Cognitive Learning and the Multimodal Memory Game: Toward Human-Level Machine Learning

2008 IEEE World Congress on Computational Intelligence (WCCI 2008)
Cognitive Architectures: Towards Human-Level Intelligence Session
June 5, 2008, Hong Kong

Byoung-Tak Zhang

Biointelligence Laboratory
School of Computer Science and Engineering
Cognitive Science, Brain Science, and Bioinformatics Programs
Seoul National University
Seoul 151-744, Korea

btzhang@cse.snu.ac.kr
http://bi.snu.ac.kr/
Talk Outline

- Human-level machine learning is a prerequisite to achieving human-level machine intelligence.
  - Differences of behaviors in humans and machines
- What principles are underlying the cognitive learning and memory in humans?
  - A proposal for three principles
- What tasks are challenging enough to study human-level machine learning?
  - A proposal for the multimodal memory game (MMG)
- Some illustrative results
  - Linguistic memory
  - Language-vision translation
- Future directions
  - Toward human-level machine intelligence
Cognitive Learning
Humans and Machines

- Humans are
  - creative,
  - compliant,
  - attentive to change,
  - resourceful, and
  - multipurpose

- Humans are
  - imprecise,
  - sloppy,
  - distractable,
  - emotional, and
  - illogical

- To achieve human-level intelligence these properties should be taken into account.
Toward Human-Level Intelligence

- Human intelligence develops situated in a multimodal environment [Gibbs, 2005].
- The human mind makes use of multiple representations and problem-solving strategies [Fuster, 2003].
- The brain consists of functional modules which are localized in subcortical areas but work together on the whole-brain scale [Grillner et al., 2006].
- Humans can integrate the multiple tasks into a coherent solution [Jones, 2004].
- Humans are versatile and come up with many new ideas and solutions to a given problem [Minsky, 2006].
Learning and Memory as a Substrate for Intelligence

It is our memory that enables us to value everything else we possess. Lacking memory, we would have no ability to be concerned about our hearts, achievements, loved ones, and incomes. Our brain has an amazing capacity to integrate the combined effects of our past experiences together with our present experiences in creating our thought and actions. This is all possible by the memory and the memories are formed by the learning process.

Principles of Learning: Early Ideas

- Aristotle: Three Laws of Association [Crowder, 1976]
  - Similarity
  - Contrast
  - Contiguity
- James Mill (1773-1836): Strength Criteria of Association
  - Permanence
  - Certainty
  - Spontaneity
  - “Mental Compounding”
- John Stuart Mill (1806-1873)
  - “Mental Chemistry”
Principles of Learning: Modern Concepts

- Types of learning: Accretion, tuning, restructuring (e.g., Rumelhart & Norman, 1976)
- Encoding specificity principle (Tulving, 1970’s)
- Cellular and molecular basis of learning and memory (Kandel et al., 1990’s)
- Conceptual blend and chemical scramble (e.g., Feldman, 2006)
Methods of Machine Learning

- Symbolic Learning
  - Version Space Learning
  - Case-Based Learning
- Neural Learning
  - Multilayer Perceptrons
  - Self-Organizing Maps
  - Support Vector Machines
- Evolutionary Learning
  - Evolution Strategies
  - Evolutionary Programming
  - Genetic Algorithms
  - Genetic Programming
- Probabilistic Learning
  - Bayesian Networks
  - Helmholtz Machines
  - Latent Variable Models
  - Generative Topographic Mapping
- Other Machine Learning Methods
  - Decision Trees
  - Reinforcement Learning
  - Boosting Algorithms
  - Kernel Methods
  - Independent Component Analysis
Three Fundamental Principles of Cognitive Learning: Our Proposal

- Continuity. Learning is a continuous, lifelong process. “The experiences of each immediately past moment are memories that merge with current momentary experiences to create the impression of seamless continuity in our lives” [McGaugh, 2003]

- Glocality. “Perception is dependent on context” and it is important to maintain both global and local, i.e. glocal, representations [Peterson and Rhodes, 2003]

- Compositionality. “The brain activates existing metaphorical structures to form a conceptual blend, consisting of all the metaphors linked together” [Feldman, 2006]
Research Platform for Cognitive Learning
Toward Human-Level Machine Learning: Multimodal Memory Game (MMG)

But, I'm getting married tomorrow. Well, maybe I am... I keep thinking about you. And I'm wondering if we made a mistake giving up so fast. Are you thinking about me? But if you are, call me tonight.

