Opportunities and Challenges in Deep Learning for Information Retrieval

Hang Li

Noah’s Ark Lab, Huawei Technologies

Talk Outline

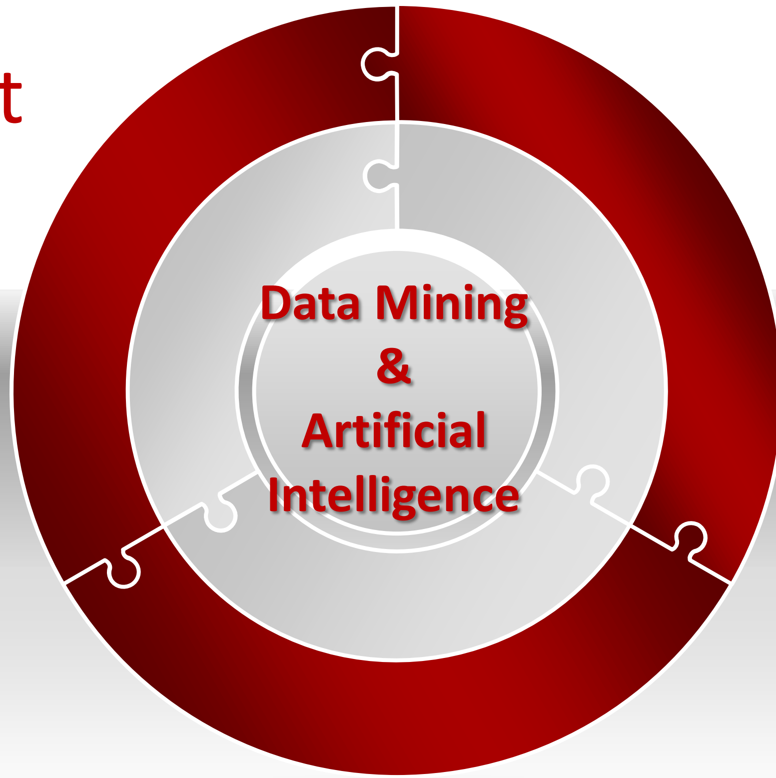
- Introduction to Huawei Noah's Ark Lab
- Deep Learning – New Opportunities for Information Retrieval
- Three Useful Deep Learning Tools
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Knowledge Base
  - Question Answering from Database
- Discussions and Concluding Remarks

