Distinguished Lecture Series

EECS COLLOQUIUM

Fall 2015

Wednesday
October 14
4:00 - 5:00pm

Refreshments
served at 3:30pm

Hewlett-Packard Auditorium
306 Soda Hall

Deep Learning: Progress in Theory and Attention Mechanisms

Yoshua Bengio
Professor of Computer Science and Operations Research
University of Montreal

Abstract

Although neural networks have long been considered lacking in theory and much remains to be done, theoretical evidence is mounting and will be discussed, to support distributed representations, depth of representation, the non-convexity of the training objective, and the probabilistic interpretation of learning algorithms (especially of the auto-encoder type, which were lacking one). Beyond theory, this talk will report about an exciting new development in deep learning research regarding the role of attention. Could some form of attention mechanism be useful in machine learning systems? We introduced a few months ago an attention mechanism in recurrent neural networks for neural machine translation in order to cope with the difficulty of handling long sequences of words in encoder-decoder neural machine translation architectures. Attention allows the network a kind of bypass to just the pieces of evidence that it needs to focus on in order to compute its next state or output. With early successes of this attention mechanism for machine translation, we decided to try the same idea (and almost the same code) on speech recognition, and later on caption generation, with surprising and quick success. In the long term, we conjecture that such mechanisms, applied to a very large set of memorized elements (and not just the words or the pixels in the current input), could be a key for bypassing the difficulty of learning very long term dependencies.

Biography

Yoshua Bengio received a PhD in Computer Science from McGill University, Canada in 1991. After two post-doctoral years, one at M.I.T. with Michael Jordan and one at AT&T Bell Laboratories with Yann LeCun and Vladimir Vapnik, he became professor at the Department of Computer Science and Operations Research at Université de Montréal. He is the author of two books and more than 200 publications, the most cited being in the areas of deep learning, recurrent neural networks, probabilistic learning algorithms, natural language processing and manifold learning.

For more information about EECS Colloquiums please see:
http://www.eecs.berkeley.edu/Colloquium/