Microsoft

Microsoft Research Faculty Summit 2013

Microsoft

Deep Machine Learning: A Panel

Organizers: Li Deng and John Platt,
Microsoft Research, Redmond

Microsoft Research
Faculty Summit

What is “Deep Learning”?

A sub-field within machine learning based on algorithms for learning multiple levels of representation in order to model complex relationships among data.

Higher-level features and concepts are thus defined in terms of lower-level ones, and such a hierarchy of features is called a deep architecture… (Wikipedia)

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous.

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child?

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts.

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people.

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss.

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket.

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible.

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases.

Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical.

The New York Times

Monday, June 25, 2012 Last Update: 11:50 PM ET

Arizona Ruling Only a Narrow Opening for Other States

By JULIA PRESTON 9:06 PM ET

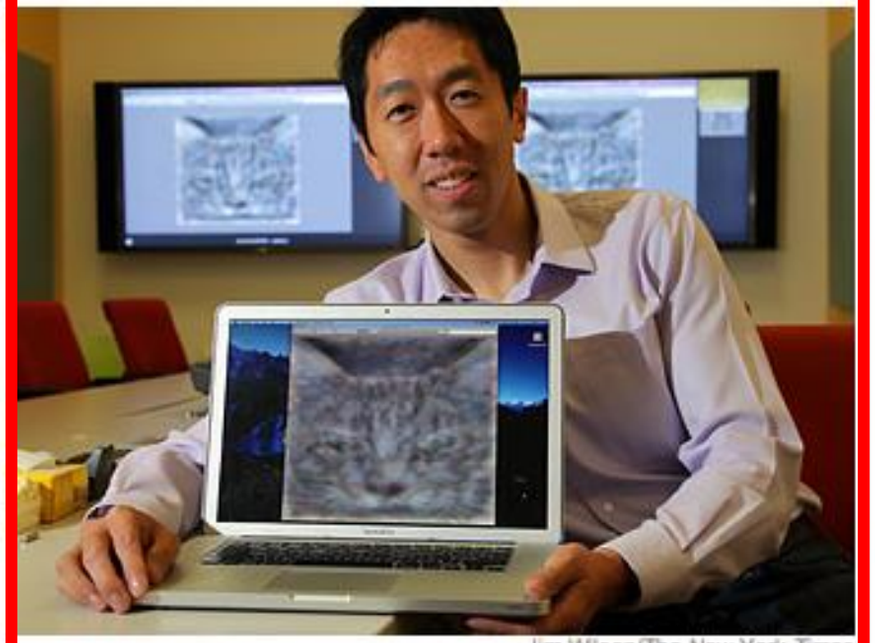
The Supreme Court's mixed decision on Arizona's immigration law does not seem likely to unleash a new wave of legislation by other states to crack down on illegal immigration.

- Blocking Parts of Arizona Law, Justices Uphold Centerpiece
- On Campaign Trail, Obama and Romney React 10:05 PM ET
- Interactive: Supreme Court Decision on Immigration Law

Justices Bar Mandatory Life Terms for Juveniles

By ADAM LIPTAK and ETHAN BRONNER 8:44 PM ET

The justices ruled that such sentencing for those under 18



Jim Wilson/The New York Times

Despite Itself, a Simulated Brain Seeks Cats

By JOHN MARKOFF 12 minutes ago

A Google research team, led by Andrew Y. Ng, above, and Jeff Dean, created a neural network of 16,000 processors that reflected human obsession with Internet felines.

Basketball Recruiters Aim for Middle Schoolers

By ADAM HIMMELSBACH and PETE THAMEL 10:03 PM ET

OPINION »

OPINIONATOR | ANXIETY

Haunted Heart

The surgeon calmly fiddled with the beating engine of my existence.

