First Neu-IR Workshop
July 21, 2016
Pisa, Italy

Does IR Need Deep Learning?

Hang Li
Huawei Noah’s Ark Lab

IR and DL

- IR (Information Retrieval) = information access tasks
- DL (Deep Learning) = human brain inspired, statistical learning tools

Our argument:
DL should be key technology for IR, and should be particularly effective for some hard IR problems

Talk Outline

- *Anatomy of IR Problems*
- Strength and Weakness of DL
- Research at Noah's Ark Lab
- Human Information Retrieval
- Discussions
- Take-away Message

IR Problems

Non-interactive (one short)

Interactive

Search (document retrieval)

Question answering (from documents)

Question answering (from knowledge base)

Image and video retrieval

Focus of
IR Research So Far
(Easy Problems)

IR Problems

Non-interactive
(one short)

Interactive

Search
(document retrieval)

Question
answering
(from documents)

Question
answering
(from knowledge base)

Image and video
retrieval

Characteristics of IR Problems

	Key concept	Main characteristics
Search or document retrieval	Topic relevance	Coarse-grained information need
Question answering from documents	Answer correctness	Fine-grained information need
Question answering from knowledge base	Answer correctness	Structured and unstructured data
Image and video retrieval	Topic relevance	Multimedia
Interactive information retrieval	Task completion	Multi-round of above actions

Talk Outline

- Anatomy of IR Problems
- *Strength and Weakness of DL*
- Research at Noah's Ark Lab
- Human Information Retrieval
- Discussions
- Take-away Message

Strength and Weakness of Deep Learning

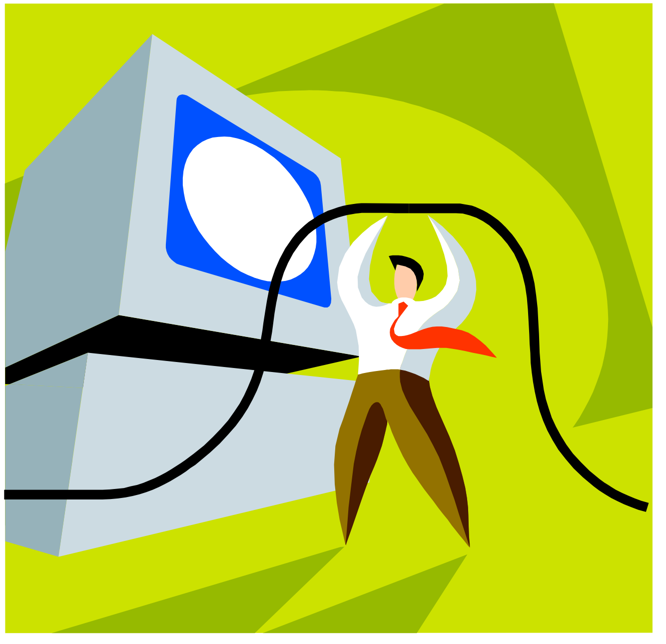
- Strength
  - Good at *pattern recognition* problems
  - Performance is high in many tasks
  - Little or no domain knowledge is needed in system construction such as feature engineering
  - Bar for entry is surprisingly low, with many tools available now
  - There are many powerful methods for supervised learning setting
- Weakness
  - Not good at *inference and knowledge* problems
  - Data-hungry and thus is not suitable when data size is small
  - Model is usually a black box and is difficult to understand
  - Still lack of theoretical foundation
  - Development of more unsupervised learning methods is needed

Two *Magic* DL Tools for IR

- Convolutional Neural Network (CNN)
- Sequence to Sequence Learning (S2SL)

- Example: image retrieval, accurate retrieval with CNN
- Example: single turn dialogue, accurate generation of response given message with S2SL

Demos

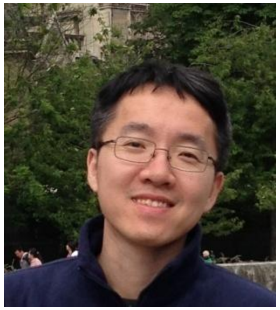


Talk Outline

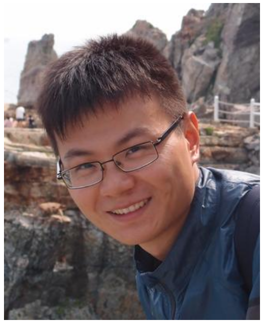
- Anatomy of IR Problems
- Strength and Weakness of DL
- *Research at Noah's Ark Lab*
- Human Information Retrieval
- Discussions
- Take-away Message

