains various ways to further fin tune the algorithm.

In the end, it explains things that RBM can do and has benefit over other methods.

Strengths:

- Paper is properly structured by explaining each concept at a time and in proper order
- Learning from unlabeled feature is what the world is looking for and it is very interesting topic
- Very well explained history or timeline of algorithm. Explained the need of RBM properly.

Drawbacks:

- Should have explained few concepts like RBM algorithm which some visual representation
- I find that you should have some prior knowledge of the content to understand the paper entirely