- Krugman: Abdication
- Edsall: Working-Class Chaos
- Editorial: Spending Cuts
- The Stone: Evolution and Our Inner Conflict
- Taking Note: Citizens United

MARKETS » At 11:48 PM ET

JAPAN		CHINA
Nikkei	HangSeng	Shanghai
8,660.36	18,901.11	2,214.11
–74.26	+3.66	–10.00
–0.85%	+0.02%	–0.45%

Data delayed at least 15 minutes

GET QUOTES My Portfolios »

Stock, ETFs, Funds Go

Geoff Hinton

The New York Times

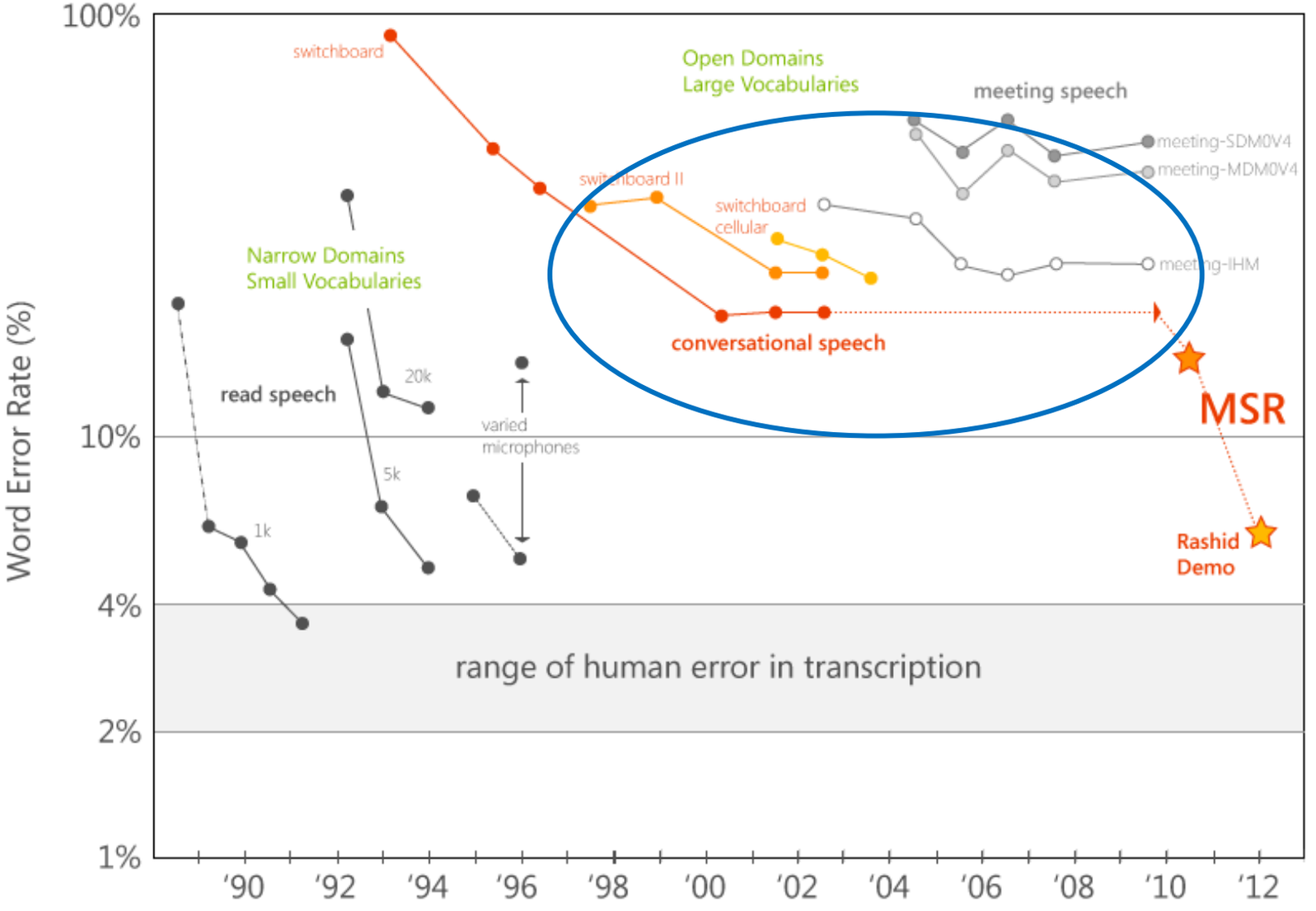
Scientists See Promise in Deep-Learning Programs