DL for NLP @Noah Lab

Zhengdong Lu



Xin Jiang



Lin Ma

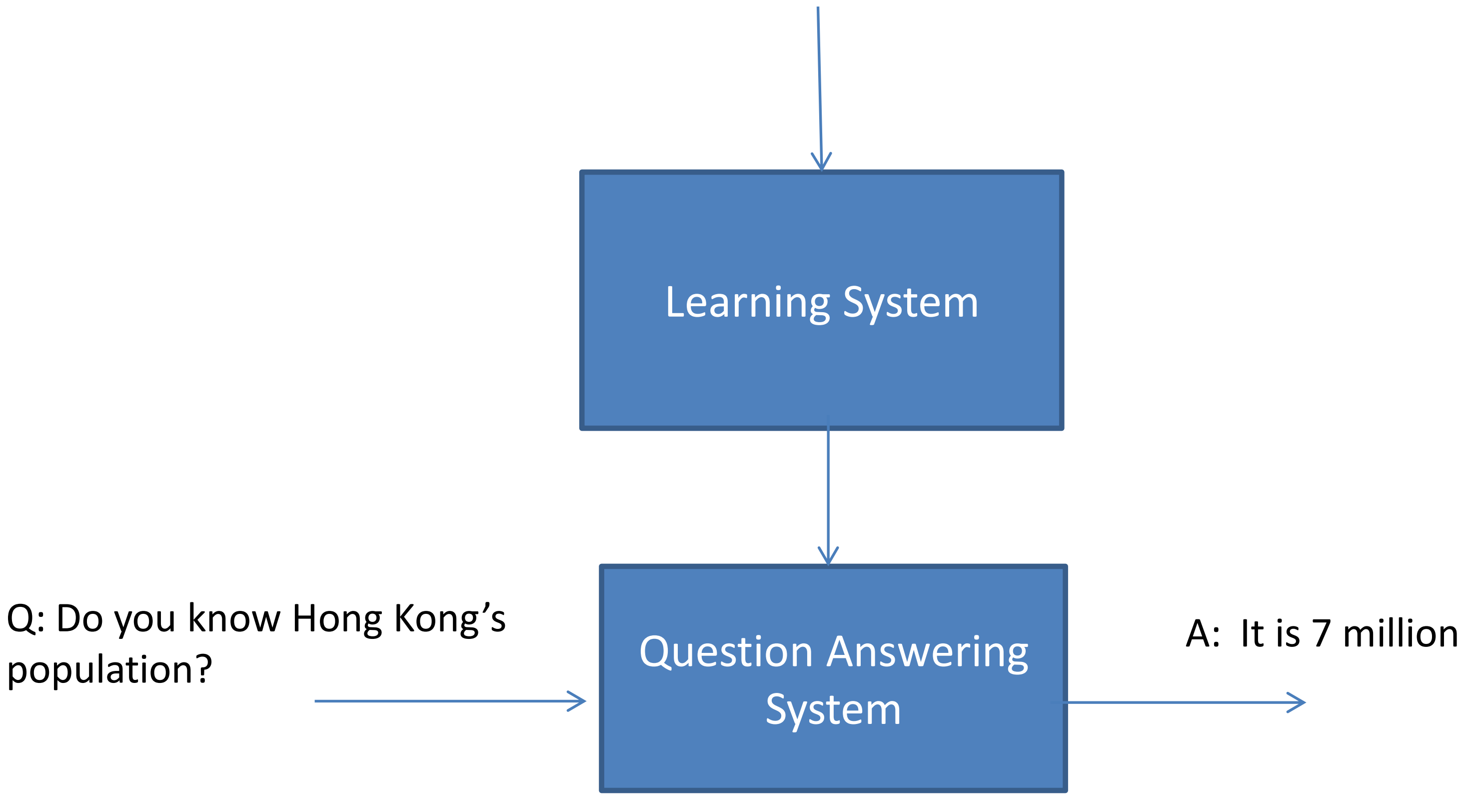
Lifeng Shang

Zhaopeng Tu

Generation-based Question Answering

Q: What is the population of Hong Kong?
A: It is 7.18 million as in 2013.

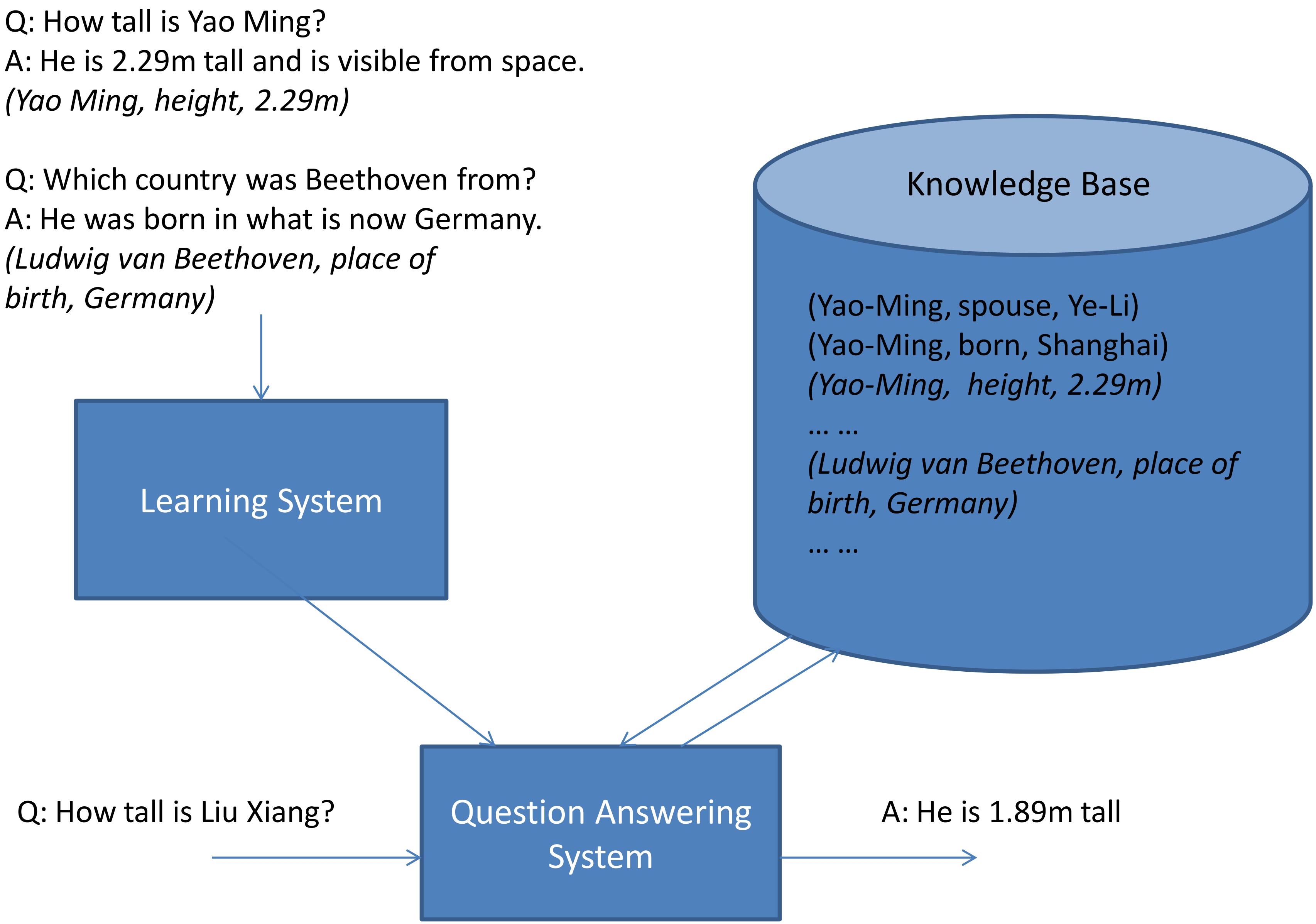
Q: How many people are there in Hong Kong?
A: There are about 7 million.



