Beyond Deep Learning: Combining Neural Processing and Symbolic Processing

Hang Li
Toutiao AI Lab
Outline

• Deep Learning for Natural Language Processing
• Neural Symbolic Processing
• Related Work
• Our Work
• Summary
Major Tasks of Natural Language Processing

• Text classification (e.g., email spam detection)
• Sentiment analysis
• Machine translation
• Information extraction
• Question answering
• Summarization
• Dialogue
Fundamental Problems of Natural Language Processing

- Classification: assigning a label to a string
  \[ S \rightarrow C \]

- Matching: matching two strings
  \[ s, t \rightarrow R^+ \]

- Translation: transforming a string to another
  \[ S \rightarrow T \]

- Structured prediction: mapping a string to its structure given knowledge
  \[ S \rightarrow S', D \]

- Sequential decision process: outputting a string given a number of strings
Fundamental Problems of Natural Language Processing

• Classification
  – Text classification
  – Sentiment analysis

• Matching
  – Search
  – Question answering
  – Dialogue (single turn)

• Translation
  – Machine translation
  – Speech recognition
  – Handwriting recognition
  – Dialogue (single turn)

• Structured Prediction
  – Chinese word segmentation
  – Part of speech tagging
  – Named entity extraction
  – Dependency parsing
  – Semantic parsing

• Sequential Decision Process
  – Dialogue (multi turn, task dependent)
Deep Learning for Natural Language Processing

• Major tools: word embedding, deep networks (recurrent neural networks, convolutional neural networks)

• For the first four problems (classification, matching, translation, structured prediction), deep learning outperforms traditional approaches

• Machine translation: paradigm shift from Statistical MT to Neural MT
Advantages of Deep Learning

- High performances in many tasks
- End-to-end training: little or no domain knowledge is needed in model construction
- Representation learning: information processing across multi-modality becomes possible
- Learning of complex patterns: complicated language processing becomes more feasible
End-to-End Training

• One can build a system only with data without human involvement
• E.g., neural machine translation
• One can build a machine translation system from parallel corpora
Representation Learning

- Vector representations of words, phrases, and sentences
- Possible to process information across multi-modality
- E.g., image retrieval using CNNs
- Not possible before

Diagram:
- Convolutional Neural Network
- Image representation
- Text representation
- Matching
- A dog is catching a ball
Learning of Complex Patterns

- Model is deep (with multi-layers)
- Layered non-linear functions can capture complex patterns
- E.g., generation-based single-turn dialogue
- Significantly better than statistical approach
Challenges of Deep Learning

• Data hungry, it needs large amount data in training
• Learning is computationally costly
• Not good at inference and decision making
• Cannot easily incorporate symbolic data into model
• Difficult to deal with the long tail phenomena
• Model is usually a black box and is difficult to interpret
Research at Huawei Noah’s Ark Lab

• Huawei is the most cited Chinese company in AI during 2012-2016, - Nikkei News

• Representative work
  – Matching: Arc Two, two dimensional matching model, Hu et al., 2014
  – Dialogue: Neural Responding Machine, generation-based dialogue model, Shang et al., 2015
  – Image retrieval: Multimodal CNN, Ma et al., 2015
  – Dialogue and summarization: CopyNet, sequence-to-sequence learning with copying mechanism, Gu et al., 2016
  – Translation: coverage in sequence to sequence learning, Tu et al., 2016
Outline

• Deep Learning for Natural Language Processing
• *Neural Symbolic Processing*
• Related Work
• Our Work
• Summary
Question Answering from Knowledge Base

• Structured prediction (semantic parsing)
• Semantic parsing: mapping language expressions into internal representations
• Challenges
  • Knowledge is not categorical
  • Language is polysemous and synonymous
• Neural symbolic processing is necessary
Intelligent Question Answering System

Learning Phase

Analysis

Language Processing Unit

Short-term Memory

Consolidation

Long-term Memory

Central Executive Unit
Intelligent Question Answering System

Use Phase

- Language Processing Unit
- Short-term Memory
- Long-term Memory

Central Executive Unit
Knowledge Is Not Categorical
- Example: Bachelor

- Bachelor: unmarried adult male
- How to judge the following?
  - Unmarried father of child
  - Man having fake marriage
  - 17 year old high school student
  - 17 year old playboy
  - Homosexual lovers
  - Arabic man with two wives to meet another fiancee
  - Bishop
- From Terry Winograd