John Markoff

November 23, 2012

Rick Rashid in **Tianjin**, October, 25, 2012

NIST Evaluations of Automatic Speech Recognition



After no improvement for 10+ years by the research community…

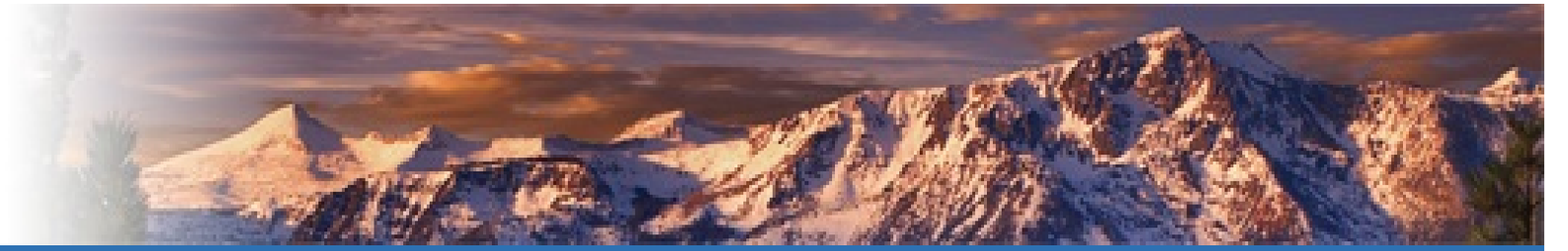
…MSR reduced error from **~23%** to **<15%** (and under 7% for Rick's demo)!

Li Deng (MSRR), Dong Yu (MSRR), & Geoffrey Hinton (Utoronto);

Frank Seide (MSRA)

MSR Partnering w. Academia in Early Days

Neural Information Processing Systems Foundation



NIPS : Conferences : 2009 : Program

- NIPS Home
- Overview
- Conference Videos
- Workshop Videos
- Program Highlights
- Tutorials
- Conference Sessions
- Workshops
- Publication Models
- Demonstrations
- Mini Symposia

Li Deng, Dong Yu, Geoffrey Hinton

Microsoft Research; Microsoft Research; University of Toronto

Deep Learning for Speech Recognition and Related Applications

7:30am - 6:30pm Saturday, December 12, 2009

Location: Hilton: Cheakamus

Abstract: Over the past 25 years or so, speech recognition technology has been dominated by a "shallow" architecture --- hidden Markov models (HMMs). Significant technological success has been achieved using complex and carefully engineered variants of HMMs. The next generation of the technology requires solutions to remaining technical challenges under diversified deployment environments. These challenges, not adequately addressed in the past, arise from the many types of variability present in the speech

Joint publications:

IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, Nov. 2012

Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury

Deep Neural Networks for Acoustic Modeling in Speech Recognition

The shared views of four research groups

INTERSPEECH 2010

Binary Coding of Speech Spectrograms Using a Deep Auto-encoder

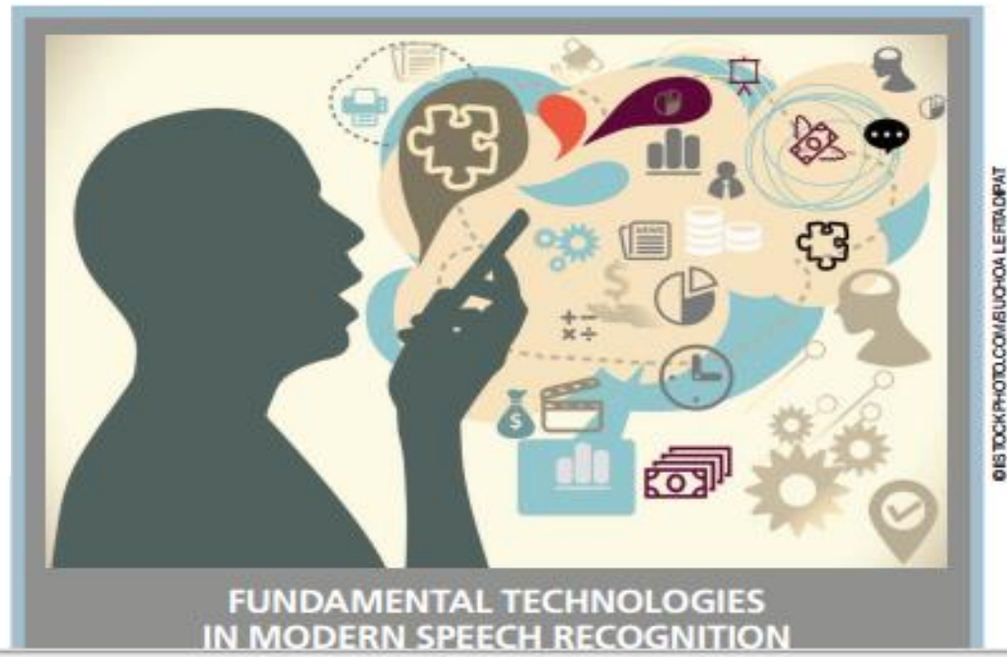
L. Deng[1], M. Seltzer[1], D. Yu[1], A. Acero[1], A. Mohamed[2], and G. Hinton[2]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98052, US
[2] University of Toronto, Toronto, Ontario, Canada
{deng|mseltzer|dongyu|alexac}@microsoft.com; {asamir|hinton}@cs.toronto.edu

