IDC LINK
Real-time IDC Research® opinion on industry news, trends and events

Baidu's Deep Learning Efforts: Notable Progress on Many Fronts

December 04, 2015

By: Robert Sorensen

IDC's Quick Take

Researchers from the Institute of Deep Learning (IDL) at China's Baidu, one of the world's largest Internet search companies, will be presenting yet another example of its broad range of activities in deep learning at the upcoming Neural Information Processing Conference (NIPS 2015) to be held in Montreal starting on December 7. This particular presentation will center on Baidu's work in image question answering (QA), the complex process whereby a user asks a computer to identify specific features in an image. This is only one element of Baidu's aggressive pursuit of a wide skill set in machine and associated deep learning in areas including machine vision, facial recognition, speech recognition, and image search capabilities that the company plans to infuse into nearly every application or service it offers.

News Highlights

The presentation will provide an update on Baidu's QA model development, one of the more complex deep learning problems that combines image recognition, natural language processing to decipher an input question, and an accurate generation of a correct written response. Baidu researchers have added complexity by developing a system that can operate in both Chinese and English. For this particular project, Baidu used a large training data set containing over 150,000 images and 310,000 freestyle Chinese question-answer pairs and their English translations. Because it is difficult to accurately quantify the correctness of answers in QA applications, the model was evaluated by human judges through a Turing Test where answers were provided by both humans and the QA model. In the most recent experiment, in 64.7% of cases, human judges could not distinguish responses between the QA model and humans. Although it is difficult to put these results in context with other counterpart research efforts as there are currently no standard benchmarks for assessing the accuracy of QA models, it is clear that the Baidu model benefits from the deep learning expertise at its IDL facility and its ambitious use of a large training set. In addition, the ability to work well in a bilingual capacity is impressive.

This Baidu presentation follows directly on the tail of a recent talk given by Dr. Bryan Catanzaro, a researcher at Baidu's Silicon Valley AI Lab, at SC15, which was recently held in Austin, Texas. In that presentation, Catanzaro discussed the complexities of scaling deep learning activities that use more powerful HPC resources and encompass larger sets of data to drive deep learning programs to new levels of performance and insight generation. A highlight of the talk was the insight that deep learning has dramatically changed the course of long-standing artificial intelligence (AI) research direction due to its ability to encompass such large data sets, allowing the field to make rapid progress on difficult problems such as improving the ability of computers to interact with a wide base of users in a more natural and easy-to-use environment. Despite these potential gains, Baidu researchers have noted that problems persist across a number of AI applications such as speech recognition where there is much improvement needed in interpreting English in a foreign accent, emotional recognition, semantic understanding, dialogue, and speech synthesis.

Baidu developers typically note that HPC capability is a key element in making progress in the development of better techniques for training AI models as it enable researchers to iterate more quickly their hypotheses and further the state of the art in deep learning algorithms. For example, Baidu researchers indicate that a typical deep learning training session — the process whereby a deep learning program is trained within a specific application domain — takes about 20 exaflops of computational resources. Due to the complexities of programming such deep learning training sessions, they can take up to six months to perform. As a result, many researchers, not just Baidu, are increasingly turning to specialized computational platforms that offer the computational support required for the rigorous demands of today's most advanced deep leaning algorithms including GPU platforms from NVIDIA and other computational accelerators.

IDC's Point of View

Baidu has made it clear that it is highly interested in advancing the state of the art in deep learning algorithms to support its ambitious agenda of expanding its Internet-based products and services, and the company is going offshore to satisfy that need. Recognizing that many of the leading developers in deep learning are located in the United States, including Facebook, Google, IBM, Microsoft, Netflix, and Yahoo, last year, the firm opened up its new Silicon Valley AI Lab to augment ongoing efforts at its Beijing Deep Learning and Big Data Labs. The firm is likely being well served by its presence in Silicon Valley, the current epicenter of deep leaning and related big data development and its demonstrated ability to attract some of the best talent in the world. For example, Baidu Chief Scientist Andrew Ng previously taught machine learning at Stanford, cofounded Coursera, and helped create the Google Brain project before leaving Google to join Baidu in early 2014.

Baidu, like many of its U.S. counterparts, will benefit by being able to draw on its large base of hyperscale centers, provisioned in many cases with in-house designed servers, and by its access to large consumer and commercial data sets on which to experiment. Baidu's stated intention is to use a strong base of HPC systems coupled with its own hypercenters as training platforms for its projected raft of deep learning programs that will then be pushed out as commercial applications for a wide base of smartphone, wearable, and IoT devices. As such, Baidu will benefit from being able to tap into a domestic smartphone market that is the largest in the world, with over 1.27 billion users and four times that of the United States, in addition to serving China's soon-to-be-burgeoning IoT market that could see China as the most data-dense IT nation in the world in the next few years. Baidu's emphasis on HPC technology as a key enabler of its vision and strategy mirrors what IDC is hearing from other leading IT companies.

Subscriptions Covered:
HPC User Forum, Technical Computing