Huawei's Vision:
Building A Better Connected World

Noah’s Ark Lab is Research Lab Working on

Intelligent Mobile Devices

Intelligent Telecommunication Networks



Intelligent Enterprise

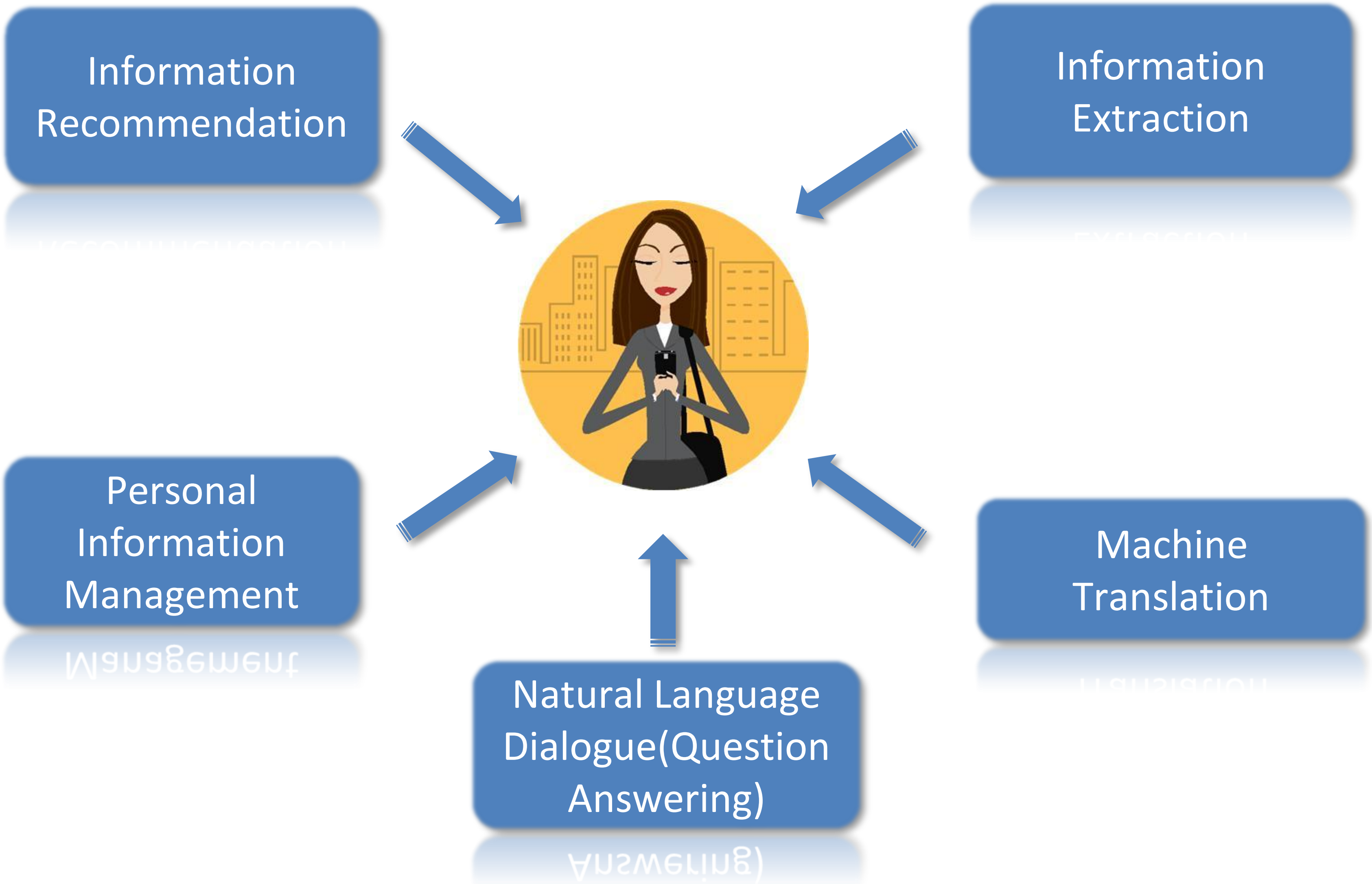
Intelligent Telecommunication Networks

Software-defined Network

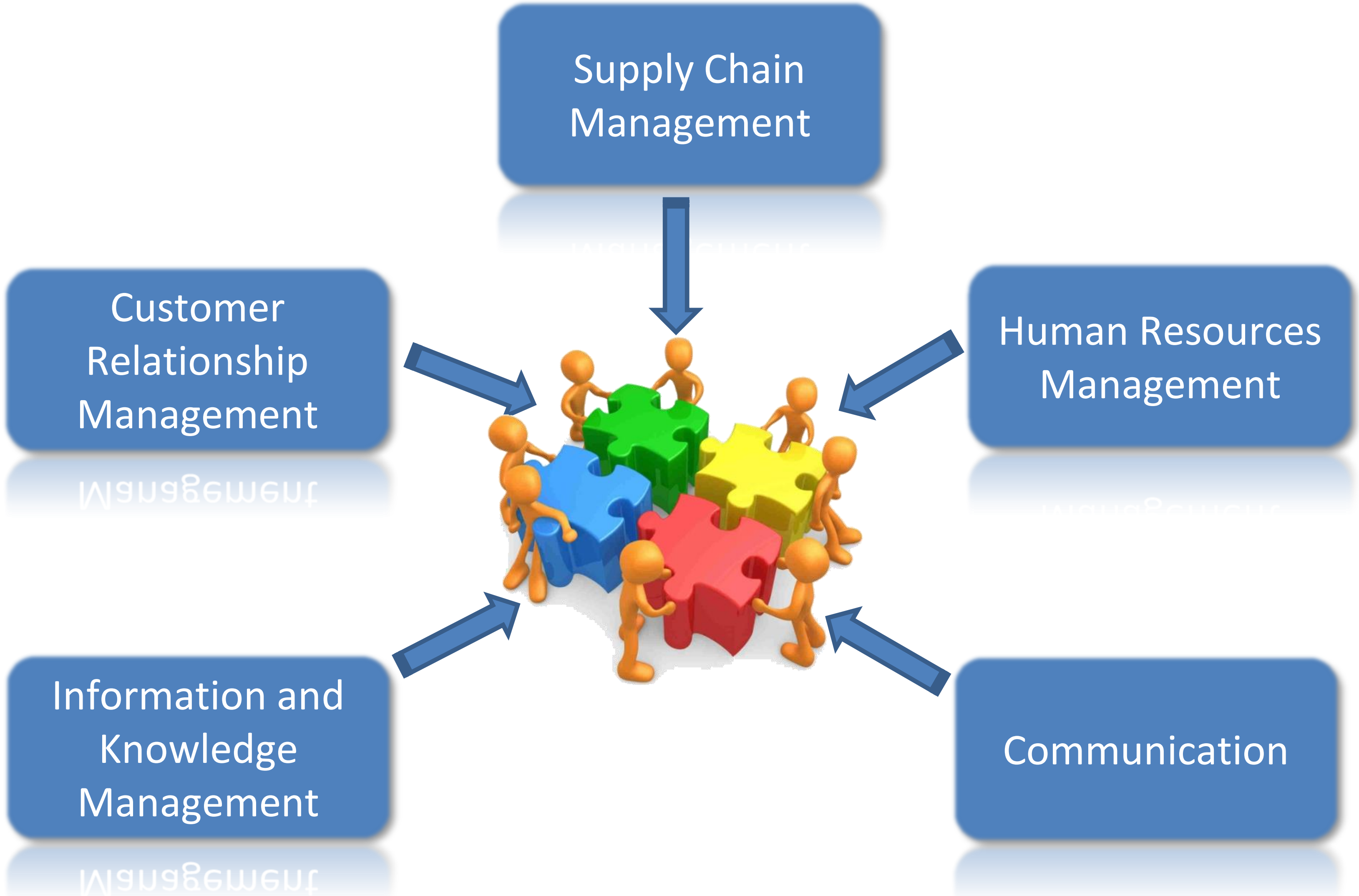
Network Maintenance

Network Planning and Optimization

Intelligent Mobile Devices



Intelligent Enterprise



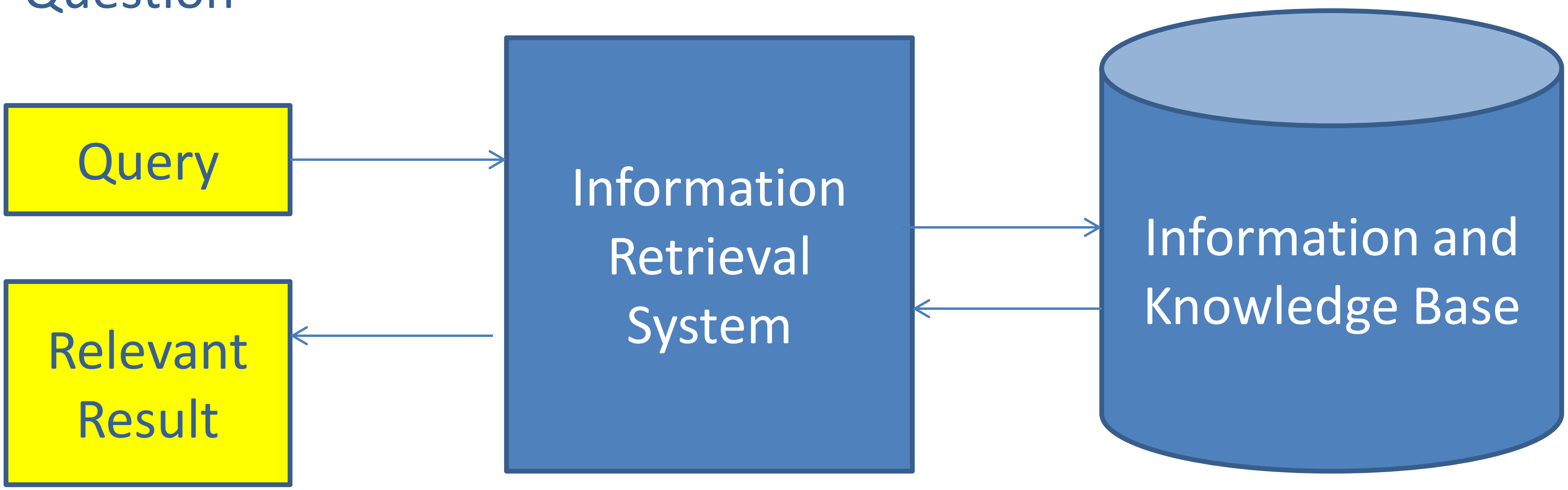
Talk Outline

- Introduction to Huawei Noah's Ark Lab
- Deep Learning – New Opportunities for Information Retrieval
- Three Useful Deep Learning Tools
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Database
  - Question Answering from Knowledge Base
- Discussions and Concluding Remarks

Overview of Information Retrieval

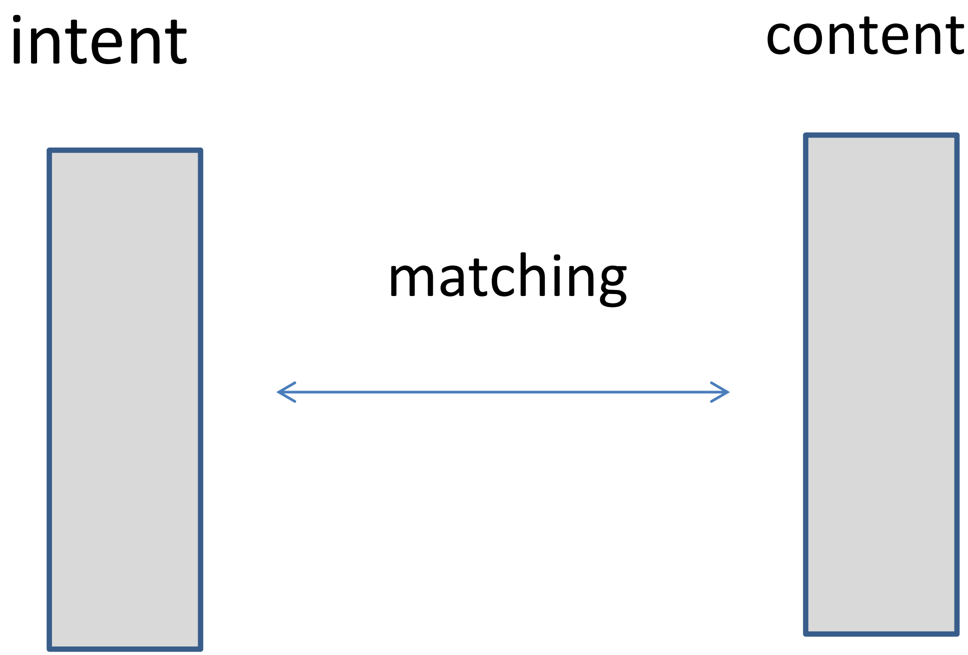
Intent:
Key words,
Question

Content:
Documents,
Images,
Knowledge



Key Questions: How to Represent Intent and Content, How to Match Intent and Content

Key is Matching



- Indexing: for efficient retrieval
- Ranking: when there are multiple result
- Generation: when only return single answer

Approach in Traditional IR

Document:

Star Wars: Episode VII
Three decades after the defeat of the Galactic Empire, a new threat arises.

Query:
star wars the force awakens reviews

$$q = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \xrightarrow{f(q,d)} d = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$f_{VSM}(q,d) = \frac{\langle q, d \rangle}{\| q \| \cdot \| d \|}$$

- Representing query and document as word vectors
- calculating cosine similarity between them

Approach in Modern IR

Query:
star wars the force awakens reviews

Document:

Star Wars: Episode VII
Three decades after the defeat of the Galactic Empire, a new threat arises.

$$
\begin{array}{llcl}
 & q & & d \\
\text{(star wars)} & v_{q1} & \xrightarrow{\vec{f}(q,d)} & v_{d1} \\
\text{(the force awakens)} & \vdots & & \vdots \\
\text{(reviews)} & v_{qm} & & v_{dn}
\end{array}
$$

- Conducting query and document understanding
- Representing query and document as multiple feature vectors
- Calculating multiple matching scores between query and document
- Training ranker with matching scores as features using learning to rank

Examples of Query Document Mismatch

Query	Document	Term Matching	Semantic Matching
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

Hard Problems in IR and NLP

How far is sun from earth? → Question Answering → The average **distance between the Sun and the Earth** is about 92,935,700 miles.

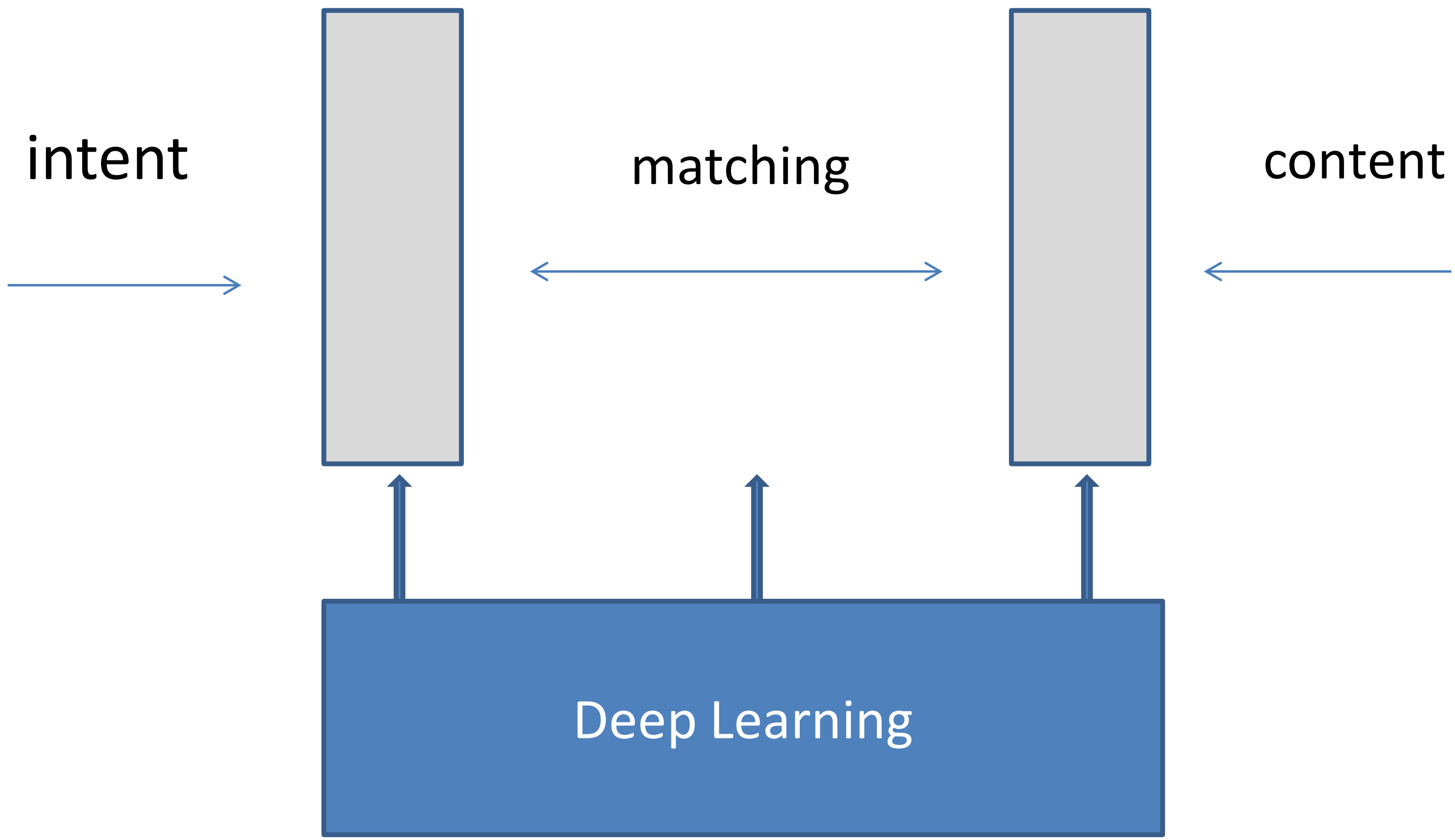
How tall is Yao Ming? → Question Answering from Relational Database →

