Natural Language Dialogue
- Future Way of Accessing Information

Hang Li
Noah’s Ark Lab
Huawei Technologies
Scientific development depends in part on a process of non-incremental or revolutionary change.

Thomas Kuhn
Development of Information Retrieval

Library Search

Web Search

Natural Language Dialogue?

1970

1990

2010

2015
Outline

• Information Access through Natural Language Dialogue
• Data-Driven Approach
• Single Turn Dialogue
• Sentence Representation Learning
• Our Approach Using Deep Learning
• Summary
Natural Language Dialogue

Information Input:
• visual 83%
• auditory 11%
• others 6%

Information Output:
• speech 70-80%

Natural language dialogue: most natural way for human machine communication
Search: Restricted Single Turn Dialogue

- Query: keywords
- Document: bag of words
- Query document matching
- Ranking of documents
- Evaluation: relevance
- Learning to rank and learning to match
Question Answering: *Simplified* Single Turn Dialogue

- Question: question representation
- Answer: answer representation
- Question answer matching
- Evaluation: accuracy
- Learning to match
Task Completion through Natural Language Dialogue

• Multi-turn dialogue
• Goal: task completion, mostly information access
• Evaluation: completion rate / cost
• Including traditional search and question answering as special cases
Example One: Hotel Booking on Smartphone

P: How may I help you?
U: I'd like to book a hotel room for tomorrow.
P: For how many people?
U: Just me. What is the total cost?
P: That would be $120 per night.
U: No problem. Book the room for one night, please.
Example Two: Auto Call Center

- **U**: hello
- **H**: hello, how can I help you?
- **U**: can you tell me how to find ABC software?
- **H**: please go to this URL to download
- **U**: how to activate the software?
- **H**: please see this document
Example Three: Trouble Shooting at Telecommunication Network

R: What is your problem?

E: Dropped call in area..

R: My diagnosis is as follows

Symptom: ..... 

Likely cause: .... 

Solution: ....
Related Work

• Academic research
  – Rule-based approach, e.g., Eliza System, Weizenbaum 1964
  – Reinforcement learning based approach, e.g., Singh et al 1999
  – Machine translation based approach, e.g., Ritter et al 2011

• Commercial products
  – Apple Siri
  – Google Now
  – MS Cortana
Outline

• Information Access through Natural Language Dialogue
• *Data-Driven Approach*
• Single Turn Dialogue
• Sentence Representation Learning
• Our Approach Using Deep Learning
• Summary
Data-Driven Approach to Dialogue

Will you attend CCIR 2015?

Yes, I will

Learning how to converse mainly from data
Data-Driven Approach

• Key Points
  – Mainly learning from data
  – Avoid difficult language understanding problem as much as possible
  – Simple fundamental mechanism

• Issues to Investigate
  – Single-turn dialogue
  – Multi-turn dialogue
  – Knowledge utilization and reasoning (lightweight)
  – Dialogue management (lightweight)
As Scientific Research Problem

• Not as engineering hacks
• Scientific methodology
  – Mathematical modeling
  – Positivism
  – Reductionism
Outline

• Information Access through Natural Language Dialogue
• Data-Driven Approach
• Single Turn Dialogue
• Sentence Representation Learning
• Our Approach Using Deep Learning
• Summary
To divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution

- Ren´e Descartes
Solving The Problem of Single Turn Dialogue Can Partially Solve the Problem

Breaking down multi-turn dialogues to single-turn dialogues
Retrieval-based Single Turn Dialogue
Generation-based Single Turn Dialogue
• Retrieval-based approach to single turn dialogue
• 5 million Chinese Weibo data, 1 million Japanese Twitter data
• Registration deadline: Oct 31, 2015
Outline

• Information Access through Natural Language Dialogue
• Data-Driven Approach
• Single Turn Dialogue
• Sentence Representation Learning
• Our Approach Using Deep Learning
Distributional Linguistics

• Distributional semantics
• *You shall know a word by the company it keeps* - John Firth

• Sylvan?
• Companying words: woods, forest, tree, living, woodland, beautiful, grove
Using high-dimensional real-valued vectors to represent the meaning of words
Using high-dimensional real-valued vectors to represent the meaning of sentences

New finding: This is possible
Recent Breakthrough in Distributional Linguistics

• From words to sentences
• Compositional
• Representing syntax, semantics, even pragmatics
How Is Learning of Sentence Meaning Possible?