Abstract

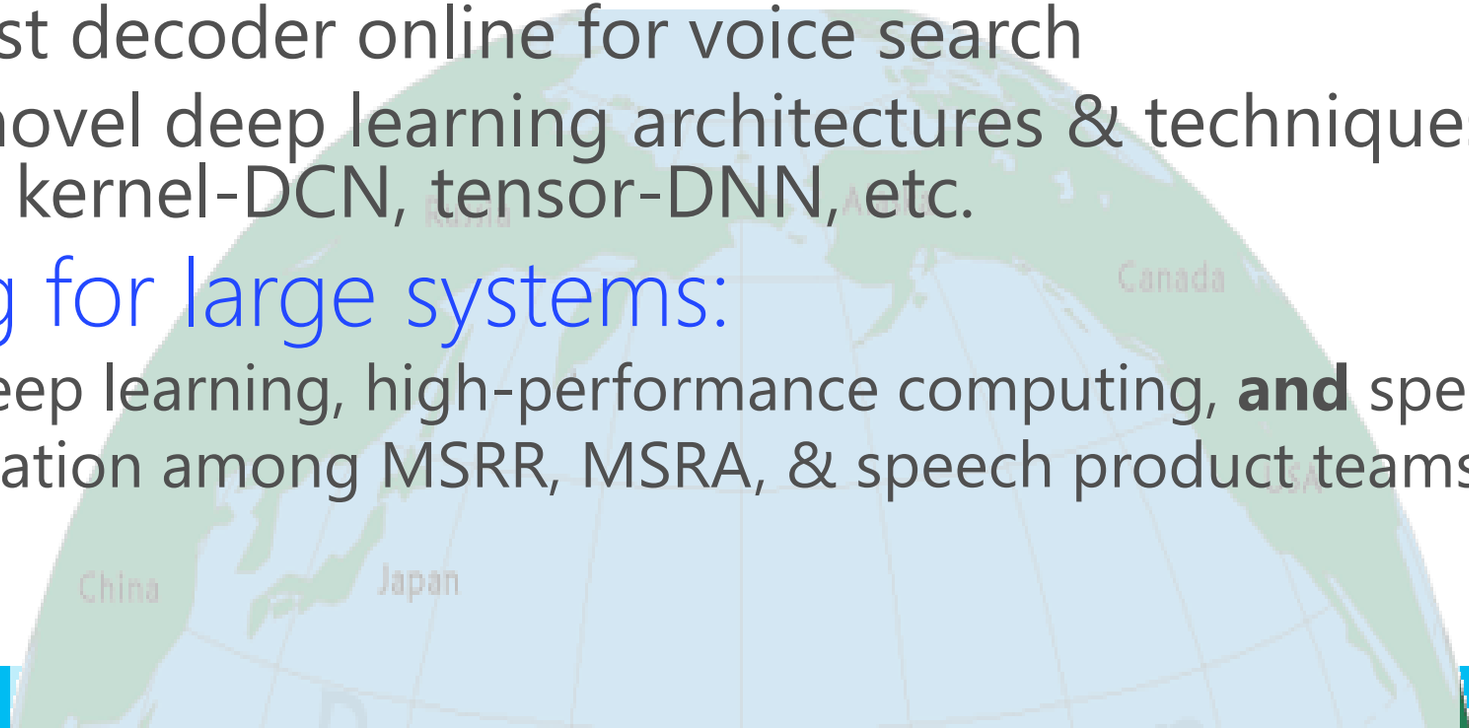
This paper reports our recent exploration of the layer-by-layer learning strategy for training a multi-layer generative model of patches of speech spectrograms. The top layer of the

The work reported in this paper was inspired by the successful use of deep auto-encoders for dimensionality reduction [8][9] and the extension of this work to the discovery of efficient binary codes in information retrieval [12]. It is also motivated by the potential benefits of using



Innovations at MSR (2009-2013)

- Scale early successes to industry-scale speech recognition tasks
  - Grow output neurons from small context-independent states to large context-dependent ones
  - Motivated initially by saving MSFT investment in huge speech decoder software infrastructure (hard to appreciate by academia)
  - Extremely fast decoder online for voice search
  - Developed novel deep learning architectures & techniques: DCN/DSN, tensor-DSN, kernel-DCN, tensor-DNN, etc.
- Engineering for large systems:
  - Expertise in deep learning, high-performance computing, **and** speech recognition
  - Close collaboration among MSRR, MSRA, & speech product teams, plus **academic interns!**



Our Speakers/Panelists Today

WIRED

Researcher Dreams Up Machines
That Learn Without Humans
06.27.13

Yoshua Bengio, University of Montreal
(Tutorial: 20 mins)

WIRED

Andrew Ng, Stanford University

The Man Behind the Google Brain: Andrew Ng and the Quest for the New AI

BY DANIELA HERNANDEZ 05.07.13 6:30 AM

Our Speakers/Panelists Today

Ruslan Salakhutdinov, University of Toronto

Honglak Lee, University of Michigan

Agenda

Short tutorial by Prof. Bengio, followed by moderated Panel Discussions on topics of:

1. What kinds of AI problems are good for deep learning paradigms to solve now? Which other aspects of AI need to be tackled to go further?
2. Relationship among deep learning, probabilistic graphical models, and other machine learning methods/models?
3. What is the status of supervised vs. unsupervised deep learning?
4. What is the role of scalability in deep learning?
5. What is the right infrastructure for deep learning? GPU/CPU clusters?
6. Review of current progress and challenges
7. Future directions of deep learning, etc.

including short presentations by panelists on these and related topics.

Audience participation is welcome!