Name	Height	Weight
Yao Ming	2.29m	134kg
Liu Xiang	1.89m	85kg

A dog catching a ball → Image Retrieval →

Key Questions: How to Represent Intent and Content, How to Match Intent and Content

Representation and Matching Are Key Problems in IR



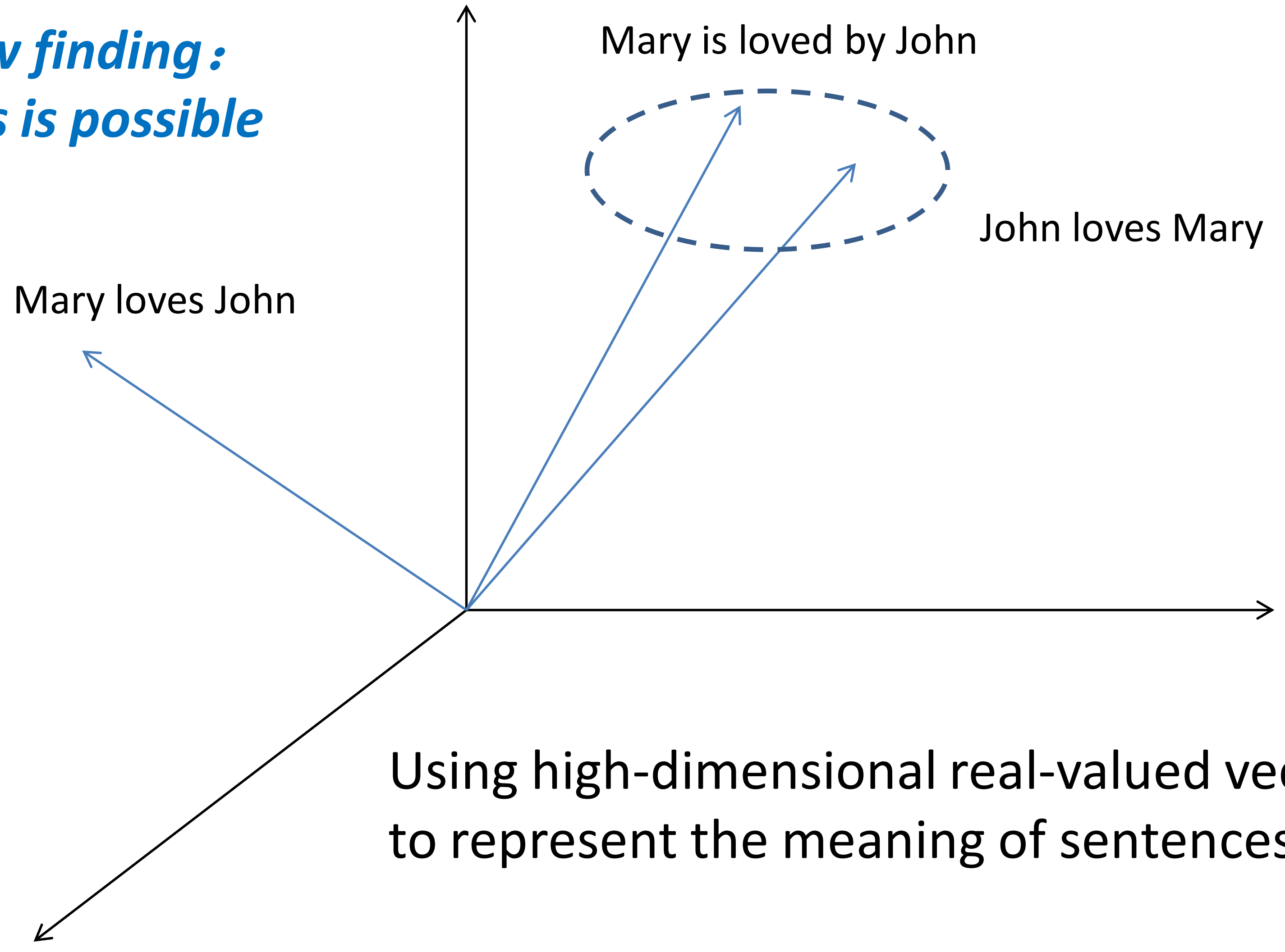
Recent Progress: Deep Learning Enables Representation Learning and Matching in IR

Talk Outline

- Introduction to Huawei Noah’s Ark Lab
- Deep Learning – New Opportunities for Information Retrieval
- Three Useful Deep Learning Tools
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Knowledge Base
  - Question Answering from Database
- Discussions and Concluding Remarks

Representation of Sentence Meaning

New finding：
This is possible



Using high-dimensional real-valued vectors
to represent the meaning of sentences

Word Representation: Neural Word Embedding (Mikolov et al., 2013)