Our Work on Generation-based Question Answering

- Neural Responding Machine: generation-based single turn dialogue system using deep learning
- Model: sequence-to-sequence learning (encoder decoder framework)
- Encoding message into representation and decoding representation into response
- Experiment
  - Trained with 4.4 million Weibo data (Chinese)
  - 95% of responses are natural as sentences, 76% of responses are appropriate as replies
  - Demo
- Shang et al. 2015

Question Answering from Knowledge Base



Our Work on Question Answering from Knowledge Base

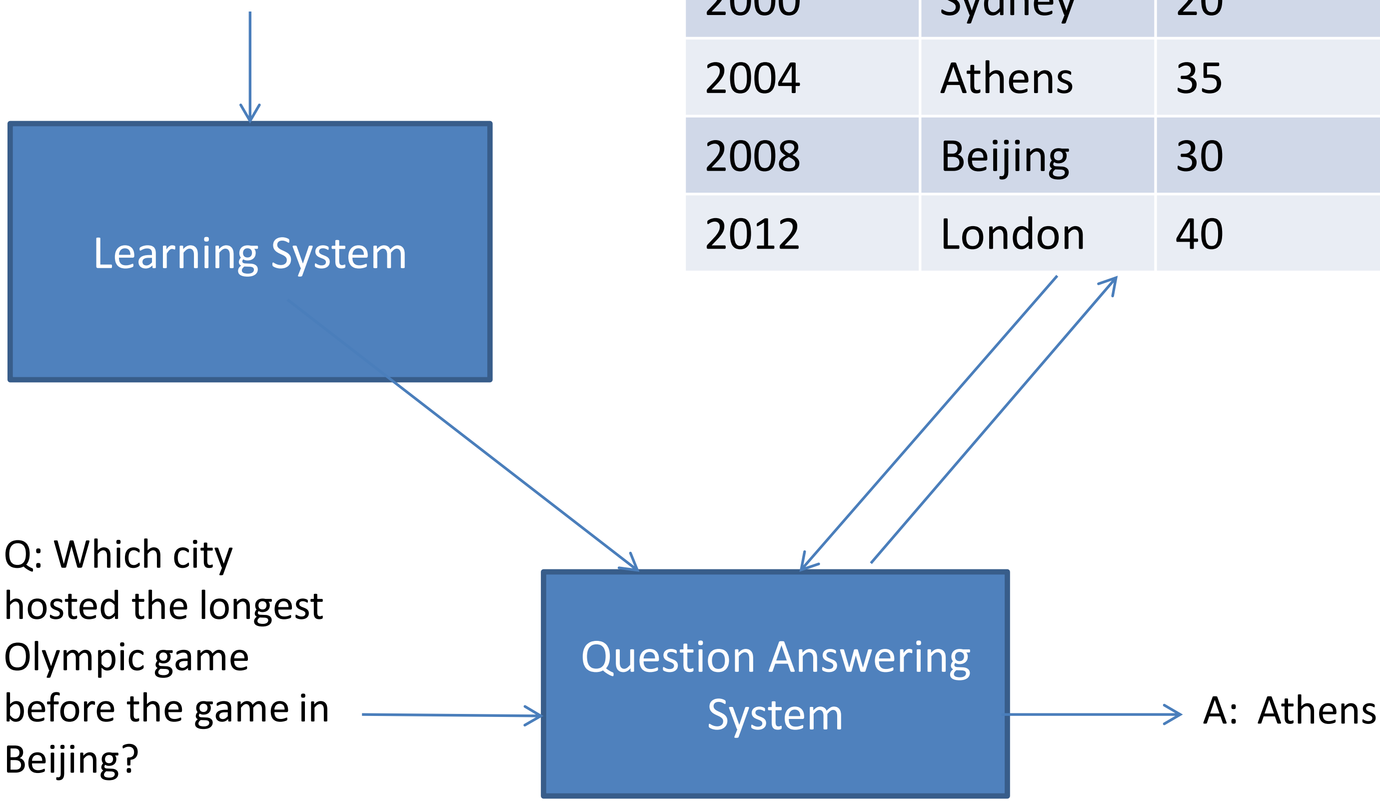
- Gen QA: encoder-decoder framework combined with knowledge base retrieval ability
- Encoding question to internal representation, retrieving triples with internal representation, generating answer using internal representation and retrieved triple
- Experiment
  - Trained with 720K question-answer pairs (Chinese) associated with 1.1M triples in knowledge-base
  - Accuracy = 52%
  - Data is still noisy
- Yin et al. 2016