- Arthur has been living happily with Alice for the last five years. They have a two-year-old daughter and have never officially married.
- Bruce was going to be drafted, so he arranged with his friend Barbara to have a justice of the peace marry them so he would be exempt. They have never lived together. He dates a number of women, and plans to have the marriage annulled as soon as he finds someone he wants to marry.
- Charlie is 17 years old. He lives at home with his parents and is in high school.
- David is 17 years old. He left home at 13, started a small business, and is now a successful young entrepreneur leading a playboy's lifestyle in his penthouse apartment.
- Eli and Edgar are homosexual lovers who have been living together for many years.
- Faisal is allowed by the law of his native Abu Dhabi to have three wives. He currently has two and is interested in meeting another potential fiancee.
- Father Gregory is the bishop of the Catholic cathedral at Groton upon Thames.
Language Is Polysemous (Ambiguity) - Example: *Climb*

- **Climb**: motion from lower level to higher level, along a path, by laborious manipulation of limbs
- **Features**: [ascend] [clamber]
- **Climb** is polysemous category consisting of several senses
- The senses are related through meaning chain A-B-C-D
- From Charles Fillmore

- The boy climbed the tree.
- The locomotive climbed the mountainside.
- The plane climbed to 30,000 feet.
- * Smoke climbed from a chimney.
- * An elevator climbed from one floor to another.
- The temperature climbed into the 90s.
- Prices are climbing day by day.
- The boy climbed down the tree and over the wall.
- We climbed along the cliff edge.
- * The locomotive climbed over the mountain.
- He climbed out of a sleeping-bag.
Language Is Synonymous (Variability)

Example: Distance between Sun and Earth

- "how far" earth sun
- "how far" sun
- "how far" sun earth
- average distance earth sun
- average distance from earth to sun
- average distance from the earth to the sun
- distance between earth & sun
- distance between earth and sun
- distance between earth and the sun
- distance from earth to the sun
- distance from sun to earth
- distance from sun to the earth
- distance from the earth to the sun
- distance from the sun to earth
- distance from the sun to the earth
- distance of earth from sun
- distance between earth sun
- how far away is the sun from earth
- how far away is the sun from the earth
- how far earth from sun
- how far earth is from the sun
- how far from earth is the sun
- how far from earth to sun
- how far from the earth to the sun
- distance between sun and earth
Combination of Neural Processing and Symbolic Processing

Symbolic Representation

- Easy to Interpret
- Easy to Manipulate

Neural Representation

- Robust to Ambiguity & Variability
- Robust to Noise

Neural Symbolic Processing
Neural Symbolic Processing for Question Answering

Language Processing Model

- Encoder
- Decoder

Short-Term Memory

- Q
- A

Long-Term Memory

- Neu.
- Sym.

Knowledge in symbolic representation & neural representation

Central Executive Unit
Outline

• Deep Learning for Natural Language Processing
• Neural Symbolic Processing
• *Related Work*
• Our Work
• Summary
Semantic Parsing

- **Executor**: execute command based on logic form and context
- **Grammar**: set of rules for creating derivations based on input and context
- **Model**: model for ranking derivations with parameters
- **Parser**: find most likely derivation under learned model
- **Learner**: learn parameters of model \( \theta \) from data \( \{(x_i, c_i, y_i)\}_{i=1}^V \)

Q: What is the largest prime less than 10?  
A: 7

Liang 2016
Memory Networks

- Long term memory + inference
- Model is learned
- Can answer factoid questions

Example
- John is in the playground.
- Bob is in the office.
- John picked up the football.
- Bob went to the kitchen.
- Q: where is the football?
- A: playground

Weston et al. 2014
Neural Symbolic Machines

- Sequence to sequence model maps utterances to programs
- LISP interpreter performs program execution
- Policy-gradient to optimize reward of structured prediction

Liang et al. 2016
Outline

• Deep Learning for Natural Language Processing
• Neural Symbolic Processing
• Related Work
• Our Work
• Summary
Q: How tall is Yao Ming?
A: He is 2.29m tall and is visible from space.
(Yao Ming, height, 2.29m)

Q: Which country was Beethoven from?
A: He was born in what is now Germany.
(Ludwig van Beethoven, place of birth, Germany)

