AI Frontiers 2017

# Schedule

8:50 – 9:00 AM > Opening Remarks

9:00 – 9:50 AM > Keynote Speech

9:50 – 11:10 AM > Autonomous Driving

11:10 – 12:30 PM > Speech-Enabled Assistants

12:30 – 1:30 PM > Lunch

1:30 – 1:50 PM > Natural Language Processing

1:50 – 2:50 PM > Computer Vision

2:50 – 3:05 PM > McKinsey Report

3:05 – 3:30 PM > Break

3:30 – 4:40 PM > Internet of Things

4:40 – 6:00 PM > Deep Learning Frameworks

6:30 – 8:30 PM > Banquet

# 9:00 – 9:50 AM > Key Note

* http://g.co/brain
* Google Brain Team
  + Research
  + Product Impact
  + Open Source Tools for ML
    - Tensor Flow
    - Magenta
    - Visualization tools
* NEED TO BUILD THE RIGHT TOOLS
  + => Enables rapid research
* What do you want in a ML sys?
  + Ease of expression
  + Scalability
  + Portability
  + Reproducibility
  + Production readiness
* TensorFlow
  + Why built?
    - Wanted flexible, scalable, production-ready
  + Goals
    - Est common platform ML ideas & systems
    - Best in the world for research and production
  + Facts and Figures
    - Auto differentiation, queues, control flow, comprehensive set of operations
    - Tutorials made system accessible
  + Stats
    - 500+ contributors
    - 12,000+ commits
    - 1M+ binary downloads
    - #15 most popular repo on GitHub
    - Used in ML classes at Toronto, Berkeley, Stanford
    - Used by: Google, DeepMind, OpenAI, Twitter, Snapchat, Airbus, Uber
* Just-In-Time Compilation via TF’s XLA, “Accelerated Linear Algebra” compiler
  + Demo: Inspect JIT Code in TensorFlow iPython Shell
* Deep Learning Impact at Google
  + Speech Recognition
    - Acoustic Input -> Deep Recurrent NN -> Text Output
      * Reduced errors by 30%+
  + Google Photo Search
    - Your Photo -> Deep Convolutional NN -> Automatic Tag
      * Search personal photos without tags
* Sequence-to-Sequence Model: Machine Translation
  + Sutskever & Vinyals & Le NIPS 2014
* Image Captions Research
* Translation as a sign of better language
  + Great quality improvements -> BUT challenging scalability issues
  + Google Neural Machine
  + Google’s Multilingual Neural Network
* Automated Machine Learning
  + Current:
    - Solution = ML Expertise + data + computation
  + Can we turn this to
    - Solution = data + 100x computation
  + **Neural Architecture Search with Reinforcement Learning, Google Brain**
    - Paper to look up!!
  + CIFAR-10 Image Recognition Task
  + NNs
    - Reduced precisions OK
    - Handful of spec computations
    - Allows specialized computations for NN
  + Tensor Processing Unit
* Example Queries of the Future
  + Which of these eye images shows symptoms of diabetic retinopathy
  + Describe this video in Spanish
  + Please fetch me a cup of tea from the kitchen
  + Find me documents related to reinforcement learning for robotics and summarize them in German
* Conclusions
  + Deep NNs are making significant strides

# 9:50 – 11:10 AM > Autonomous Driving

## CrowdFlower – How a Driverless Car Sees the World

* Progress
  + 2009 – 2014
  + Shifting focus from highway to city street
  + Prototype cars announced in 2014
  + Lombard Street Video
    - Fully able to navigate through it
* Project Outline
  + Map everything we can
  + Create simple object representations
  + Don’t try to understand what others are doing
  + Avoid vision, it is just not robust enough

## Tesla, Autonomous Driving Revolution, Junli Gu

* Outline
  + Vehicle Revolution
  + Challenges: When Big Data meets ML in the car
    - Sensor System collects Big Data
    - ML perceives the real world
    - Car is a Digital Agent
  + Conclusion
* Transportation Evolution History
  + P0 Primitive
  + P1 Analog Device
  + P2 Digitalization
  + P3 Intelligence
* A revolution from analog to digital to intelligence
  + All parts are digital controllable
  + Complex sensor systems
  + Driving through perception
  + Controlled by computers
  + A car will become an agent with autonomous behavior
* Big Data meets ML in the Car
  + Real world is a big data problem
  + Driving in human world required intelligent perception in of the world
* Hybrid sensor system (SS) collects Big Data
  + SS composed of cameras, radar, ultrasound, lidar
  + Sensor fusion technology is not mature yet
  + Different sensor data is mostly computed separated
* Case study in combining sensors
  + Google self drivin gcars
  + Radar info is too thin to differentiate objects of same size, same speed
  + Cameras see rich semantics
    - WYSIWYG
* Machine learning perceives the real world
  + Large scale for 2D object recognition is better than human
    - WHAT objects are around
  + 3D scene understanding and modeling
    - WHERE is the object
  + Semantic segmentation
    - EXTENTof the obstacles
  + Reinforcement learning
    - POLICY (reward or penalty based learning)
  + End to end learning
    - From raw data direct to behavior, like a robot
* Deep learning based 2D image recognition
  + Large scale object recognition 98% accuracy
* 3D scene understanding
  + Real world is 3D
  + Depth information is critical for driving
  + 3D model more complicated
* Semantic segmentation
  + Understand the extent of the object and each pixel of it
  + Advanced requirement based on recognition and detection
  + Eventually 3D world segmentation
* Reinforcement learning
  + Goal: to win
    - Policy based learning
  + Learning task: Next move at arbitrary states
  + Algorithm: reinforcement learning + DNN
  + Similar methodology can be applied to driving
    - Driving policy: Given safety, spend least time to get to destination
* End to end learning
  + Use ML as the only step from raw data to control
    - Without human interference noise in the middle steps
  + DaveNet: AI teaches the car how to drive
* Other challenges in the car
  + Embedded computers in the car
  + Cloud solution
    - Agent+ system
    - Challenges: network reliability and real time response
  + Vehicle external
  + High speed internal interconnect
* Conclusions
  + Cars are transforming from analog to digitalization and intelligence
  + When big data meets ML in the car
    - Complex SS collects the big data info
    - ML algorithms will be the key to perceive the real world
    - Computing in the car is another challenge
    - Cars will be connected to each other and the cloud
  + Whoever addresses the technical challenge will harvest the influence