Question Answering from Relational Database

Q: How many people participated in the game in Beijing?
A: 4,200
SQL: select #_participants, where city=beijing
Q: When was the latest game hosted?
A: 2012
SQL: argmax(city, year)

Relational Database

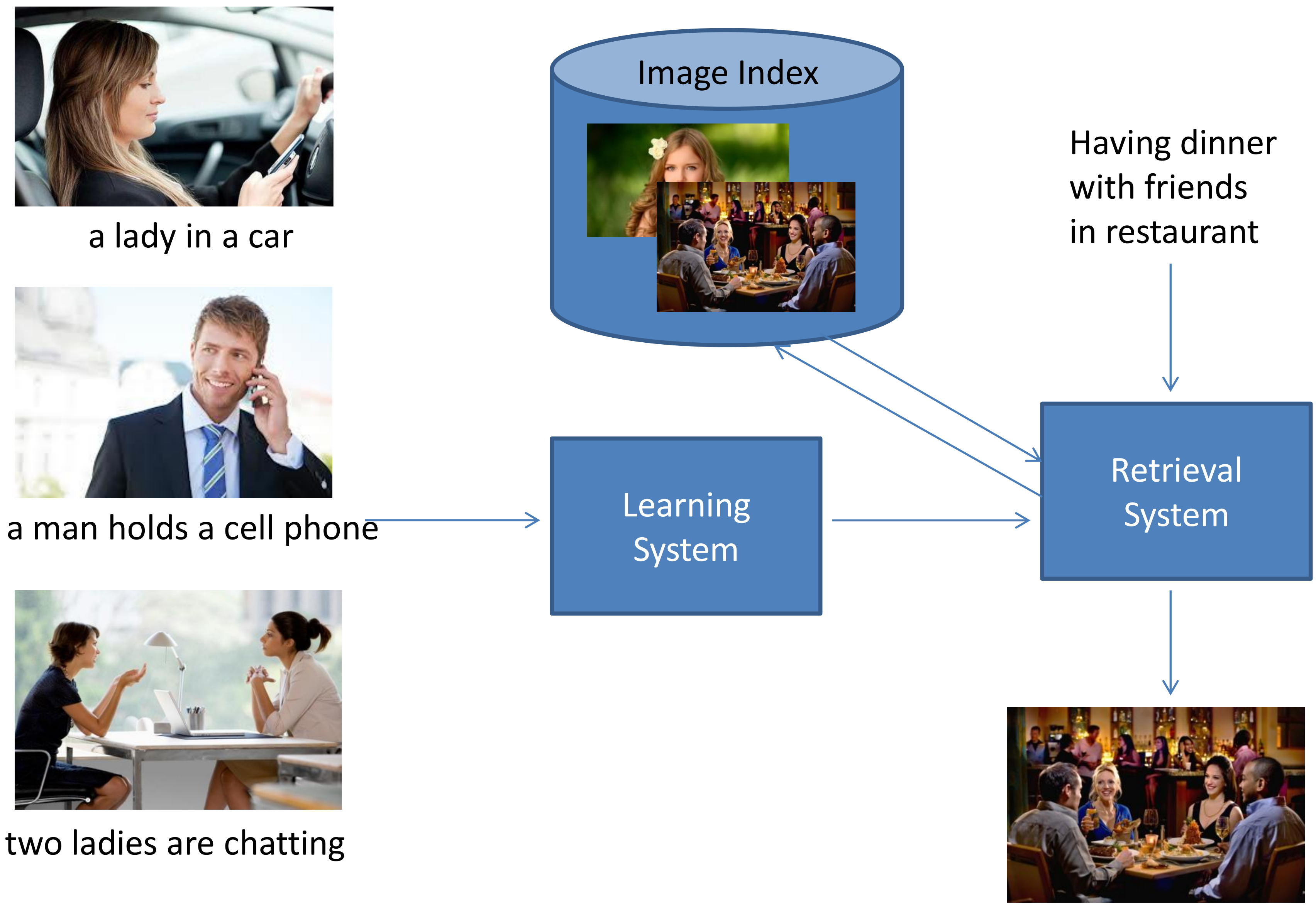
year	city	#_days	#_medals
2000	Sydney	20	2,000
2004	Athens	35	1,500
2008	Beijing	30	2,500
2012	London	40	2,300



Our Work on Question Answering from Relational Database

- Neural Inquirer: multi-step matching model
- Encoding query and table entries, conducting matching at several steps, saving intermediate results at external memories, training model in end-to-end fashion
- Experiment
  - Olympic database
  - Trained with 25K and 100K synthetic data
  - Accuracy: 84% on 25K data, 91% on 100K data
  - Significantly better than SemPre (semantic parser)
  - Criticism: data is synthetic
- Yin et al. 2016

Image Retrieval



Our Work on Image Retrieval

- Multimodal CNN: best performing image retrieval model
- Model: deep matching model using CNN
- Modeling image using CNN, modeling text using CNN
- Experiment
  - Trained with 30K Flickr data
  - R@1=26.2, R@5=56.3, R@10=69.6
  - Outperforming other state-of-the-art models
  - Demo
- Ma et al. 2015