M	c_1	c_2	c_3	c_4	c_5
w_1	**10**		**1**	**5**	
w_2		2			1
w_3	2.5			1	

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

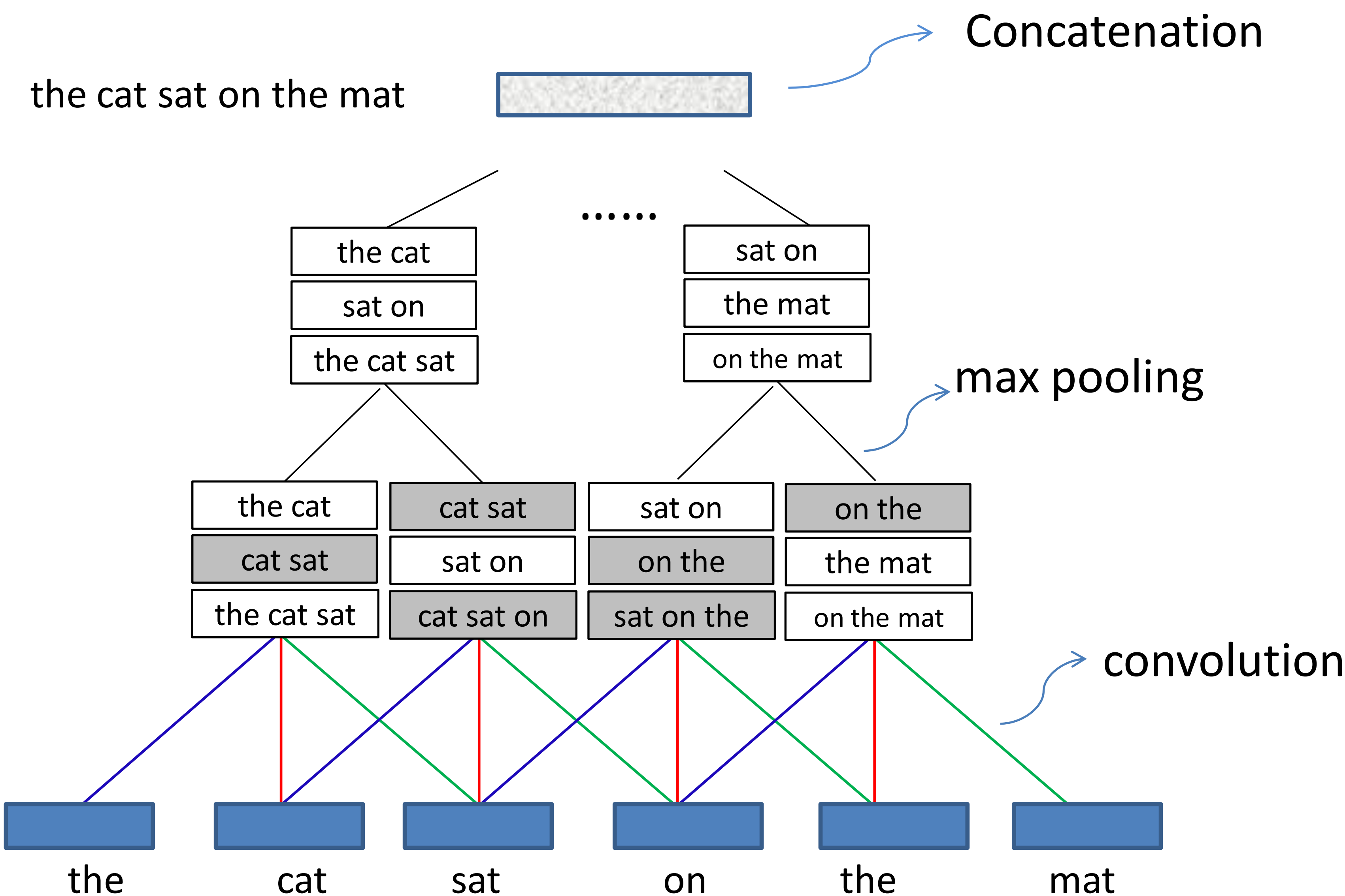
$$M = WC^T$$

matrix factorization

W	t_1	t_2	t_3
w_1	**7**	**1**	**1**
w_2		2	3
w_3	1	1.5	2

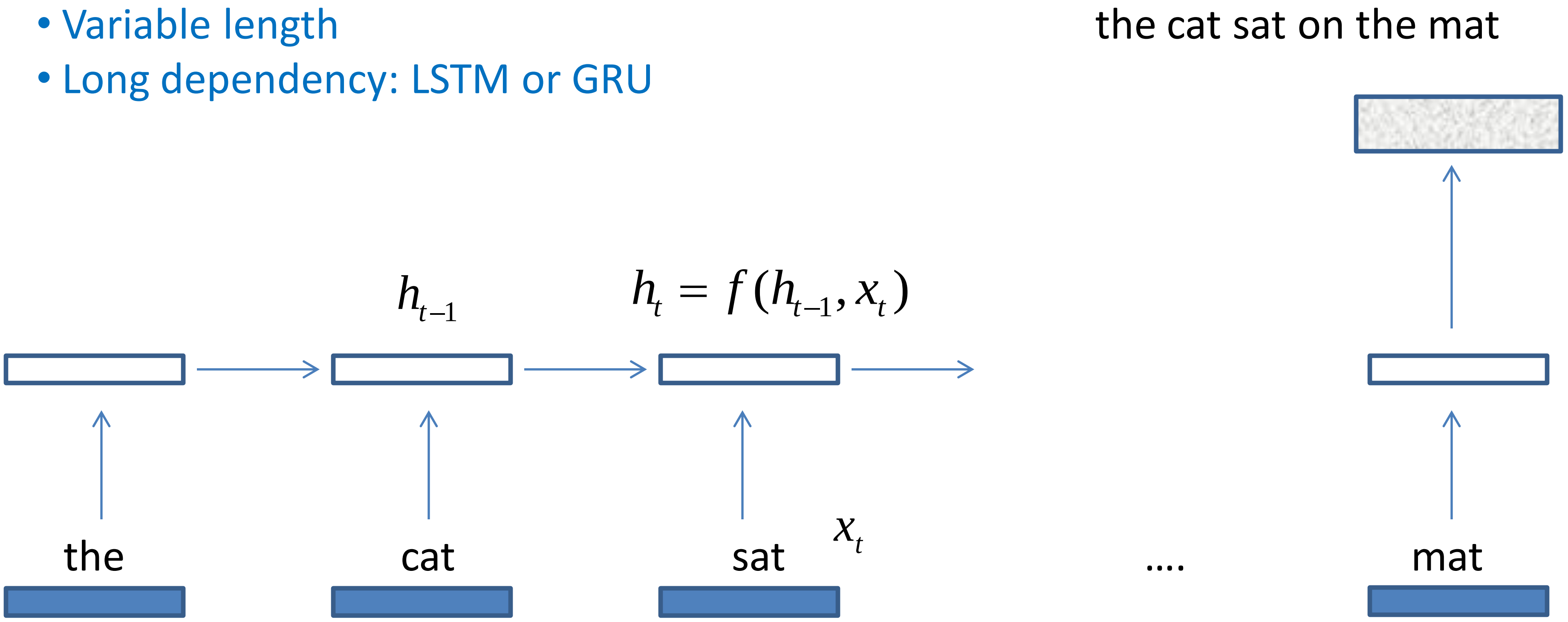
word embedding or word2vec

Convolutional Neural Network (CNN)
(Hu et al., 2014)



Recurrent Neural Network (RNN) (Mikolov et al. 2010)

- On sequence of words
- Variable length
- Long dependency: LSTM or GRU

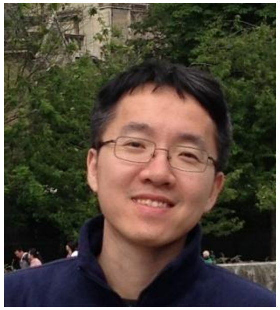


Talk Outline

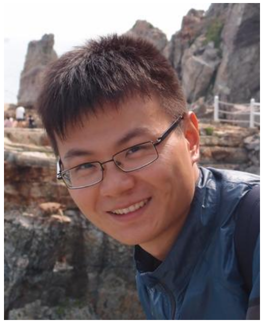
- Introduction to Huawei Noah's Ark Lab
- Deep Learning – New Opportunities for Information Retrieval
- Three Useful Deep Learning Tools
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Knowledge Base
  - Question Answering from Database
- Discussions and Concluding Remarks

DL for NLP @Noah Lab

Zhengdong Lu



Xin Jiang



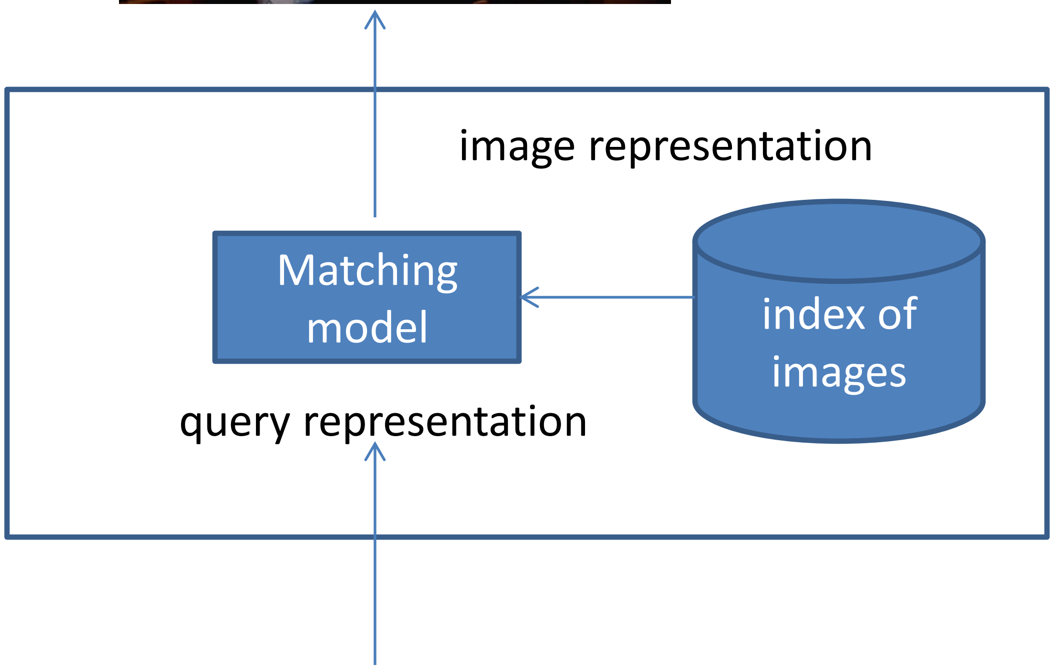
Lin Ma

Lifeng Shang

Zhaopeng Tu

Image Retrieval

Image Retrieval

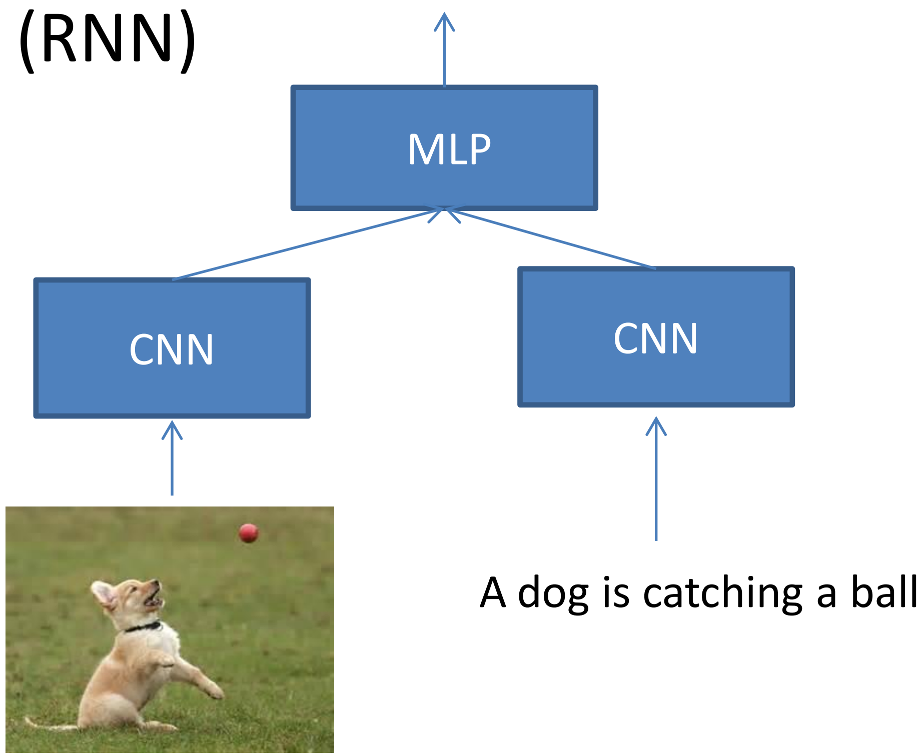


Find the picture that I had dinner with my friends at an Italian restaurant in Hong Kong

- Scenario
  - Image search on smartphone
  - Key: matching queries to images
- Technology
  - Deep model for matching text and image

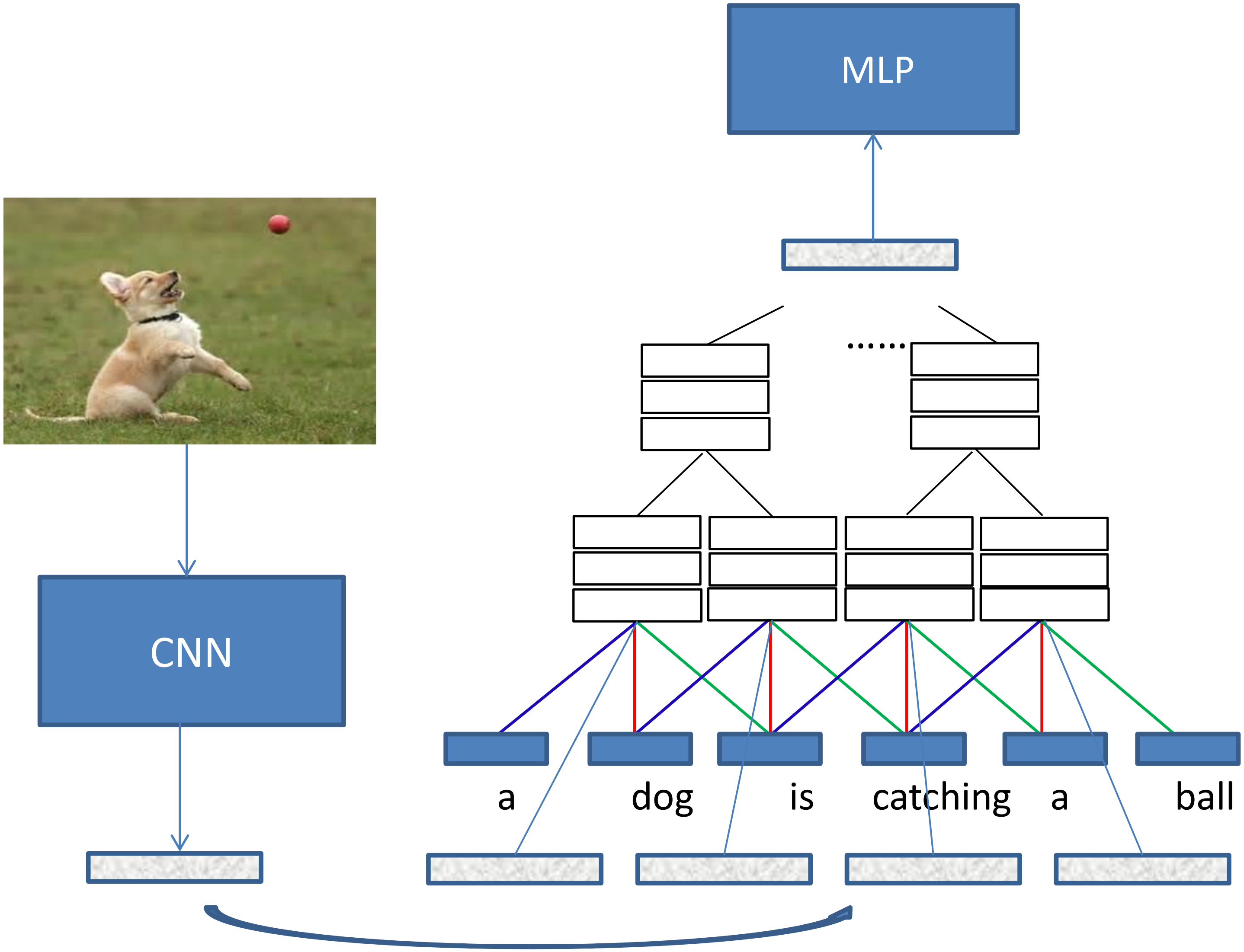
Deep Match Model for Image and Text

- Represent text and image as vectors and then match the two vectors
- Word-level matching, phrase-level matching, sentence-level Matching
- Our model (CNN) work better than state of the art models (RNN)