## Baidu, Head of Self-Driving Cars

* Three Laws of Autonomous Driving
  + 1. Safety
  + 2. Efficiency
  + 3. Economy
* Graph: Internet leads the breakthrough of AI
  + Deep learning permeates the internet
* Perception Performance
  + 95% Pedestrian Detection
  + 99.9% Signal Light Recognition

## Autonomous Cars Panel

* Challenge:
  + Tesla: Many algorithms and issues with control flow and cross-domains
  + Baidu: Don’t know too much about cars

# 11:10 – 12:30 PM > Speech-Enabled Assistants

## Three Generations of Spoken Dialogue Systems (Bots), Li Deng, Chief Scientist of AI, Microsoft

* Conversational UIs, Bots, Dialogue Systems
  + Speech-based vs text-based
    - Spoken dialogue system = | != speech recognition + text-based dialogue system
      * Integrated learning vs simply pipelined
      * Para-linguistic cues in speech signal (prosody, emotion, speaker, etc)
      * Depending on users
  + Narrow domain vs wide domains
  + Speech recognition maturing with DL
    - Spoken systems no longer limited to narrow domains
  + 3 generations of dialogue technology since early 90s
* Diagram: Bots Landscape
* Diagram: AI Tools: NLP, ML, Speech & Voice Recognitions
* \*\* Article: How deep reinforcement learning can help chatbots (Li Deng)
* Generation I: Symbolic Rule/ Template Based
  + Centered on grammatical rule & ontological design by human experts
  + Easy interpretation, debugging and system update
  + Popular before late 90s
  + Still in use in commercial systems and by bots startups
  + Limitations:
    - Reliance on experts => \*\* EXPERT SYSTEMS DESIGN!
    - Hard to scale over domains
    - Data used only to help design rules, not for learning
  + Example system next
* Fig. 3. A generic block diagram for a typical conversational interface
* Generation II: Data Driven, (shallow) Learning
  + Data used not to design rules of NLU and action, but to learn statistical parameters in dialogue systems
    - Reduce cost of hand-crafting complex dialogue manager
    - Robustness against speech recognition errors in noisy environment
    - MDP/POMDP
    - Discriminative (CRF) & generative (HMM) methods for NLU
    - Popular in academic research until 2015 (before DL arrived at dialogue world); in parallel with Generation (BBN, AT&T, etc)
  + Paper: POMDP-based statistical spoken dialogue systems: A Review, Steve Young, Cambridge University
* Generation III: Data-Driven Deep Learning
  + Like Gen-II, data used to learn everything in dialogue systems
* What is reinforcement learning
* Stateful Model for RL
* Deep Reinforcement Learning to Optimized Dialogue Policy
  + Can formulate all 3 types of bots into reverse learning
* Research Frontiers
  + Speech-based vs text-based
    - Errors in speech recognition NOT just as “noise”
* Contact him for slides!
  + Lots of great references!

## AI For 100 Million People with Deep Learning, Adam Coates, Baidu Research

* First goal: Speech recognition everywhere
* Speech will transform mobile device interface
* Speech recognition: Traditional ASR
  + Traditional speech systems build on standard ML + engineering practices
    - Features -> Acoustic Model -> Sequence Model -> Language Model -> Transcription
    - Features: Compute Features
    - Acoustic Model: Predict Phonemes
    - Sequence Model: Combine into state sequence
    - Language Model: Merge with pronunciation and language data
    - Transcription
* Deep Learning
  + Major advantage of deep learning: scalability
* “Deep Speech”
  + Pour effort into data + computation
    - Try to catch up to scalability
  + Train giant neural networks to predict characters from audio
    - Train “end to end” [e.g. Graves et al., 2006]
  + Hard part is training at scale and searching for best model
  + Need data + computing power
* Raw Training Data
  + Need lots of data for end to end DL systems: use read speech
* Data augmentation
  + Many forms of distortion that model should be robust to:
    - Reverb, noise, far field effects, echo, compression artifacts, changes in tempo
  + Learn to be robust by training from data with distortions
    - Easier to engineer data pipeline than to engineer recognition pipeline
* Augmented Dataset
  + Augmentation greatly expands
* Compute
  + 1 experiment requires 10s of ExaFLOPs)
    - Titan X GPU: ~6 TeraFLOPS
  + “Speed of light” = 3-6 weeks on 1 GPU
  + Scale out to large numbers of GPUs (e.g. 8-64)
* Deep Speech for Mandarin
  + Deep Speech is driven by data

12:30 – 1:30 PM > Lunch

1:30 – 1:50 PM > Natural Language Processing

1:50 – 2:50 PM > Computer Vision

2:50 – 3:05 PM > McKinsey Report

3:05 – 3:30 PM > Break

3:30 – 4:40 PM > Internet of Things

4:40 – 6:00 PM > Deep Learning Frameworks

6:30 – 8:30 PM > Banquet