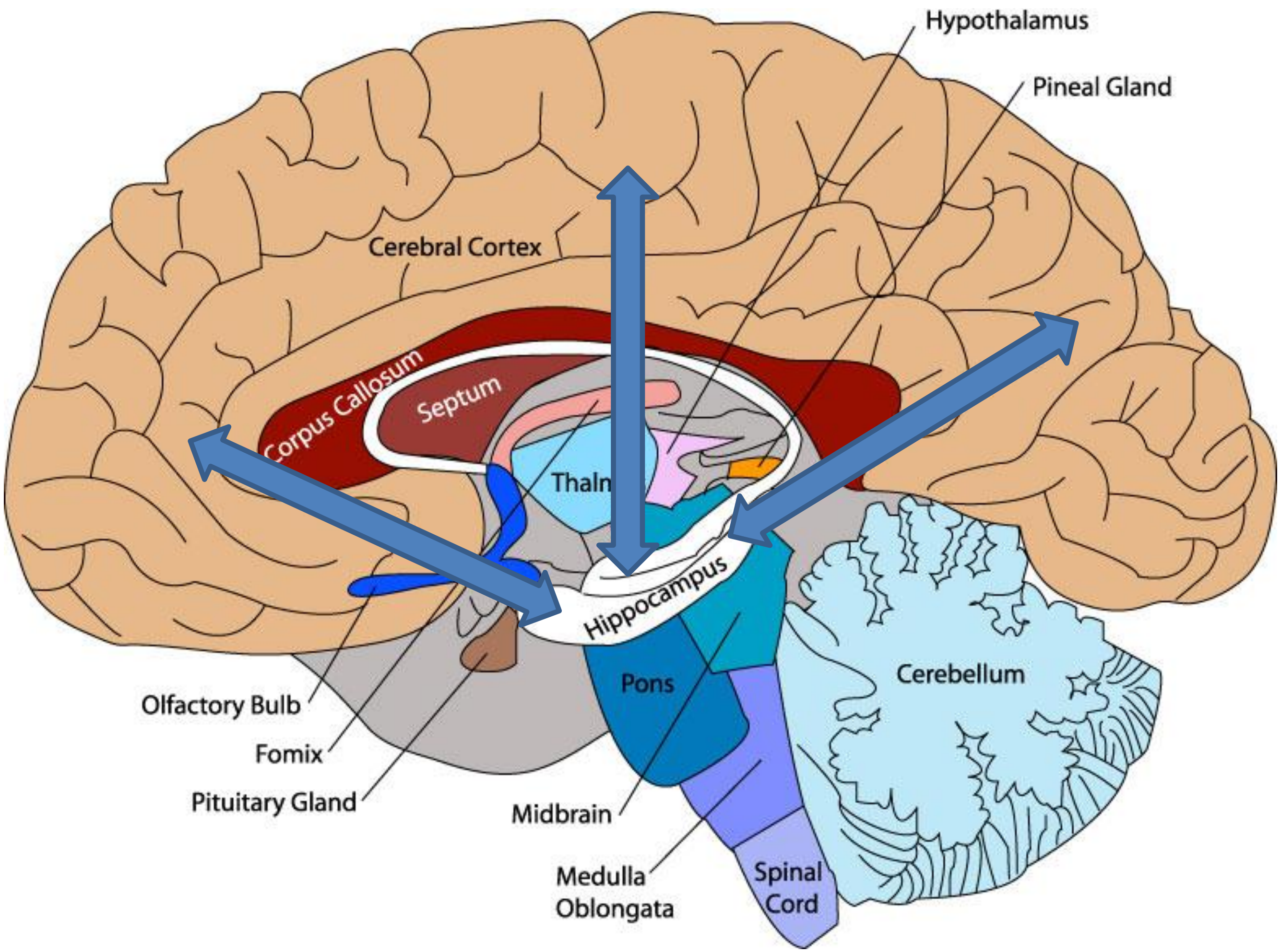
Talk Outline

- Anatomy of IR Problems
- Strength and Weakness of DL
- Research at Noah's Ark Lab
- *Human Information Retrieval*
- Discussions
- Take-away Message