Text Generation Game (from Image)

Image Generation Game (from Text)

Learning by Viewing

Game Manager

I2T

T2I

Image

Sound

Text

Some Experimental Results
Three Experiments

● Sentence Generation
  ♦ Learn: a linguistic recall memory from a sentence corpus
  ♦ Given: a partial or corrupt sentence
  ♦ Generate: a complete sentence

● Image-to-Text Translation
  ♦ Learn: an image-text joint model from an image-text pair corpus
  ♦ Given: an image (scene)
  ♦ Generate: a text (dialogue of the scene)

● Text-to-Image Translation
  ♦ Learn: an image-text joint model from an image-text pair corpus
  ♦ Given: a text (dialogue)
  ♦ Generate: an image (scene of the dialogue)
Experiment 1: Learning Linguistic Memory

- Dataset: scripts from dramas
  - Friends
  - House
  - 24
  - Grey Anatomy
  - Gilmore Girls
  - Sex and the City
- Training data: 289,468 sentences
- Test data: 700 sentences with blanks
- Vocabulary size: 34,219 words

## Sentence Completion Results

<table>
<thead>
<tr>
<th>Original Sentence</th>
<th>Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>? gonna ? upstairs ? ? a shower</td>
<td>I'm gonna go upstairs and take a shower</td>
</tr>
<tr>
<td>? have ? visit the ? room</td>
<td>I have to visit the ladies' room</td>
</tr>
<tr>
<td>? still ? believe ? did this</td>
<td>I still can't believe you did this</td>
</tr>
<tr>
<td>? ? ? decision</td>
<td>to make a decision</td>
</tr>
<tr>
<td>? appreciate it if ? call her by ? ?</td>
<td>I appreciate it if you call her by the way</td>
</tr>
<tr>
<td>Would you ? to meet ? ? Tuesday ?</td>
<td>Would you nice to meet you in Tuesday and</td>
</tr>
<tr>
<td>Why ? you ? come ? down ?</td>
<td>Why are you go come on down here</td>
</tr>
<tr>
<td>We ? ? a lot ? gifts</td>
<td>We don't have a lot of gifts</td>
</tr>
<tr>
<td>? ? don't need your ?</td>
<td>if I don't need your help</td>
</tr>
<tr>
<td>? ? a dream about ? In ?</td>
<td>I had a dream about you in Copenhagen</td>
</tr>
<tr>
<td>What ? ? ? here</td>
<td>What are you doing here</td>
</tr>
<tr>
<td>? you ? first ? of medical school</td>
<td>Are you go first day of medical school</td>
</tr>
<tr>
<td>I'm standing ? the ? ? cafeteria</td>
<td>I'm standing in the one of the cafeteria</td>
</tr>
<tr>
<td>? think ? I ? met ? somewhere before</td>
<td>I think but I am met him somewhere before</td>
</tr>
</tbody>
</table>
Experiments 2 & 3: Crossmodal Translation

- Dataset: scenes and corresponding scripts from two dramas
  - Friends
  - Prison Break
- Training data: 2,808 scenes and scripts
- Scene (image) size: $80 \times 60 = 4800$ binary pixels
- Vocabulary size: 2,579 words

- The order ($k$) of hyperedge
  - Text: Order 2~4
  - Image: Order 10~340
- The method of creating hyperedges from training data
  - Text: Sequential sampling from a randomly selected position
  - Image: Random sampling in 4,800 pixel positions
- Number of samples from an image-text pair
  - From 150 to 300

Where am I giving birth
I know it's been really hard for you
So when you guys get in there

Image-to-Text Translation Results

Query: There's a kitty in my guitar case
Completion: There's a kitty in ...
Completion: in my guitar case
Answer: There's a kitty in my guitar case

Query: Maybe there's something I can do to make sure I get pregnant
Completion: Maybe there's something I ...
Completion: ... I get pregnant
Answer: Maybe there's something I can do to make sure I get pregnant

I don't know what happened
I don't know what happened

Text-to-Image Translation Results

<table>
<thead>
<tr>
<th>Query</th>
<th>Matching &amp; Completion</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don't know what happened</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Take a look at this</td>
<td></td>
<td></td>
</tr>
<tr>
<td>There's a kitty in my guitar case</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maybe there's something I can do to make sure I get pregnant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypernetwork Architecture for Cognitive Learning
Learning

