Nirajan Koirala

CSC 9010_001

1. Summary

The paper describes the wide use of a branch of Machine Learning called Deep Learning in today's world. These systems are used from image recognition to natural language processing and many other applicable fields. A history of the previous systems in performing such predictive task is provided and the authors explain how deep-learning which are representation-learning methods, perform the same tasks with higher accuracies using multiple level of abstraction levels. One of the major advantages of deep-learning systems is that the features are not designed by humans, but they are learned from the pattern in data by these systems. This kind of general-purpose learning procedure using data and computation can be very promising since these systems can take advantage of the increasing amount of data and computation power in the future.

The learning procedure of deep-learning systems can be divided into two parts: supervised learning and unsupervised learning and in the paper supervised learning methods are described. In supervised learning where labels are present in the dataset, the system adjusts its unit's weights in the network using stochastic gradient descent and tries to reduce the difference between the actual output and the desired output. These systems when deployed in multiple layers can detect minute details from the input and perform good discriminations. Backpropagation is used to compute the gradient of the objective function with respect to the weights in a multilayer system and it is carried out using basic calculus techniques like chain rule. Backpropagation can be applied repeatedly to propagate gradients through all the modules starting from the output layer through middle layers to the input layer. This method was commonly thought to get trapped in the local minmas so it was largely abandoned by the research communities until 2006 when a group of researchers used the backpropagation technique to achieve remarkable results in recognizing handwritten digits using very limited labelled data. With the advent of GPUs, record-breaking results were obtained in speech recognition tasks using similar techniques. ConvNets came into the picture as they could handle data in the form of multiple arrays like images, audio and videos remarkably well. In a series of stages, results of local weighted sum is passed through ReLU and patterns in array data like images are easily detected. After the detection of patterns in the convolution layer, semantically similar features are merged into one in the pooling layer. Training of the network in ConvNets can be performed using backpropagating gradients which is similar to the regular deep network training. As hierarchies exists in speech and text, this type of network can take advantage of this hierarchical property as higher level features could be obtained by composing lower-level ones. Therefore, ConvNets have been used in numerous systems like document-reading, speech recognition, image processing and face recognition and gained the attention of major technological companies like Google, Facebook and Microsoft as well as start-ups which require research and development projects to deploy ConvNets. Deep nets have also exponential advantages over the classical learning algorithms since they use distributed representations. In predictive tasks, representations of the elements and their configurations correspond to the variation seen in data and this greatly helps in increasing the prediction accuracy. So, the features are discovered and learned automatically by the neural networks. Also, neural network differs from the logic-inspired systems as they use big

activity vectors instead of single symbols to perform fast 'intuitive' inference. Neural network models exploit distributed representations and generalize well across semantically related sequence of elements in data. For tasks involving sequential inputs such as speech and language, RNN perform better than other kind of neural networks. RNN are able to learn from prior inputs while generating outputs and due to this mechanism, they can produce different output for same inputs depending upon the previous input results. RNNs are powerful dynamic systems and are very good at predicting next character in a text or the word in a sequence. Since RNN works by remembering the previous inputs for long tern dependencies, theoretical evidence has shown that it is difficult for these systems to learn and store information for very long. LSTM networks augment RNN by using special hidden units which remember inputs for a long time. Using several layers in LSTM enables entire speech recognition systems and they perform well in machine translation as well. Several other techniques have been tried to augment RNN by using tape-like memory which are called Neural Turing Machine and they have yielded excellent performance on standard question-answering benchmarks. Neural Turing Machines are being used for reasoning and symbol manipulation tasks and among other things these machines can also be taught 'algorithms'. Even though most of the deep-learning is supervised till now, one should expect unsupervised learning in the future to become very pervasive since human and animal learning is largely unsupervised.

The intended audience of this paper are Machine Learning enthusiasts, students, and anyone interested in the field of machine learning.

Main points made in the paper are how different kind of Deep Learning models are suited for different learning mechanisms. It has listed the history of such models and how are they being used in today's world.

2. Strengths

This review paper is able to touch almost all paradigms of Deep Learning and has explained its branches and uses in today's world remarkably well. All the examples of uses of deep-learning systems employed in today's world are very intuitive and specially the example provided in Figure 3 is very informative and visually helpful.

3. Weaknesses

The paper doesn't delve into unsupervised learning even though it mentions about it in the beginning part. The paper mentions that it is difficult to train RNN because of backpropagating gradients but it doesn't mention the exact mechanism of new ways of training them. Some real-world examples of the usage of combining ConvNets and RNN could have made that particular topic easier to comprehend.

4. Points of confusion

The architecture of ConvNets is a little hard to grasp as no figures are provided and also the workings of convolutional and pooling layers in particular are confusing.

5. Discussion questions

a) How is a Neural-Turing Machine designed and are they better at performing other deep-learning tasks besides natural language processing?

b) Can deep-learning systems employed to predict human behavior to avoid potential crimes or other undesirable activities?

c) Even though human learning is largely unsupervised, children are first taught before they start making observations. Does that mean the current supervised systems can evolve into unsupervised systems after learning to a certain degree?