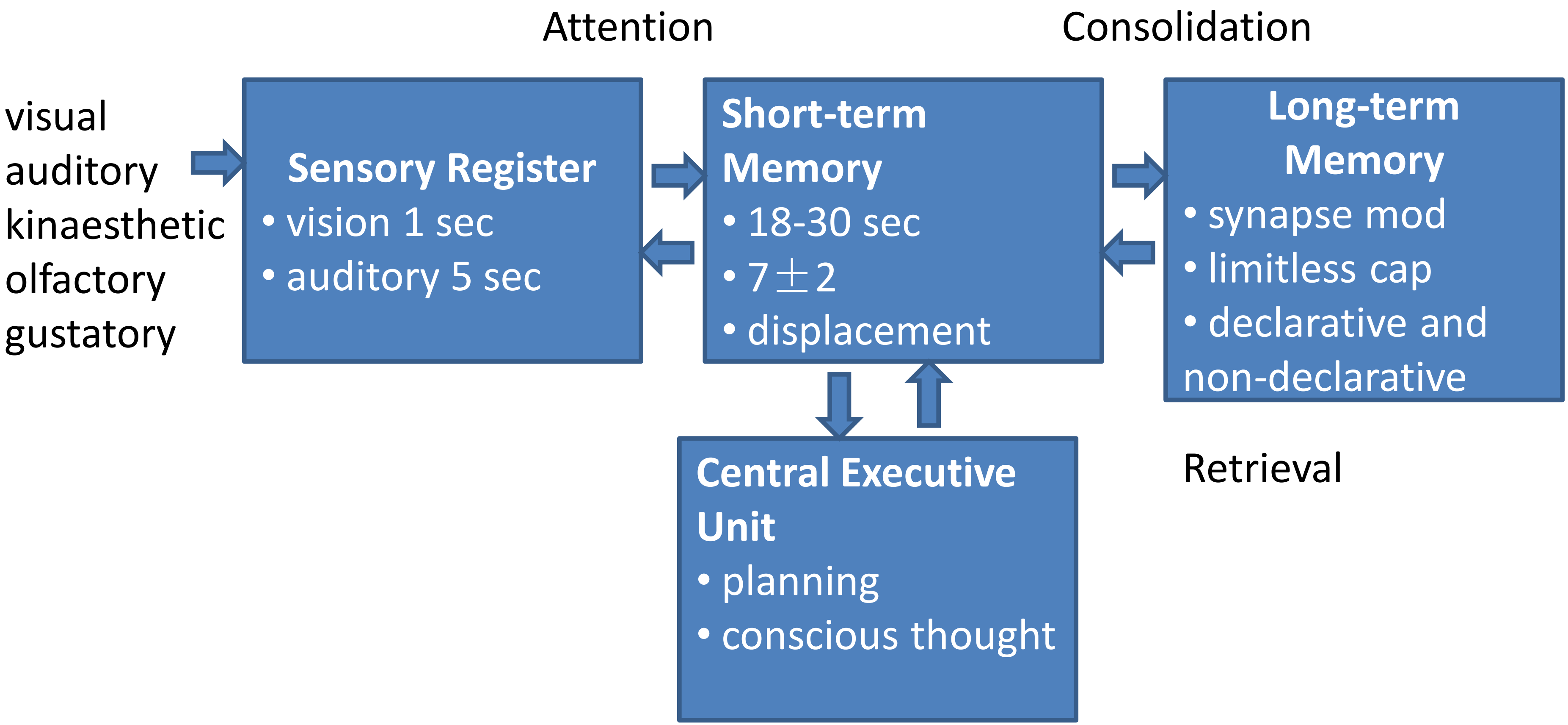
Human Information Retrieval

- Hippocampus (short term memory) ←→ Cerebral Cortex (long term memory)
- Information and knowledge is stored in long term memory
- Hebb’s hypothesis: **fire together wire together**
- Consolidation: create connections between neurons (patterns) in long term memory
- Retrieval: activate related neurons through connections in long term memory

Human Brain


Hypothalamus
Pineal Gland
Cerebral Cortex
Corpus Callosum
Septum
Thalm
Hippocampus
Olfactory Bulb
Fomix
Pituitary Gland
Midbrain
Pons
Medulla
Oblongata
Spinal
Cord
Cerebellum

Encoding, Storage, and Retrieval of Information in Human Brain



Modified from Frank Longo 2010

Information Retrieval in Human Brain v.s. Information Retrieval on Computer

	Brain	Computer
Computing paradigm	Parallel processing	Sequential processing
Capability	Mathematically Ill-posed problems	Mathematically well-formed problems
Representation of information	Represented in neurons and synapses	Represented by *digitized* symbols, numbers, data structures
Language to encode information	Mentalese (hypothetical language of thought, cf. Pinker)	Mainly in Natural language
Means of retrieval	Association of neurons	*IR models*

Talk Outline

- Anatomy of IR problems
- Strength and Weakness of DL
- Research at Noah's Ark Lab
- Human Information Retrieval
- *Discussions*
- Take-away Message

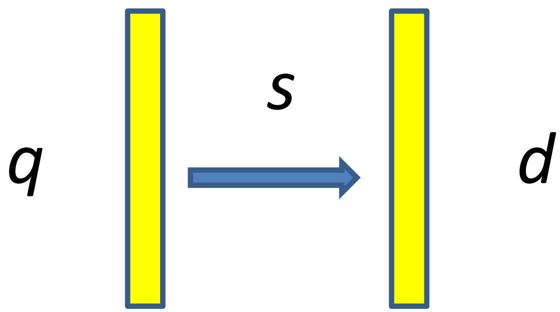
Breakdown of Argument

- DL should be key technology for IR
- DL can improve, but may not significantly enhance document retrieval
- DL should be powerful for hard IR problems (question answering, image retrieval, etc)
- DL might not be enough for interactive IR

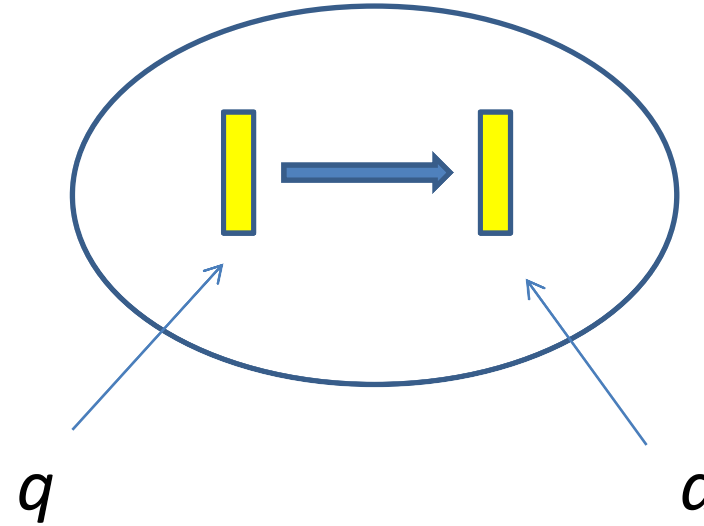
DL and IR

- DL is strongly connected with IR
- Many IR models can be viewed as approximate implementation of Hebb's rule
- Almost all IR models represent association (matching) from query to document or question to answer
- Past work shows
  - non-linear models work better than linear models
  - machine learning models work better than human defined models
- *DL models are nothing but extensions of IR models*