4 Data Items

Round 3
Hypernetwork of DNA Molecules

[Zhang, DNA-2006]
Hypernetwork as Chemical Associative Memory

The hypernetwork is defined as
\[ H = (X, S, W) \]
\[ X = (x_1, x_2, \ldots, x_j) \]
\[ S = \sum_i S_i, \quad S_i \subseteq X, \quad k = |S_i| \]
\[ W = (W^{(2)}, W^{(3)}, \ldots, W^{(K)}) \]

Training set:
\[ D = \{ x^{(n)} \}_{j=1}^N \]

The energy of the hypernetwork
\[ E(x^{(n)}; W) = -\frac{1}{2} \sum_{i_1, i_2} W^{(2)}_{i_1 i_2} x^{(n)}_{i_1} x^{(n)}_{i_2} - \frac{1}{6} \sum_{i_1, i_2, i_3} W^{(3)}_{i_1 i_2 i_3} x^{(n)}_{i_1} x^{(n)}_{i_2} x^{(n)}_{i_3} - \ldots \]

The probability distribution
\[ P(x^{(n)} | W) = \frac{1}{Z(W)} \exp[-\beta E(x^{(n)}; W)] \]
\[ = \frac{1}{Z(W)} \exp \left[ \frac{1}{2} \sum_{i_1, i_2} W^{(2)}_{i_1 i_2} x^{(n)}_{i_1} x^{(n)}_{i_2} + \frac{1}{6} \sum_{i_1, i_2, i_3} W^{(3)}_{i_1 i_2 i_3} x^{(n)}_{i_1} x^{(n)}_{i_2} x^{(n)}_{i_3} + \ldots \right] \]
\[ = \frac{1}{Z(W)} \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_1, i_2, \ldots, i_k} W^{(k)}_{i_1 i_2 \ldots i_k} x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_k} \right], \]
where the partition function is
\[ Z(W) = \sum_{x^{(n)}} \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_1, i_2, \ldots, i_k} W^{(k)}_{i_1 i_2 \ldots i_k} x^{(m)}_{i_1} x^{(m)}_{i_2} \ldots x^{(m)}_{i_k} \right] \]

[Zhang, DNA-2006]

For more details: Zhang, B.-T., IEEE Computational Intelligence Magazine, August 2008 (in press)

Image to Text (Recall Rate)

Recall Rate

Rate

Perfect Recall

Tolerant Recall

Image Order

Text to Image (Recall Rate)

Toward Human-Level Intelligence
From Mind to Molecules and Back

Mind

Brain

Cell

Molecule

∞ memory

10^{11} cells

>10^3 molecules

Paradigms for Computational Intelligence

<table>
<thead>
<tr>
<th></th>
<th>Symbolism</th>
<th>Connectionism</th>
<th>Dynamicism</th>
<th>Hyperinter-actionism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metaphor</strong></td>
<td>symbol system</td>
<td>neural system</td>
<td>dynamical system</td>
<td>biomolecular system</td>
</tr>
<tr>
<td><strong>Mechanism</strong></td>
<td>logical</td>
<td>electrical</td>
<td>mechanical</td>
<td>chemical</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>syntactic</td>
<td>functional</td>
<td>behavioral</td>
<td>relational</td>
</tr>
<tr>
<td><strong>Representation</strong></td>
<td>localist</td>
<td>distributed</td>
<td>continuous</td>
<td>collective</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>structural</td>
<td>connectionist</td>
<td>differential</td>
<td>combinatorial</td>
</tr>
<tr>
<td><strong>Adaptation</strong></td>
<td>substitution</td>
<td>tuning</td>
<td>rate change</td>
<td>self-assembly</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td>sequential</td>
<td>parallel</td>
<td>dynamical</td>
<td>massively parallel</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>procedure</td>
<td>network</td>
<td>equation</td>
<td>hypergraph</td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td>logic, formal</td>
<td>linear algebra,</td>
<td>geometry, calculus</td>
<td>graph theory,</td>
</tr>
<tr>
<td></td>
<td>language</td>
<td>statistics</td>
<td></td>
<td>probabilistic logic</td>
</tr>
<tr>
<td><strong>Space/time</strong></td>
<td>formal</td>
<td>spatial</td>
<td>temporal</td>
<td>spatiotemporal</td>
</tr>
</tbody>
</table>