Q: How tall is Liu Xiang?
A: He is 1.89m tall
GenQA

Language Processing Module
Encoder creates question representation, decoder generates answer

Decoder

Short-Term Memory
Matches and retrieves most relevant answer representation

Q'  A'

Long-Term Memory
Triples in symbolic representations (indexed) & neural representations

End-to-End Training

Triples

Index
Decoding in GenQA

• Generating response using attention-based encoder-decoder
• At each position, deciding whether to generate a word or to insert a word from the retrieved result

He is 2.29m

<eos> He is 2.29m

encoding decoding

How tall is Yao Ming?

tall 2.29m

generate/insert

< yao ming, height, 2.29m>

retrieved result

is

attentive context

internal representation
Q: How many people participated in the games in Beijing?
A: 4,200

Q: When was the latest Olympic games held?
A: 2012

Q: Which city hosted the Olympic games before the games in Beijing?
A: Athens
Question Answering System

Language Processing Module

Encoder creates question representation, decoder simply returns answer

Short-Term Memory

Matches question representation to table representations to find answer

Long-Term Memory

Features and values are in symbolic representations and neural representations

Q' is matched to the question representation, and A' is the answer returned by the decoder.
Q: How long are the games for which the area of the host city is the largest?

<table>
<thead>
<tr>
<th>year</th>
<th>city</th>
<th>area</th>
<th>duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Sydney</td>
<td>4775</td>
<td>16</td>
</tr>
<tr>
<td>2004</td>
<td>Athens</td>
<td>1131</td>
<td>16</td>
</tr>
<tr>
<td>2008</td>
<td>Beijing</td>
<td>6490</td>
<td>16</td>
</tr>
<tr>
<td>2012</td>
<td>London</td>
<td>607</td>
<td>16</td>
</tr>
</tbody>
</table>

Sequence of operators

- $q$
- $h_1$
- $h_2$
- $h_3$
- $\text{argmax area}$
- $\text{select duration}$
- $\text{EOE}$

Recurrent neural network
Q: How long are the games for which the area of the host city is the largest?

Sequence of executors:
- Executor: Deep neural network
- Executor: Deep neural network
- Executor: Deep neural network

Answer:
- Year: 2008
  - City: Beijing
  - Area: 6490
  - Duration: 16
• Five executors, except last one, each one has reader, annotator, memory
• Reader fetches important representation for each row
• Annotator encodes result representation for each row
Coupling Symbolic Enquirer & Neural Enquirer

Symbolic Enquirer

$q \xrightarrow{argmax \ area} h_1 \xrightarrow{select \ duration} h_2 \xrightarrow{EOE} \text{Intermediate Result}

$h_1$ is generated by the symbolic enquirer.

$h_2$ is generated by the neural enquirer.

$h_3$ is generated by the neural enquirer.

$h_3$ is the final result.

Neural Enquirer

$f \xrightarrow{Annotation} h_1 \xrightarrow{Annotation} h_2 \xrightarrow{Annotation} h_3

$h_1$, $h_2$, and $h_3$ are created by the neural network.

EOE stands for End Of Execution.
Symbolic, Neural, and Coupled Enquirers

- Symbolic enquirer: model is not differentiable, policy-gradient (reinforcement learning)
- Neural enquirer: model is differentiable, gradient based learning
- Coupled: first train neural enquirer and then use the result to train symbolic enquirer
- Coupled: take advantage of two models

<table>
<thead>
<tr>
<th></th>
<th>Symbolic</th>
<th>Neural</th>
<th>Coupled</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Efficiency</strong></td>
<td>Low</td>
<td>Fair</td>
<td>Fair</td>
</tr>
<tr>
<td><strong>Execution Efficiency</strong></td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Model Interpretability</strong></td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Execution Accuracy</strong></td>
<td>Fair</td>
<td>Fair</td>
<td>High</td>
</tr>
</tbody>
</table>
Outline

• Deep Learning for Natural Language Processing
• Neural Symbolic Processing
• Related Work
• Our Work
• Summary
Summary

• Five main problems in NLP: classification, matching, translation, structured prediction, sequential decision process
• Deep Learning for NLP is making significant progress, particularly in the first four problems
• Advantages: end-to-end training, representation learning, learning of complex patterns
• Neural Symbolic Processing is necessary and important for NLP, particularly question answering
• Our proposals: GenQA, Neural Enquirer, Coupled Enquirers
References


Thank You!