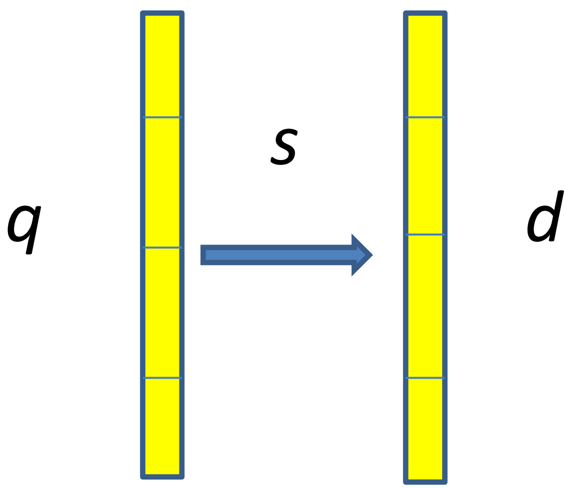
Symbolic Matching Models



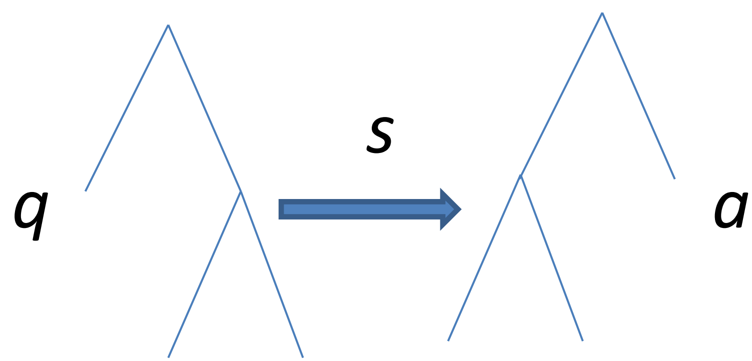
Vector Space Model,
BM25, Language Model for IR



Linear Projection into Latent Space
Wu et al. 2012

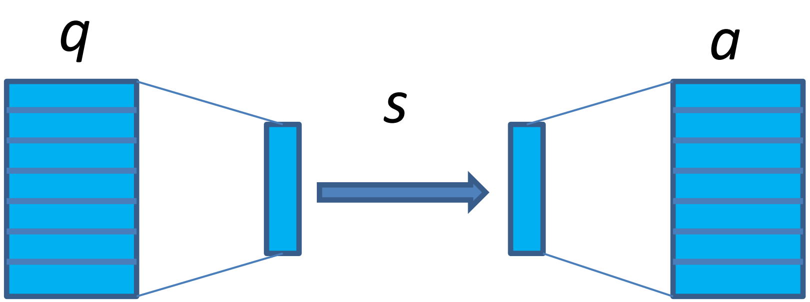


N-gram Matching Model,
Xu et al. 2010, Bu et al. 2012

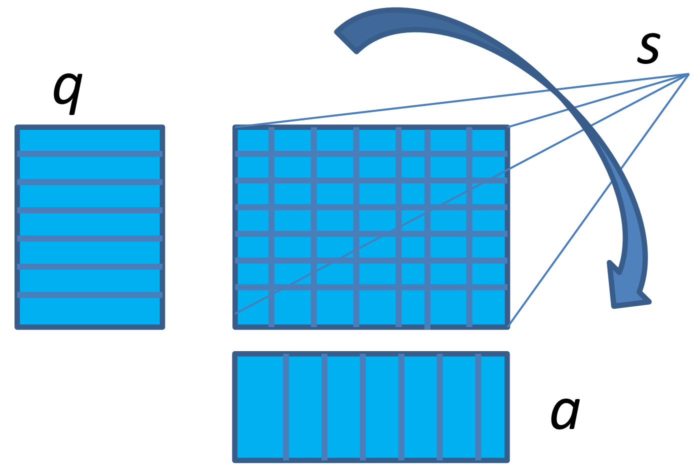


Tree Kernel,
Moschitti et al. 2007

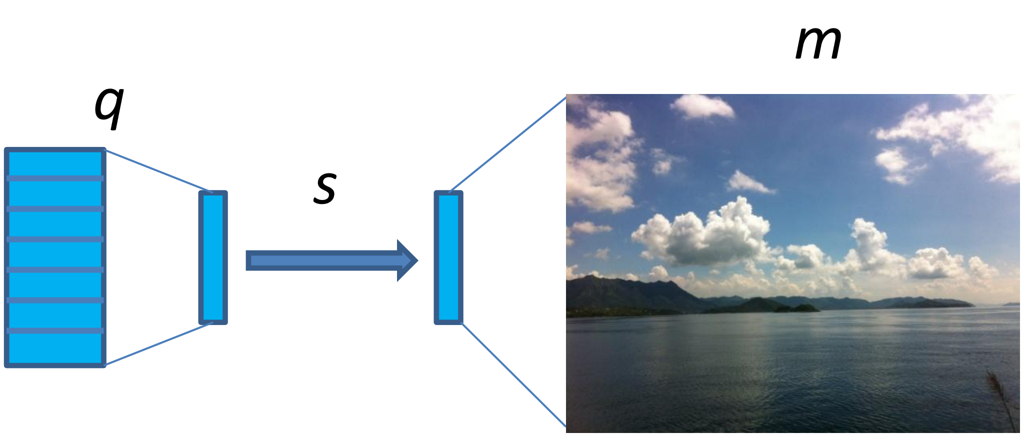
Neural Matching Models



Deep Match Models (DNN, CNN, RNN),
e.g., Shen et al. 2013



Two Dimensional Deep Match Model, Hu et al. 2014



Multimodal Match Model (CNN),
Ma et al. 2015

Neural matching models are natural extension of symbolic matching models

DL for Document Retrieval

- Symbolic processing is mostly enough, because
  - words are quite strong to indicate topics
  - key is topical relevance
  - term mismatch usually occurs at tail
- It appears that DL can improve precision, but improvement may not be so significant
- DL can improve recall, but it requires more training data

