Rahul Thapa

Summary

In this survey paper, the authors describe a very popular branch of machine learning called Deep Learning. The paper starts by talking about various fields of machine learning that deep learning has been able to contribute on substantially such as web searches to content filtering and recommendation systems to cameras and smartphones. One distinguishing feature that deep learning has that other traditional machine learning algorithm does not have is this idea of representation learning. According to the paper, the representation learning is a set of methods that allows a machine to automatically discover the representations needed for giving the output by just feeding the input through the input layers. In other word, deep learning harnesses the power of hidden units which has the potential to learn some important features from the dataset without explicitly customizing those units. This is where the concept of back-propagation comes into play. Back-propagation minimizes the total error in the output units by adjusting the weights of the connection in the network, which allows the hidden units to capture those important features. The paper mentions that it is possible through the practical application of the chain rule of derivative, which the old machine learning procedures could not utilize.

The paper also talks about supervised learning which is basically this idea of training the machine leaning model in a supervised environment and using that trained model to make predictions on unseen input values. Supervised environment just means that you give the model large number of labelled training data, so that the network can figure out a complex function that maps from input to the desired output.

In the recent years, deep learning has got really popular specially because of its application in the computer vision using Convolutional Neural Network. ConvNets are designed to process the data that come in the form of multiple arrays for example a color image that contains pixel data in the form of red blue and green color content. The authors mention that the architecture of such networks consists of convolutional layer, pooling layer, and fully connected layer, each of which has a distinct function. For example, pooling layer is responsible for merging semantically similar features into one so that as we go deep into the network, we do not loose the connection between those features.

An important feature that deep learning lacked in the past was this idea of remembering the history of all the past elements of the sequence. Recurrent neural network solved this problem by processing one element at a time and maintaining state vector in their hidden state. According to the paper, this property is especially useful in sequence modeling for natural language, audio, which requires context from the past state. One problem with this network was that it was not able learn long-term dependencies because of vanishing gradient problem. This is solved by a procedure called Long Short-Term Memory (LSTM) that uses special hidden units for remembering inputs for a long time.

Strengths:

The paper is nicely worded. It is not to0 complicated to follow and therefore, very good for a beginner in deep learning. The paper also has a lot of references. So, if the reader wants to learn more about the background work, they can easily find the resources. The figures are descriptive and clear, and he even uses his own experiments to illustrate some procedures, which gives reader an idea on how those concepts can be applied in practice.

Weakness:

Some of the captions of the figures are very length. Some of that information can be included in the main text. Since this is a survey paper, the authors have not included many mathematically equations, however, it will be helpful in some cases to include some mathematical formulas, that could help the reader understand the concept better.

Confusion:

I was confused about the unsupervised learning procedure that the authors briefly touch upon in page 28. The authors also mention this idea of "pre-training" which I did not fully understand. I also did not fully understand what the task of convolutional layer is in the ConvNets. In the figure 3, the author uses CNN and RNN to caption the image. I was little confused about how the author uses RNN in this application.

Discussion:

1. What are the functions of various layers in CNN?
2. Why exactly is RNN useful in sequence modeling?
3. What are some ways and applications in which ConvNets and RNNs can be used together with reinforcement learning? How might it make the existing system better?