Summary and Conclusion

- We argue that understanding and implementing the principles of cognitive learning and memory is a prerequisite to achieving human-level intelligence.
- Suggested three principles as the most fundamental to cognitive learning.
  - Continuity, glocality, compositionality
- Proposed the multimodal memory game (MMG) as a research platform for studying the architectures and algorithms for cognitive learning.
- Presented the hypernetwork model as a cognitive architecture for learning in an MMG environment.
- Showed some experimental results to illustrate the usefulness of the platform.
  - Linguistic recall memory or sentence completion
  - Language-vision crossmodal translation tasks
- Future work can extend the experimental setups in various dimensions, such as corpus size, kinds of modality, and learning strategies.
Derivation of the Learning Rule

\[
\frac{\partial}{\partial w_{i_1 \ldots i_s}^{(s)}} \ln P \left( \{x^{(n)}\}_{1}^{N} \mid W \right)
\]

\[
= \frac{\partial}{\partial w_{i_1 \ldots i_s}^{(s)}} \sum_{n=1}^{N} \left\{ \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_{i_1}, \ldots, i_{i_s}} w_{i_1 \ldots i_s}^{(k)} x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} \right] - \ln Z(W) \right\}
\]

\[
= \sum_{n=1}^{N} \left\{ \frac{\partial}{\partial w_{i_1 \ldots i_s}^{(s)}} \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_{i_1}, \ldots, i_{i_s}} w_{i_1 \ldots i_s}^{(k)} x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} \right] - \frac{\partial}{\partial w_{i_1 \ldots i_s}^{(s)}} \ln Z(W) \right\}
\]

\[
= \sum_{n=1}^{N} \left\{ x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} - \left\langle x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} \right\rangle_{P(x \mid W)} \right\}
\]

\[
= N \left\{ \left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{\text{Data}} - \left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{P(x \mid W)} \right\}
\]

where

\[
\left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{\text{Data}} = \frac{1}{N} \sum_{n=1}^{N} \left[ x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} \right]
\]

\[
\left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{P(x \mid W)} = \sum_{x} \left[ x_{i_1} x_{i_2} \ldots x_{i_s} P(x \mid W) \right]
\]
The number of completion increases while the number of training sentences become larger.
Completion of Two Missing Words (2/3)

- The number of completions increases until the number of missing words equals the order – 1. (ex) Orders 3 and 4
The number of completions rapidly decreases if the number of missing words becomes larger than the order. (ex) Orders 2 and 3.
Multimodal Memory Game as a Platform for Cognitive Machine Learning

Image

Sound

Text

I2T

Learning by Viewing

T2I
## Illustrative Results

<table>
<thead>
<tr>
<th>Query</th>
<th>Completion</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>who are you</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? are you</td>
<td>what are you</td>
<td>Friends</td>
</tr>
<tr>
<td>who ? you</td>
<td>who are you</td>
<td>Friends</td>
</tr>
<tr>
<td>who are ?</td>
<td>who are you</td>
<td>Friends</td>
</tr>
<tr>
<td>you need to wear it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? need to wear it</td>
<td>I need to wear it</td>
<td>24</td>
</tr>
<tr>
<td>you ? to wear it</td>
<td>you want to wear it</td>
<td>24</td>
</tr>
<tr>
<td>you need ? wear it</td>
<td>you need to wear it</td>
<td>24</td>
</tr>
<tr>
<td>you need to ? it</td>
<td>you need to do it</td>
<td>House</td>
</tr>
<tr>
<td>you need to wear ?</td>
<td>you need to wear a</td>
<td>24</td>
</tr>
</tbody>
</table>

Image to Text Translation

Image

Learning by Viewing

Text
- Where am I giving birth
- You guys really don't know anything
- So when you guys get in there
- I know it's been really hard for you
- ...

Where ? I giving ?

User

Answer:
Where am I giving birth

Text Corpus

Text to Image Translation

Image Corpus

Learning by Viewing

Text
- Where am I giving birth
- You guys really don't know anything
- So when you guys get in there
- I know it's been really hard for you
- ...

Question:
You've been there

Answer: