Neural Machine Translation

Seminar Report-NLP
Arul, Ehsan
Rheinische Friedrich-Wilhelms-Universität Bonn
July 6, 2017

Abstract—Abstract

I. INTRODUCTION

II. DEEP LEARNING - AN INTRODUCTION

In the recent years, Deep learning has been making a big impact in various fields of computer science. This impact is profound in the perceptual learning task in the fields like computer vision, natural language processing, etc, [4]. Though the idea of training a multi layered neural network to perform function approximation is known to the research community for at least a couple of decades[5], due to the nature of training problem requiring large computational resources, the multi layered neural network remained less practical. With the advent of general purpose graphical processing units(GPGPUs) and the availability of training data, the neural networks are now a practical solution for many perceptual tasks. One of the early success of deep learning was in the field of image classification. In the annual ImageNet classification challenge [1], AlexNet [3] showed a remarkable improvement in the state-of-the-art accuracy. Within a few years, more sophisticated architectures like Google Inception Network [6], Deep Residual Network [2], etc., has improved the accuracy to be comparable with a human in that task. The generality of the neural network made it easily possible to be used for a wide variety of tasks. With more people working on deep learning and the ideas arising in solving on the problems being easily transferable, the deep learning is the state-of-the-art for many learning tasks. In the following section, we will briefly summarize the basics of deep learning.

The basic building block of a Multilayer network, or in more technical term Multilayer perceptron(MLP) is a perceptron. The preceptron 3 closely resembles a neuron in a human brain 1. A preceptron takes a set of input values, computes a weighted sum of the inputs, and outputs a scalar value of a after applying an activation function (mostly non-linear). The weights for computing the weighted sum and the threshold in the activation function are initially set to random values and are learned from the training data. Mathematically, a perceptron with the weight matrix A and the threshold T for an step activation function, for the input vector x, outputs the following,

$$f(x) = \begin{cases} 1, & \text{if } Ax > c \\ 0, & \text{otherwise} \end{cases}$$

A layer has many number of neurons and many layers are stacked one on top of the other to form a MLP. An MLP with very many layers is called as Deep Neural Network(DNN). In
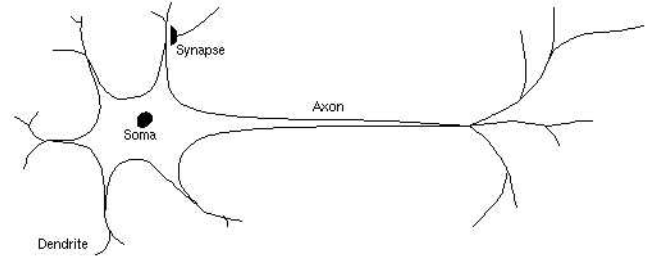

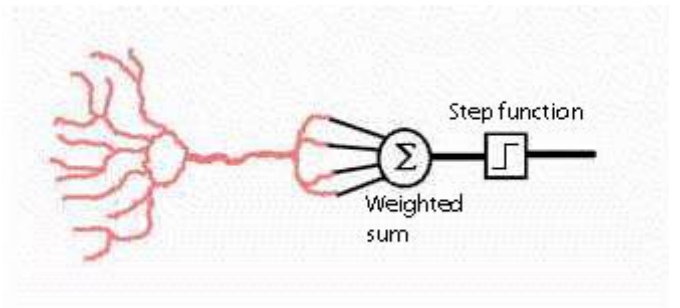
Fig. 1: A biological neuron.

Fig. 2: An artificial neuron (perceptron)

the following section we will discuss the properties of DNNs and how they are useful in the context of supervised machine learning.

A. Supervised learning

Supervised learning is the problem of predicting an new output $y\prime$ for a new input $x\prime$, and set of training data that provides a set of input and output $\{x \mapsto y\}$. Most of the supervised learning problems can be formulated as either a classification or regression problem. Classification is the problem of predicting the label for a given $x\prime$ among a set of given labels $\{Y\}$ while regression of predicting a continuous valued output for a given input. The usual way of doing supervised learning is to extract features from the given data and train a model to perform classification or regression.

$$x \xrightarrow[\text{Feature extraction}]{\mathbb{D}(x)} x\prime_{intermediate} \xrightarrow[\text{classifer/regressor}]{\mathbb{F}(x\prime|\theta)} y$$

The power of the DNNs lie in the fact that the models learns the features directly from the given training data and
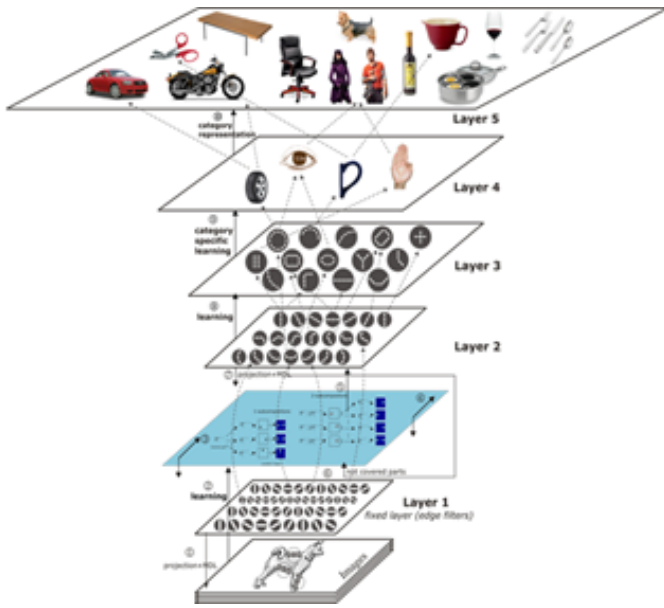
Fig. 3: Layered representation in an MLP

avoids hand engineered features.

$$x \xrightarrow[\text{hierachical}]{\mathbb{M}(x|\Theta)} y$$

The DNNs with neurons in one layers being connected to all the neurons in the next layer are called Feed Forward Networks. The feed forward networks are harder to train since they have lots of parameters. To make a the DNNs learn useful representation for performing supervised learning, we need special types of connections between the neurons. Two of the most commonly used DNN architectures are Convolutional Neural Networks(CNN) and Recurrent Neural Networks(RNN). In the following section we will discuss these architectures in detail.

REFERENCES

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[5] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.