Powering AI Robots with Deep Learning

NVIDIA Deep Learning Day 2017
COEX Intercontinental Hotel, Seoul, Korea, Tuesday, October 31, 2017

Byoung-Tak Zhang

Seoul National University
School of Computer Science and Engineering &
Surromind Robotics, Inc.
http://bi.snu.ac.kr/
Outline

AI Robots Come of Age  ........................................ 3
  ▪ Autonomous Mobile Robots
  ▪ Home Robots, Human-Like Robots

Deep Learning for AI Robots  ......................... 12
  ▪ Perception, Action, Cognition
  ▪ RoboCup@Home Challenge

New AI  ................................................................. 30
1. AI Robots Come of Age
Humans (NI) and Machines (AI)

- **Introspectionism** - 1920
  
  **Psyche**

- **Behaviorism** ↔ **Cybernetics** 1920 - 1950

  **Mind (= Computer)**

- **Cognitivism** ↔ **Symbolic AI** 1950 - 1980

  **Brain**

- **Connectionism** ↔ **Neural Nets (ML)** 1980 – 2010

  **Body**

- **Action Science** ↔ **Autonomous Robots** 2010 –

**Environment**

*Embodied Mind | Mind Machine (= Smart Machine)*
Early AI Robots
(Autonomous Mobile Robots)

Shakey (SRI)

RHINO (U. Bonn)

Cart (Stanford Univ.)

CoBot (CMU)
RoboCup (1997~)
Home Robots

PR2 Fetches Beer (Willow Garage)

Dash at Hotel (Sevioke)

PR2 Making Popcorns (TU Munich)

SpotMini (Boston Dynamics)
Human-Like Robots

Humanoid Robot Nao (Aldebaran)

Life-Like Robots (Hanson Robotics)

Emotion Robot Pepper (SoftBank)

Atlas (Google Boston Dynamics)
Robot Life in a City

Obelix (University of Freiburg, Germany)

https://www.youtube.com/watch?v=gPzC88HkgcU&t=80s
AI Robots for the 4th Industrial Revolution

“Cognitive” Smart Machines

Body (HW, Device)

Mind (SW, Data)
Enabling Technologies for AI Robots

- **Perception**
  - Object recognition
  - Person tracking

- **Control**
  - Manipulation
  - Action control

- **Navigation**
  - Obstacle avoidance
  - Map building & localization

- **Interaction**
  - Vision and voice
  - Multimodal interaction

- **Computing Power**
  - Cloud computing
  - GPUs, parallel computing
  - Neural processors
2. Deep Learning for AI Robots
Traditional Machine Learning vs. Deep Learning

Traditional Machine Learning Flow

Deep Learning Flow
Deep Learning Revolution

- Big Data + Parallel Computing + Deep Learning
- From programming to learning
- Automated- or self-programming
- Paradigm shift in S/W
- Self-improving systems
- Intelligence explosion
**Power of Deep Learning**

- Multiple boundaries are needed (e.g. XOR problem)  
  ➔ Multiple Units

- More complex regions are needed (e.g. Polygons)  
  ➔ Multiple Layers

---

### Big Data + Deep Learning = Automatic Programming
AI / Deep Learning Growth

NIPS Growth

Total Registrations: 3755

- Tutorials: 2,584
- Conference: 3,262
- Workshops: 3,006

Year-wise AI Acquisition Trends:
- 2011: 6
- 2012: 9
- 2013: 16
- 2014: 23
- 2015: 34
- 2016 YTD: 21

Major AI Acquirers:
- Google
- Apple
- Salesforce
- Yahoo
- Intel

Growing Use of Deep Learning at Google:
- Across many product lines: Android, Apps, drug discovery, Gmail, Image understanding, Maps, Natural language understanding, Photos, Robotics research, Speech, Translation, YouTube, and many others.
Deep Learning for Voice and Dialogue

- Speech LSTM-RNN (Recurrent Neural Networks)
- End-to-End Memory Networks (N2N MemNet)
- CNN + RNN for Question Answering


