LARGE-SCALE DEEP LEARNING

Speaker: Quoc V. Le
Research Scientist, Google

Until recently, manually feature engineering is a crucial step in many pattern recognition and machine learning algorithms. Key to this approach is to manually design features to transform raw input data into a new space such that the data can be learned effectively. This approach has been widely used in computer vision (e.g., SIFT and HOG features), speech recognition (e.g., Spectrogram and MFCC features), and information retrieval (e.g., BM25 and TFIDF features). The main weakness with this approach is that it does not scale well for new problems: to solve a new problem, we need to invent a new set of features.

Since 2009, researchers have been working on techniques that attempt at learning features from data. Key to our approach is to use neural networks with some simple algorithmic modifications (also known as “Deep Learning”). Despite its promise to transform many fields, Deep Learning is usually slow and only shown to work well on small problems.

In 2011, the Google Deep Learning project (the other name is “Google Brain”) was founded as an attempt to scale up Deep Learning so that it can work for large scale problems. In this setting, Deep Learning begins to shine. The outcome is a breakthrough: not only we save many years of manual feature engineering, the algorithms are much better in terms of accuracy. In late 2011, we show that Deep Learning can learn the concept of cat from watching YouTube videos. What surprises us is that the network can learn to invent this concept of cat without being told ahead of time about the concepts to learn.

Since then, the system has made significant improvements in many areas, especially computer vision, speech recognition, and language understanding. In computer vision, neural networks, especially convolutional neural networks developed in the 90’s, are now becoming the state of art systems. In speech recognition, our system plays a key role in the acoustic model and is now used in Voice Search for Android. In language understanding, our system can learn vectors for words which represent the semantics of the words. For example, vector for word “love” and “like” are close and both far away from “tree”. Combining the word vectors into a recurrent network allows us to also represent a sentence as a vector such that semantically similar sentences have similar vectors.

In this talk, I will describe our approach to scaling up Deep Learning and highlight some of the progress in these areas (object recognition, speech recognition and language understanding) thanks to the scaling up approach.