• Deep neural networks (complicated non-linear models)
• Big Data
• Task-oriented
• Error-driven and gradient-based
Natural Language Tasks

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

• Generation: creating a string
  \[ \rightarrow S \]

• Matching: matching two strings
  \[ s, t \rightarrow \mathbb{R}^+ \]

• Translation: transforming one string to another
  \[ S \rightarrow t \]

• Structured prediction: mapping string to structure
  \[ S \rightarrow S' \]
Natural Language Applications Can Be Formalized as Tasks

- Classification
  - Sentiment analysis
- Generation
  - Language modeling
- Matching
  - Search
  - Question answering
- Translation
  - Machine translation
  - Natural language dialogue (single turn)
  - Text summarization
  - Paraphrasing
- Structured Prediction
  - Information Extraction
  - Parsing
Learning of Representations in Tasks

• Classification
  \[ S \rightarrow r \rightarrow c \]

• Generation
  \[ \rightarrow s(r) \]

• Matching
  \[ s, t \rightarrow r \rightarrow \mathbb{R}^+ \]

• Translation
  \[ S \rightarrow r \rightarrow t \]

• Structured Prediction
  \[ S \rightarrow s' + r \]
Deep Learning Tools for Learning Sentence Representations

- Neural Word Embedding
- Recurrent Neural Networks
- Recursive Neural Networks
- Convolutional Neural Networks
## Word Representation: Neural Word Embedding

$$\text{log} \frac{P(w,c)}{P(w)P(c)}$$

$$M = WC^T$$

### Matrix Factorization

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W</strong></td>
<td><strong>t_1</strong></td>
<td><strong>t_2</strong></td>
<td><strong>t_3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>W_1</strong></td>
<td>7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>W_2</strong></td>
<td></td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>W_3</strong></td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Word Embedding or word2vec

$$M = WC^T$$ is the matrix factorization method used in neural word embedding, where $W$ represents the word embedding matrix and $C$ represents the context embedding matrix.
Recurrent Neural Network (RNN) (Mikolov et al. 2010)

- Variable length
- Long dependency: long short term memory (LSTM)

\[ h_t = f(h_{t-1}, w_t) \]

the cat sat on the mat
Recursive Neural Network (RNN) (Socher et al. 2011)

- Based on parse tree
- Learning for parsing and representation
- Max margin parsing

The cat sat on the mat
Convolutional Neural Network (CNN) (Hu et al. 2014)

- Robust parsing
- Shared parameter
- Fixed length, zero padding

the cat sat on the mat

max pooling

convolution

the

the cat
sat on
the cat sat

the cat
sat on
the cat sat

sat on
the mat
on the mat

sat on
on the
the mat
on the mat

the
cat
sat
on
the
mat
Outline

• Information Access through Natural Language Dialogue
• Data-Driven Approach
• Single Turn Dialogue
• Sentence Representation Learning
• *Our Approach Using Deep Learning*
• Summary
Collaborators

Zhengdong Lu  
Baotian Hu  
Mingxuan Wang

Lifeng Shang  
Qingcai Chen  
Qun Liu
Our Approach to Single-Turn Dialogue Using Deep Learning

• Retrieval-based
  – Deep Match CNN (Hu et al., NIPS 2015)
  – Deep Match Tree (Wang et al., IJCAI 2015)

• Generation-based
  – Neural Responding Machine (Shang et al., ACL 2015)
Natural Language Dialogue System
- Retrieval based Approach

message

online

offline

retrieved messages and responses

matching

ranked responses

best response

index of messages and responses

matching models

ranking model

retrieval

matched responses

ranked responses
Deep Match CNN
- Architecture I

• First represent two sentences, and then match
Deep Match CNN - Architecture II

- Represent and match two sentences simultaneously
- Two dimensional convolution and pooling
Deep Match Tree

• Constructing deep neural network, with first layer corresponding mined patterns
Retrieval based Approach: 
Accuracy = 70%+

It is very hot in Shanghai today, just like Singapore.
It is unusually hot.

I want to go to Mountain Wudang, is there anybody going together with me?

Haha, I want to go with you, handsome boy.
Natural Language Dialogue System - Generation based Approach

- Encoding messages to intermediate representations
- Decoding intermediate representations to responses
- Recurrent Neural Network (RNN)
Neural Responding Machine

Global Encoder

Local Encoder

Context Generator

Attention Signal

Decoder

140 million parameters
Generation based Approach
Accuracy = 70%+

占中终于结束了。  Occupy Central is finally over.

下一个是陆家嘴吧？  Will Lujiazui (finance district in Shanghai) be the next?

我想买三星手机。  I want to buy a Samsung phone

还是支持一下国产的吧。  Why not buy our national brands?
Demo: Neural Responding Machine
Advantages and Disadvantages

• Advantages
  – Powerful modeling capability
  – No human intervention

• Disadvantages
  – Black box
Open Questions and Future Work

• How to incorporate knowledge
• How to combine with logic
• How to extend to multi-turn dialogue
Outline

- Information Access through Natural Language Dialogue
- Data-Driven Approach
- Single Turn Dialogue
- Sentence Representation Learning
- Our Approach Using Deep Learning
- Summary
Take-away Messages

• Natural language dialogue = new paradigm for information retrieval
• Data-driven approach is key
• Current focus is single turn dialogue
• NTCIR STC task: call for participations
• Sentence representation learning is possible
• Progress being made on both retrieval-based and generation-based single-turn dialogue with deep learning and big data
References


Thank you!