Interaction: Conversational Interface

Amazon Echo  Google Home  SKT Nugu
Deep Learning for Robotic Grasping

(Levine et al, 2016)

(C) 2015-2016, SNU Biointelligence Lab, http://bi.snu.ac.kr/
Deep Reinforcement Learning for Action Control

BRETT (Univ. of California, Berkeley)
ImageNet Large-Scale Visual Recognition Challenge

- Image Classification/Localization
- 1.2M labeled images, 1000 classes
- Convolutional Neural Networks (CNNs) has been dominating the contest since..
  - 2012 non-CNN: 26.2% (top-5 error)
  - 2012: (Hinton, AlexNet) 15.3% (Using GPUs)
  - 2013: (Clarifai) 11.2%
  - 2014: (Google, GoogLeNet) 6.7%
  - (pre-2015): (Google) 4.9%
    - Beyond human-level performance
Use 3D CNNs to model the temporal patterns as well as the spatial patterns.
Deep Learning for Autonomous Driving (NVIDIA)
Visual Question Answering (VQA) is a new dataset containing open-ended questions about images. These questions require an understanding of vision, language and common sense to answer.

- 254,721 images (MSCOCO and abstract scenes)
- 3 questions per image (764,163 total)
- 10 ground truth answers per question
- 3 plausible (but likely incorrect) answers per question
- Open-ended and multiple-choice answering tasks

Winner (UC Berkeley & Sony) 66.9% accuracy on real-image open-ended QA. Naver, Samsung, SNU, Postech (현재 SNU 1등 중)
Cambot (SNU)

[Kim et al., NIPS-2016]
Deep Hypernets for Visual Dialogue

- Deep hypernetworks with hierarchical concept structure are used as knowledge base for Q&A

Hierarchical formulation

\[ P(x) = \sum_{h_n} \cdots \sum_{h_1} P(h_n \mid h_{n-1}) \cdots P(h_2 \mid h_1) P(h_1 \mid x) P(x) \]

Joint probability of hidden variables \( h_s^{(s)} \) in the \( s \)th layer

\[ P(h_s) = \frac{\exp(-E(h_s))}{\sum_j \exp(-E(h_j))} \]

\[ E(h_j) = h(s(h_j)) \]

\[ s(h_j) = \sum_i w_i^{(j)} h_i^{(j)} + \sum_{i \neq j} w_i^{(j)} h_i^{(j)} + \cdots + \sum_{i \neq j} w_i^{(j)} h_i^{(j)} \]

Learning is done by adjusting \( s(h_j) \) towards maximizing likelihood \( P(x \mid W) \)

\[
\nabla_{W_{h_1}^{(k)}, \ldots, h_k} \ln \prod_{n=1}^{N} P(x^{(n)} \mid W) = \nabla \left[ \left[ x_{n_1}, x_{n_2}, \ldots, x_{n_k} \right]_{\text{Data}} - \left[ x_{n_1}, x_{n_2}, \ldots, x_{n_k} \right]_{P(x\mid W)} \right]
\]

Learning from Cartoon Videos

Image 개수: 3,600
Word 개수: 1,900
Episode 개수: 160

[Ha et al., AAAI-2015]
Pororobot (SNU)

[Ha et al., AAAI-2015]
AUPAIR: Autonomous Personal AI Robot

AUPAIR (SNU & Surromind Robotics)

Winning the RoboCup@Home 2017

https://www.youtube.com/watch?v=a2ZJTpbMWsQ  http://mnews.joins.com/article/21823070#home
3. New AI
Human Intelligence and Artificial Intelligence

1. The accelerating pace of change...
   - Agricultural Revolution: 8,000 years
   - Industrial Revolution: 120 years
   - Light bulb: 90 years
   - Moon landing: 22 years
   - World Wide Web: 9 years
   - Human genome sequenced: Present

2. ...and exponential growth in computing power...
   - Computer technology has progressed dramatically by powers of 10, now doubling more each hour than it did in its entire first 90 years.