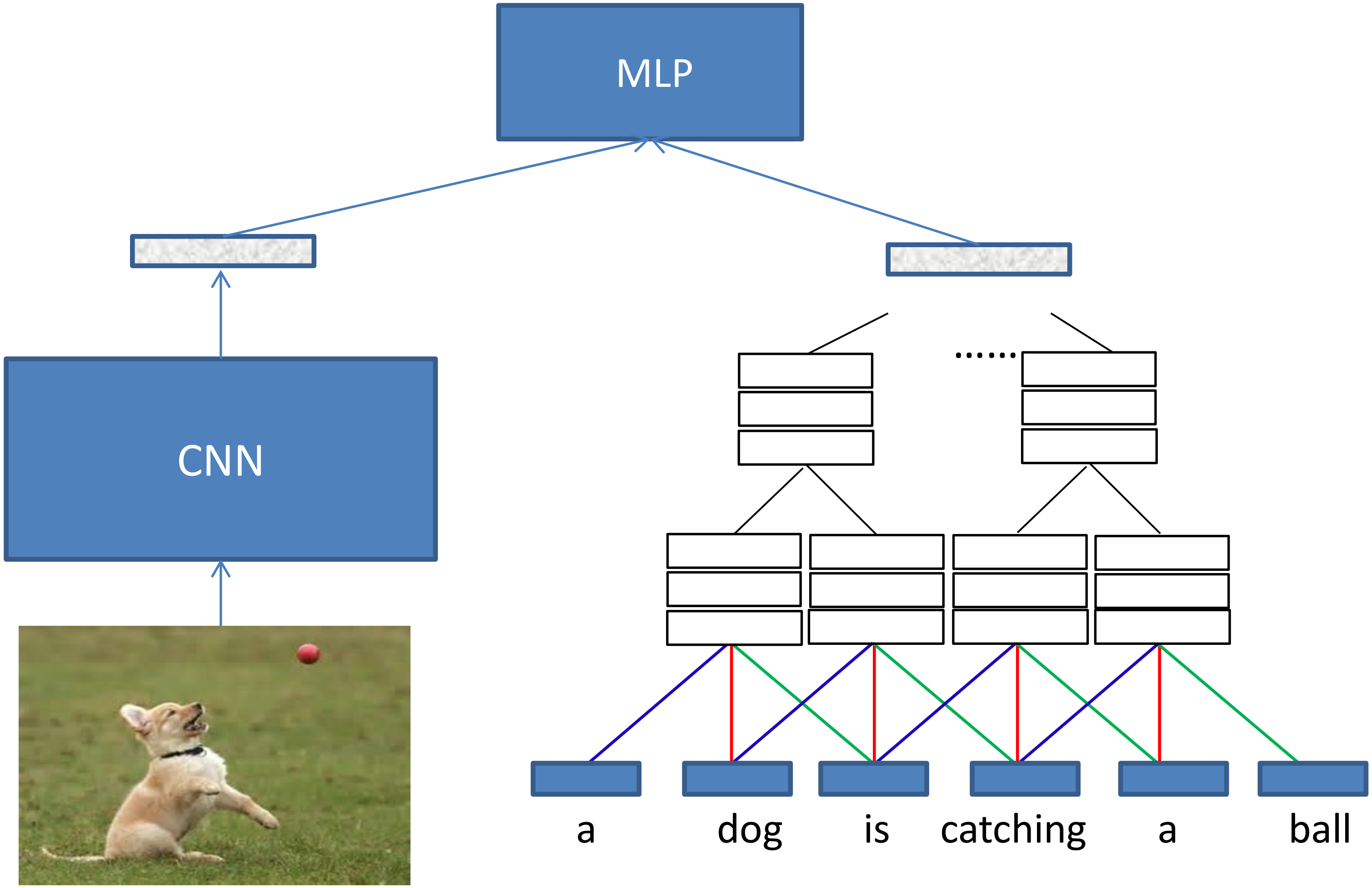
Ma et al., ICCV 2015

Word-level Matching Model



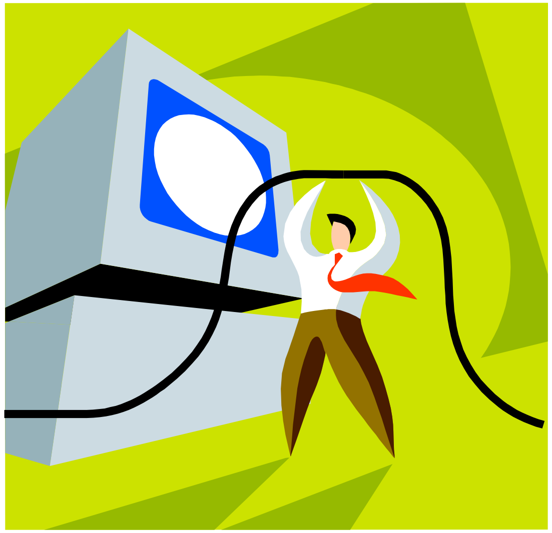
- Adding image vector to word vectors

Sentence-level Matching



- Combing image vector with sentence vector

Demo



Experimental Result

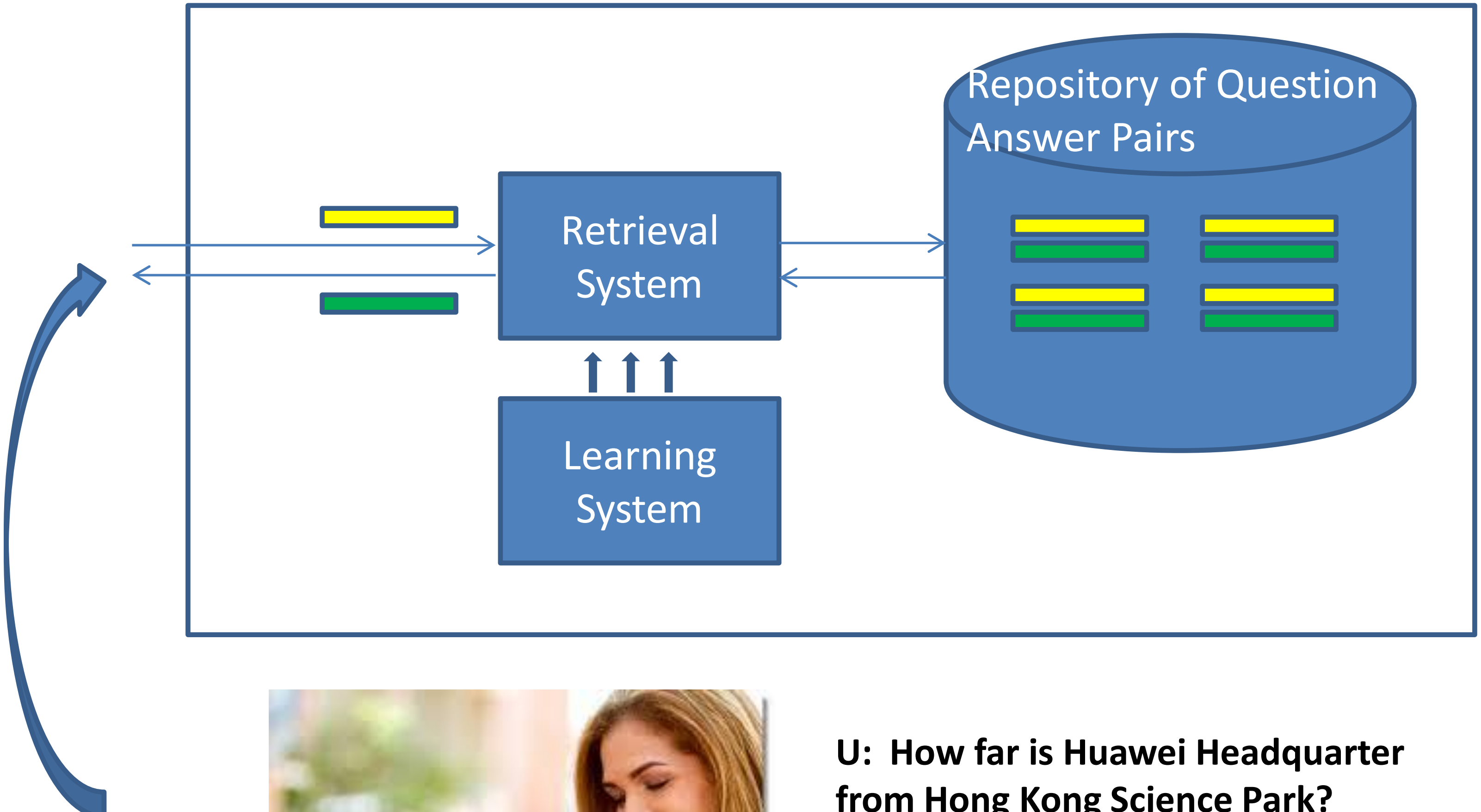
Flickr 30K images

	Sentence Retrieval				Image Retrieval			
	R@1	R@5	R@10	Med r	R@1	R@5	R@10	Med r
Random Ranking	0.1	0.6	1.1	631	0.1	0.5	1.0	500
DeViSE [6]	4.5	18.1	29.2	26	6.7	21.9	32.7	25
SDT-RNN [30]	9.6	29.8	41.1	16	8.9	29.8	41.1	16
MNLM [20]	14.8	39.2	50.9	10	11.8	34.0	46.3	13
MNLM-VGG [20]	23.0	50.7	62.9	5	16.8	42.0	56.5	8
m-RNN [24]	18.4	40.2	50.9	10	12.6	31.2	41.5	16
m-RNN-VGG [23]	**35.4**	63.8	73.7	**3**	22.8	50.7	63.1	5
Deep Fragment [16]	14.2	37.7	51.3	10	10.2	30.8	44.2	14
RVP (T) [3]	11.9	25.0	47.7	12	12.8	32.9	44.5	13
RVP (T+I) [3]	12.1	27.8	47.8	11	12.7	33.1	44.9	12.5
DVSA (DepTree) [17]	20.0	46.6	59.4	5.4	15.0	36.5	48.2	10.4
DVSA (BRNN) [17]	22.2	48.2	61.4	4.8	15.2	37.7	50.5	9.2
NIC [34]	17.0	*	56.0	7	17.0	*	57.0	7
LRCN [5]	*	*	*	*	17.5	40.3	50.8	9
OverFeat [28]:								
m-CNN$_{wd}$	12.7	30.2	44.5	14	11.6	32.1	44.2	14
m-CNN$_{phs}$	14.4	38.6	49.6	11	12.4	33.3	44.7	14
m-CNN$_{phl}$	13.8	38.1	48.5	11.5	11.6	32.7	44.1	14
m-CNN$_{st}$	14.8	37.9	49.8	11	12.5	32.8	44.2	14
m-CNN$_{ENS}$	20.1	44.2	56.3	8	15.9	40.3	51.9	9.5
VGG [29]:								
m-CNN$_{wd}$	21.3	53.2	66.1	5	18.2	47.2	60.9	6
m-CNN$_{phs}$	25.0	54.8	66.8	4.5	19.7	48.2	62.2	6
m-CNN$_{phl}$	23.9	54.2	66.0	5	19.4	49.3	62.4	6
m-CNN$_{st}$	27.0	56.4	70.1	4	19.7	48.4	62.3	6
m-CNN$_{ENS}$	33.6	**64.1**	**74.9**	**3**	**26.2**	**56.3**	**69.6**	4

