Microsoft Research Faculty Summit 2014

15TH ANNUAL

Deep Learning For Text Processing

Jeffrey A. Bilmes

Professor
Departments of Electrical Engineering
& Computer Science and Engineering
University of Washington, Seattle
`http://melodi.ee.washington.edu/~bilmes`

Monday, July 14th, 2014

Deep Neural Networks

Aspects:

- Deep Generative Models and Deep Non-linear Mappings (classification & regression)
- Convolutional Deep Models and Aggregation methods (sum, max pooling, etc.)
- Non-Linear Learning of Feature Representations, Autoencoders
- Distributed Representations (efficiency of parameter utilization).

Applications and Successes:

- Computer Vision (image, video, motion capture)
- Speech (acoustic modeling in speech, speech perception, continuous representations of words/phrases, etc.) and NLP (language modeling, sentiment classification, etc.)
- Information Retrieval, learning similarity, deep CCA, etc.

Deep Neural Networks

Aspects:

- Deep Generative Models and Deep Non-linear Mappings (classification & regression)
- Convolutional Deep Models and Aggregation methods (sum, max pooling, etc.)
- Non-Linear Learning of Feature Representations, Autoencoders
- Distributed Representations (efficiency of parameter utilization).

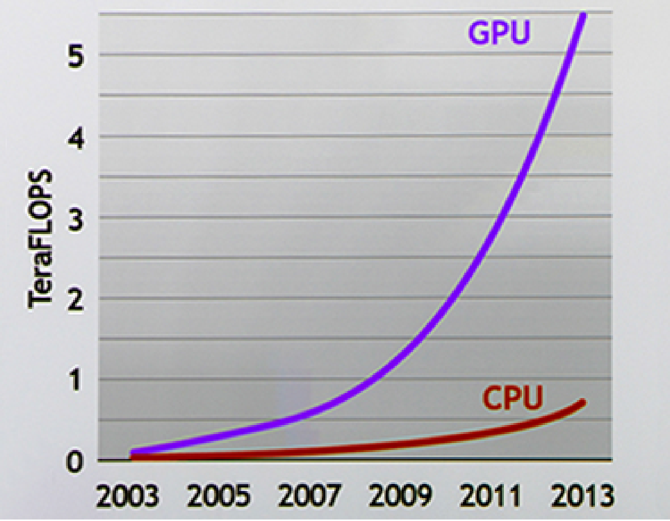
Applications and Successes:

- Computer Vision (image, video, motion capture)
- Speech (acoustic modeling in speech, speech perception, continuous representations of words/phrases, etc.) and NLP (language modeling, sentiment classification, etc.)
- Information Retrieval, learning similarity, deep CCA, etc.

Resources:

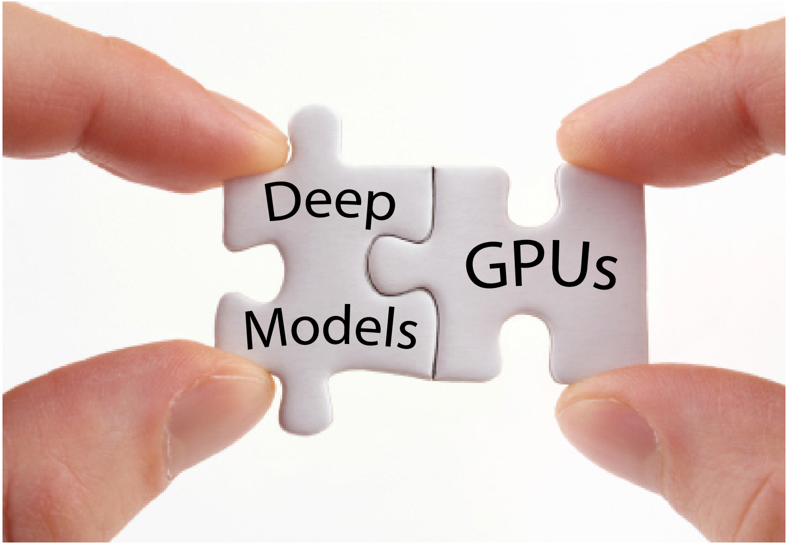
- Big data - big data is different data, big neural networks are different neural networks
- Big computation - CPU only computation is insufficient, while GPUs are feasible

GPU vs. CPU Performance Over Time



Source: GTC (GPU Tech. Conference) Keynote, NVIDIA CEO Jen-Hsun Huang, 3/25/2014

Deep Models and GPUs



Questions

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.
- GPUs: commodity high-performance supercomputing (thanks to gaming industry). Hence, inexpensive and ubiquitous.

Questions

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.
- GPUs: commodity high-performance supercomputing (thanks to gaming industry). Hence, inexpensive and ubiquitous.
- Part of the success of deep models is due to wave of GPU performance increase, and that deep models' computational needs are well-matched with what GPUs offer.

Questions

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.
- GPUs: commodity high-performance supercomputing (thanks to gaming industry). Hence, inexpensive and ubiquitous.
- Part of the success of deep models is due to wave of GPU performance increase, and that deep models' computational needs are well-matched with what GPUs offer.
- Q: Could other machine learning methods (e.g., kernel machines, random forests, graphical models, etc.) have a correspondingly well matched, inexpensive, ubiquitous, and high performance hardware platform to run on?

Questions

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.
- GPUs: commodity high-performance supercomputing (thanks to gaming industry). Hence, inexpensive and ubiquitous.
- Part of the success of deep models is due to wave of GPU performance increase, and that deep models' computational needs are well-matched with what GPUs offer.
- Q: Could other machine learning methods (e.g., kernel machines, random forests, graphical models, etc.) have a correspondingly well matched, inexpensive, ubiquitous, and high performance hardware platform to run on?
- Q: Or is the nature of such neural computations (matrix operations) such that it will always have an advantage due its regularity and hardware optimizability?

Questions

- Deep Learning needs big data to train, and is well-matched to GPUs (for matrix operations) to do so.
- GPUs: commodity high-performance supercomputing (thanks to gaming industry). Hence, inexpensive and ubiquitous.
- Part of the success of deep models is due to wave of GPU performance increase, and that deep models' computational needs are well-matched with what GPUs offer.
- Q: Could other machine learning methods (e.g., kernel machines, random forests, graphical models, etc.) have a correspondingly well matched, inexpensive, ubiquitous, and high performance hardware platform to run on?
- Q: Or is the nature of such neural computations (matrix operations) such that it will always have an advantage due its regularity and hardware optimizability?
- Q: If so, can text/NLP applications like parsing, SMT be made just as regular?