3. ...will lead to the Singularity
   - Apple II: The compact machine was one of the first massively popular personal computers.
   - Power Mac G4: The first personal computer to deliver more than 1 billion floating-point operations per second.

COMPUTER RANKINGS

- Analytical engine: Never fully built, Charles Babbage’s invention was designed to solve computational and logical problems.
- Colossus: The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WWII.
- UNIVAC I: The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.

By calculations per second per $1,000:
- 1900: 100
- 1920: 200
- 1940: 1,000
- 1960: 10,000
- 1980: 100,000
- 2000: 1,000,000
- 2045: 10,000,000,000

Notable developments:
- 2045: Surpasses brainpower equivalent to that of all human brains combined.
- 2023: Surpasses brainpower of human brain.
- 2015: Surpass brainpower of mouse.
Dual Process Theories of Mind

Kahneman, *Thinking, Fast and Slow* (2011)

- **System 1**
  - Automatic, Fast, Unconscious
  - Heuristic, “Hot”
    - Emotions, Stereotypes
- **System 2**
  - Controlled, Slow, Conscious
  - Algorithmic, “Cold”
    - Logical, Systematic
- **System 1 Usually Wins the Race**
## New AI (System 1) and Old AI (System 2)

<table>
<thead>
<tr>
<th>System 1</th>
<th>System 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconscious reasoning</td>
<td>Conscious reasoning</td>
</tr>
<tr>
<td>Judgments based on intuition</td>
<td>Judgments based on critical examination</td>
</tr>
<tr>
<td>Processes information quickly</td>
<td>Processes information slowly</td>
</tr>
<tr>
<td>Hypothetical reasoning</td>
<td>Logical reasoning</td>
</tr>
<tr>
<td>Large capacity</td>
<td>Small capacity</td>
</tr>
<tr>
<td>Prominent in animals and humans</td>
<td>Prominent only in humans</td>
</tr>
<tr>
<td>Unrelated to working memory</td>
<td>Related to working memory</td>
</tr>
<tr>
<td>Operates effortlessly and automatically</td>
<td>Operates with effort and control</td>
</tr>
<tr>
<td>Unintentional thinking</td>
<td>Intentional thinking</td>
</tr>
<tr>
<td>Influenced by experiences, emotions, and memories</td>
<td>Influenced by facts, logic, and evidence</td>
</tr>
<tr>
<td>Can be overridden by System 2</td>
<td>Used when System 1 fails to form a logical/acceptable conclusion</td>
</tr>
<tr>
<td>Prominent since human origins</td>
<td>Developed over time</td>
</tr>
<tr>
<td>Includes recognition, perception, orientation, etc.</td>
<td>Includes rule following, comparisons, weighing of options, etc.</td>
</tr>
</tbody>
</table>
Humans (NI) and Machines (AI)

- Introspectionism  - 1920

Psyche
- Behaviorism ↔ Cybernetics  1920 - 1950
- Mind (= Computer)
- Cognitivism ↔ Symbolic AI  1950 - 1980

Brain
- Connectionism ↔ Neural Nets (ML)  1980 – 2010

Body
- Action Science ↔ Autonomous Robots  2010 –

Environment

*Embodyed Mind | Mind Machine (= Smart Machine)*
Autonomous Machine Learning

1G: Supervised Learning (1980s~2000)

- Decision Trees
- Kernel Methods
- Multilayer Perceptrons

2G: Unsupervised Learning (2000~Present)

- Deep Networks
- Markov Networks
- Bayesian Networks

3G: Autonomous Learning (Next Generation)

- Complex Adaptive Systems
- Perception-Action Cycle
- Lifelong Continual Learning

Future of AI

Narrow AI

AI with Deep Learning

Cognitive AI

Human-Level AI

Superhuman AI

Free Will

Agency

Time

Parallel Computing

Autonomous

Technology

(Embodied Brain-Like)

Sequential

Reactive

1980  1990  2010  2020  2030  2050

Follows given goals and methods

Works out own methods, follows given goals

Works out own goals

Follows given goals and methods

Modified from Eliezer Yudkowsky & David Wood