Our CNN Model outperforms all existing models using RNN

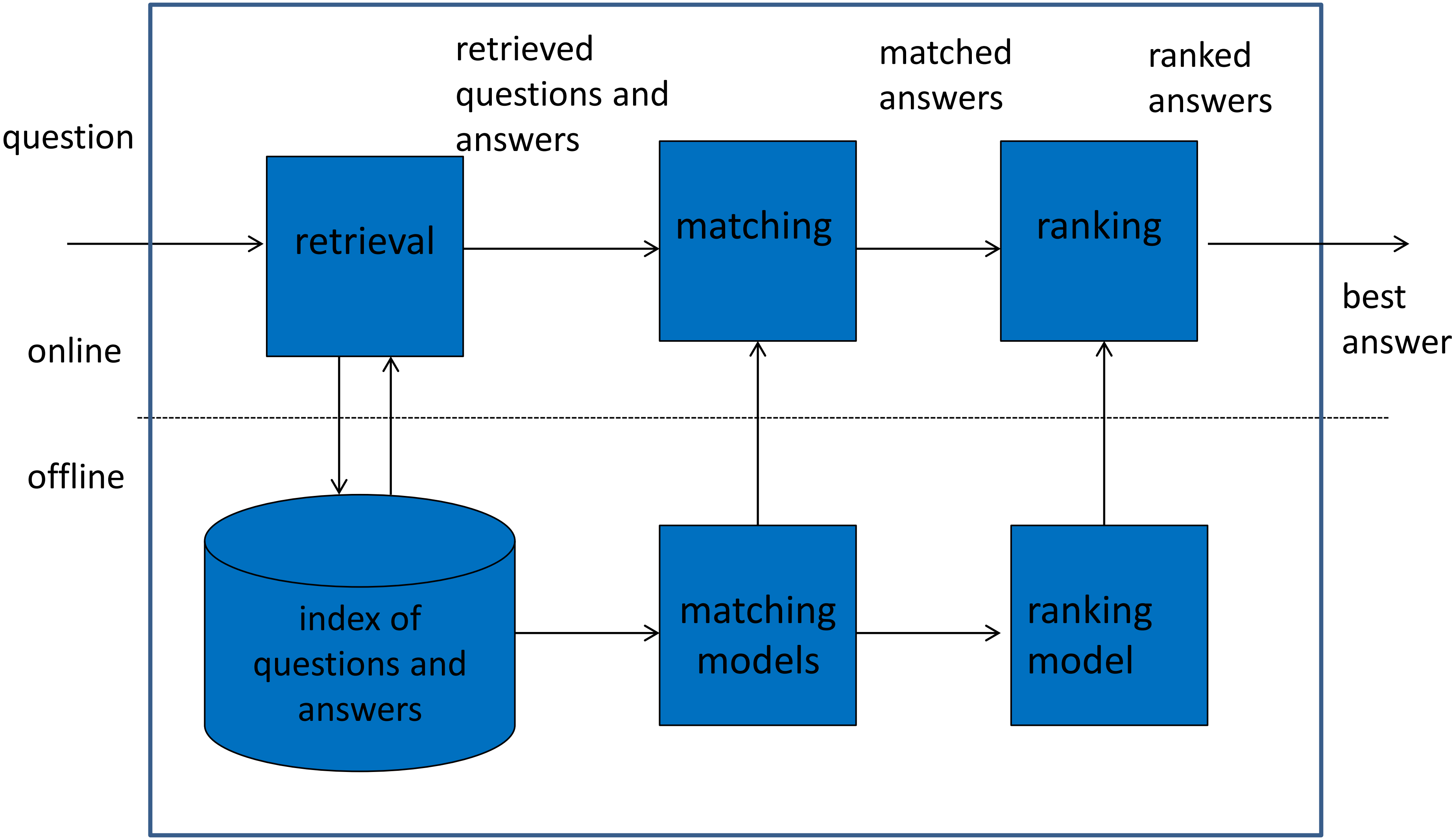
Retrieval based Question Answering

Retrieval-based Question Answering



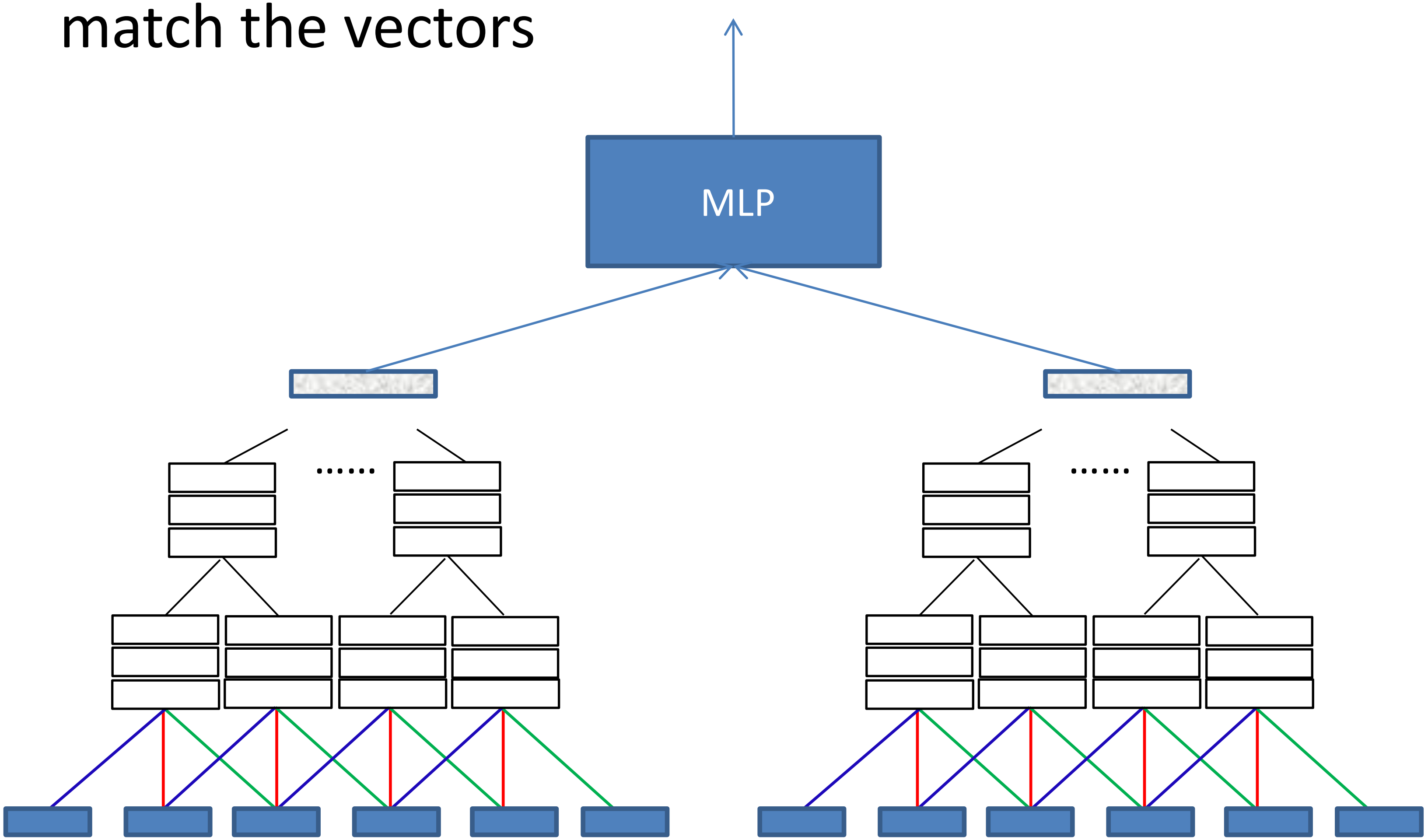
U: How far is Huawei Headquarter from Hong Kong Science Park?
P: About 30km
U: How can I get there?
P: You can first take MTR train to Lo Ma Chow and then take a taxi

Question Answering System - Retrieval based Approach



Deep Match CNN
- Architecture I

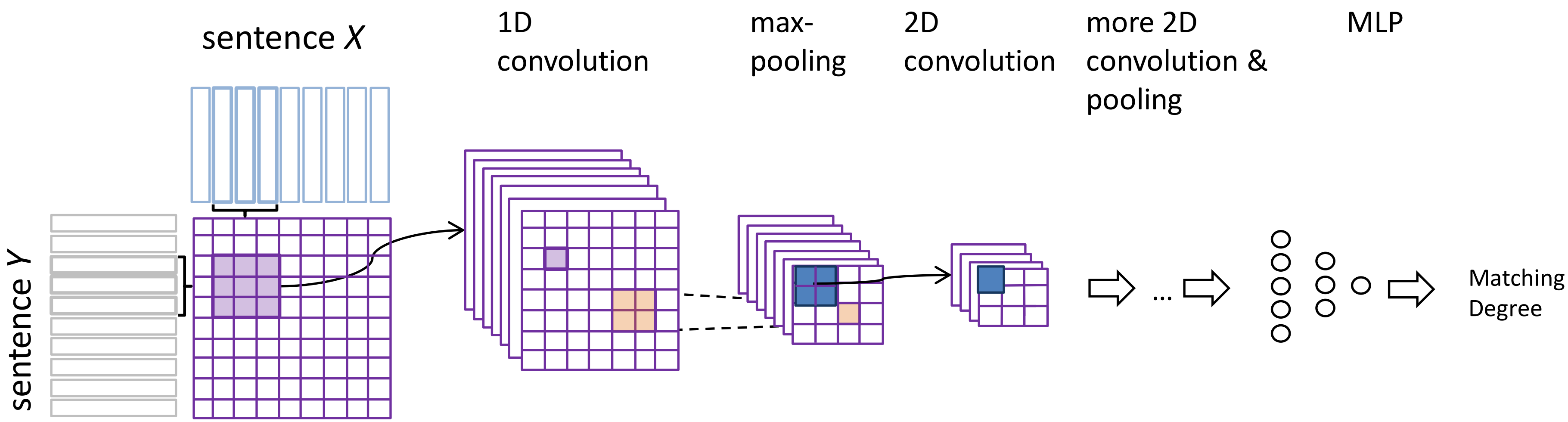
- First represent two sentences as vectors, and then match the vectors



Hu et al., NIPS 2014

Deep Match CNN - Architecture II

- Represent and match two sentences simultaneously
- Two dimensional model



Retrieval based Approach:

Accuracy = 70%

上海今天好熱，堪比新加坡。

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

上海今天热的不一般。

It is unusually hot.