Mini Experiment: Five Randomly Shuffled Sentences

1. pulled out text phone to he Damian his
2. text unpleasant back ahead to brought him message Jule's him task of the
3. sat Darian down text a to shook head her and she sending
4. support will not careful text study this of view the
5. worsening mood text his up popped another

1. he pulled out his phone to text Damian.
2. Jule's text message brought him back to the unpleasant task ahead of him.
3. she shook her head and sat down, sending a text to Darian.
4. careful study of the text will not support this view.
5. another text popped up, worsening his mood.

Observations

- Bag of words can carry a large proportion of information
- Humans can get the topic of each sentence from bag of words
- Structures of sentences are important for understanding the exact meaning
- Symbolic representations are not enough, however, because language is polysemous and synonymous, e.g. “pulled out” vs “took out”

Challenge in Document Retrieval: Mismatch between Query and Document in the Tail

Query	Document	Term Match	Semantic Match
seattle best hotel	seattle best hotels	no	yes
pool schedule	swimmingpool schedule	no	yes
natural logarithm transformation	logarithm transformation	partial	yes
china kong	china hong kong	partial	no
why are windows so expensive	why are macs so expensive	partial	no

DL for Question Answering and Image Retrieval

- Key of QA is *exact* semantic matching
- Key of image retrieval is matching *across media*
- Both were difficult for previous approaches based on symbolic processing
- DL can make a big difference
- Requirement: large scale training data

Deep Learning and Interactive Information Retrieval

- DL may not be enough for interactive IR, a special case of natural language dialogue
- Key is dialogue management, including dialogue control and dialogue modeling
- Involvement of multiple “modules”, each having multiple “states”
- Recent work tries to use deep learning (e.g., Wen et al. 2016)
- There are many open questions

Mini Experiment: Comparison between Single-turn QA and Multi-turn QA by Humans

Single-turn QA

- **Q:** How tall is Yao Ming?
- **A:** He is 2.29m tall.

Multi-turn QA

- **Q:** How tall is Yao Ming?
- **A:** He is 2.29m tall.
- **Q:** Who is taller, Yao Ming or Liu Xiang?
- **A:** He is taller, and I think that Liu Xiang is only 1.89m tall.

- Single turn QA is only related to fact retrieval and answer generation.
- Multi-turn QA needs fact retrieval and answer generation, as well as other *mental processing*. More modules in human brain are involved.

Talk Outline

- Anatomy of IR Problems
- Strength and Weakness of DL
- Research at Noah's Ark Lab
- Human Information Retrieval
- Discussions
- *Take-away Message*

Take-away Message

- We should take a broad view on IR; QA, image retrieval, etc are all IR problems
- DL is key technology for IR
- IR models can be viewed implementation of Hebb's rule
- There is continuity from traditional IR models to DL models for IR
- Document retrieval: symbolic approach is almost sufficient, DL can help further improve, but the improvement may not be so significant
- DL is particularly effective for hard IR problems (complicated matching, multimodal matching)
- Interactive IR is important topic for future research, DL may only play a partial role

References

- Hang Li, Zhengdong Lu, Deep Learning for Information Retrieval. SIGIR 2016 Tutorial.
- Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. ACL-IJCNLP 2015.
- Lin Ma, Zhengodng Lu, Lifeng Shang, Hang Li . Multimodal Convolutional Neural Networks for Matching Image and Sentence, ICCV 2015.
- Pengcheng Yin, Zhengdong Lu, Hang Li, Ben Kao. Neural Enquirer: Learning to Query Tables. arXiv, 2015
- Jun Yin, Xin Jiang, Zhengdong Lu, Lifeng Shang, Hang Li, Xiaoming Li. Neural Generative Question Answering. arXiv, 2015.
- Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, Larry Heck. Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data. CIKM 2013.
- Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, Steve Young. A Network-based End-to-End Trainable Task-oriented Dialogue System. arXiv:1604.04562, 2016.

References

- Jun Xu, Hang Li, Chaoliang Zhong. Relevance Ranking Using Kernels. AIRS 2010.
- Hang Li, Jun Xu, Semantic Matching in Search. Foundation and Trends in Information Retrieval. 2014.
- Fan Bu, Hang Li, and Xiaoyan Zhu. String Re-writing Kernel. ACL 2012.
- Wei Wu, Zhengdong Lu, Hang Li. Learning Bilinear Model for Matching Queries and Documents. Journal of Machine Learning, 2013.
- Alessandro Moschitti and Fabio Massimo Zanzotto. Fast and Effective Kernels for Relational Learning from Texts. ICML 2007.
- Frank Longo, Learning and Memory: How It Works and When It Fails. Stanford Lecture. 2010.
- Steven Pinker. The Language Instinct: How the Mind Creates Language. 1994.

Thank you!

Hang Li
hangli.hl@huawei.com