想去武当山 有想同游的么？

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

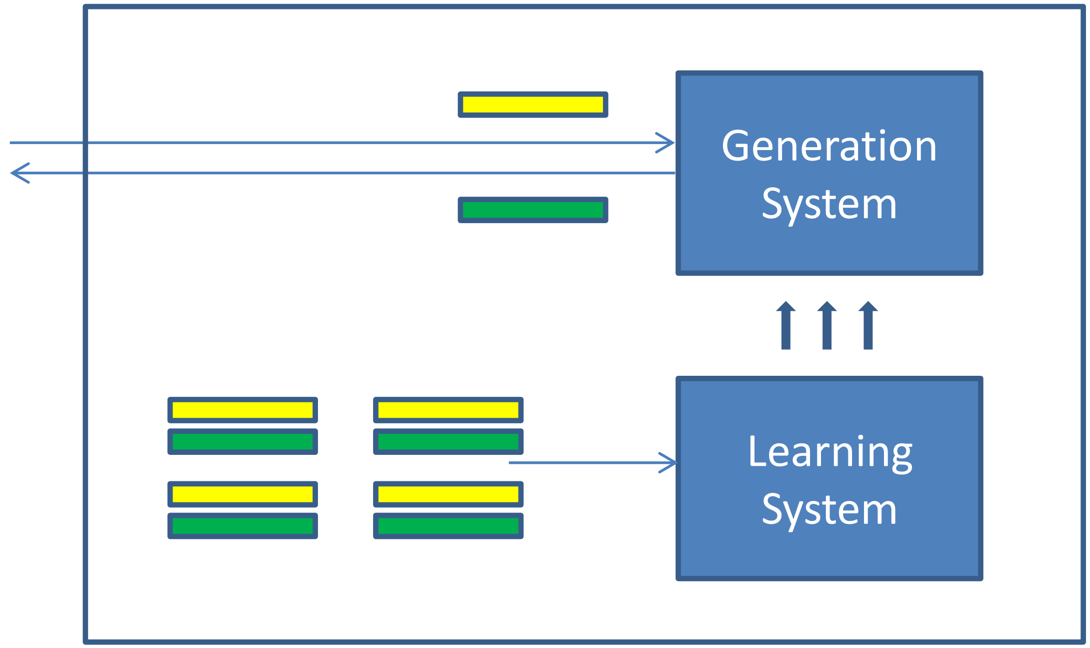
我想跟帅哥同游~哈哈

Haha, I want to go with you, handsome boy

Using 5 million Weibo Data

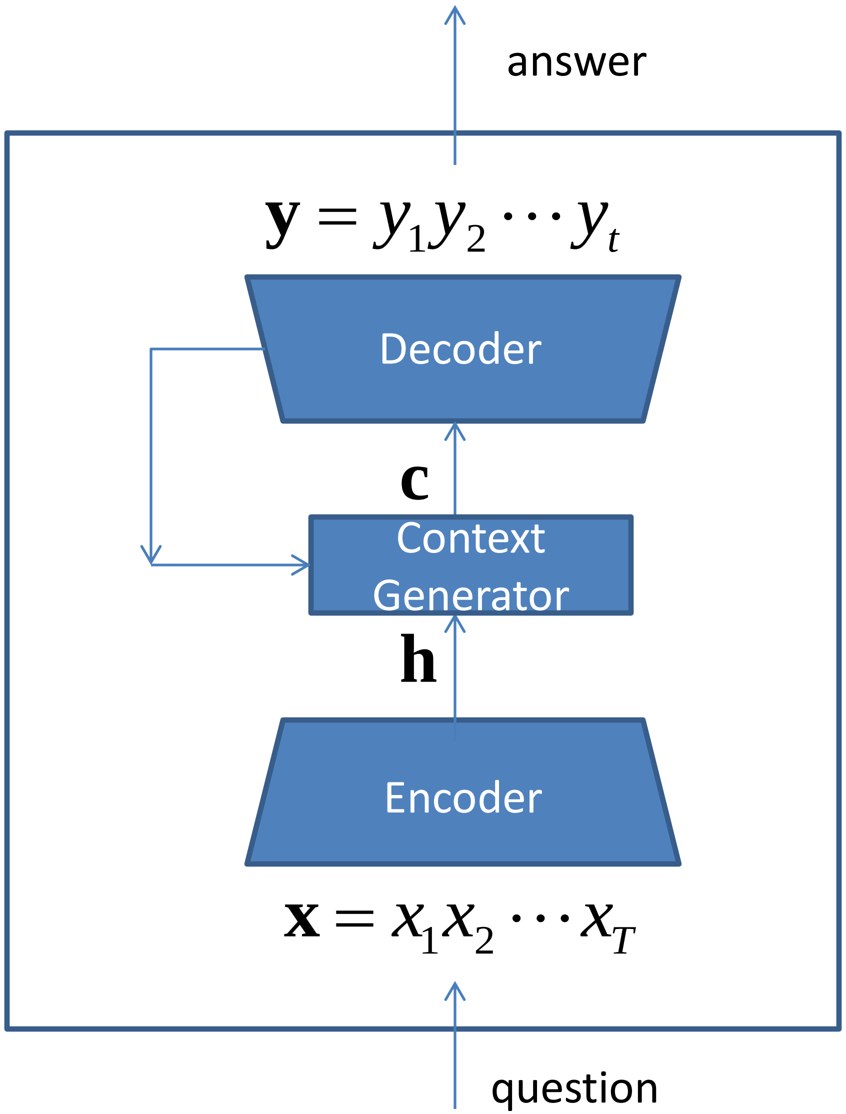
Generation based Question Answering

Generation-based Question Answering



U: How far is Huawei Headquarter from Hong Kong Science Park?
P: About 30km
U: How can I get there?
P: You can first take MTR train to Lo Ma Chow and then take a taxi

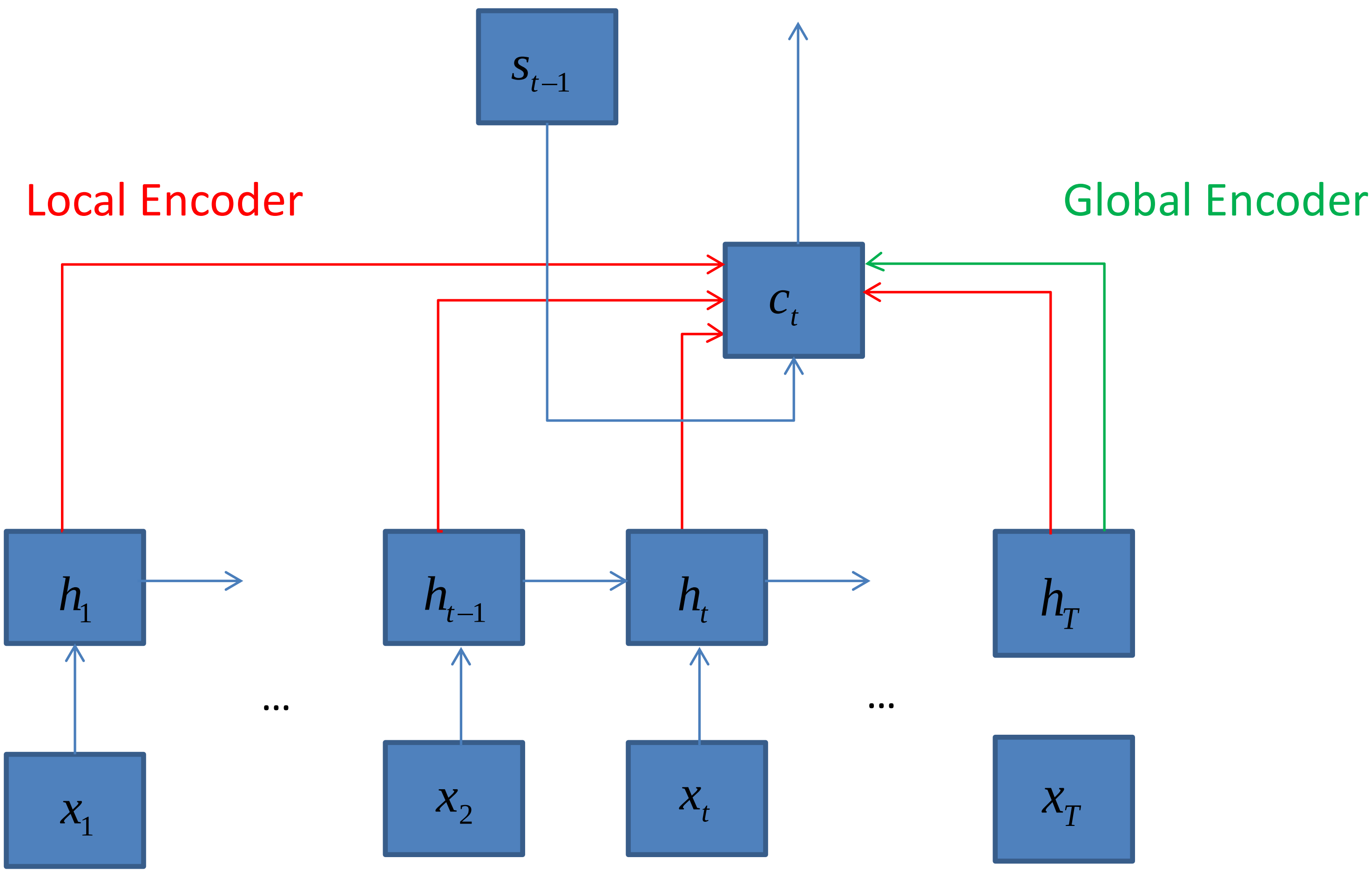
Natural Language Dialogue System - Generation based Approach



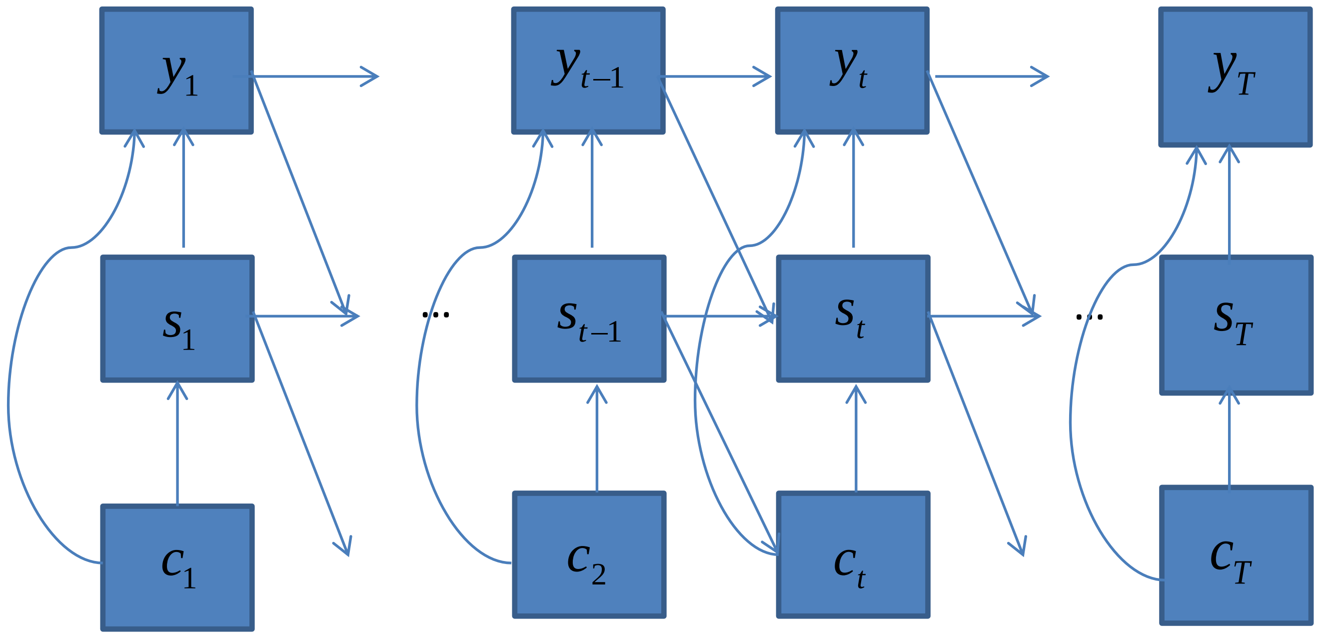
- Encoding questions to intermediate representations
- Decoding intermediate representations to answers
- Recurrent Neural Network (RNN)

Shang et al., ACL 2015

Encoder



Decoder



Generation based Approach

Accuracy = 76%

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?

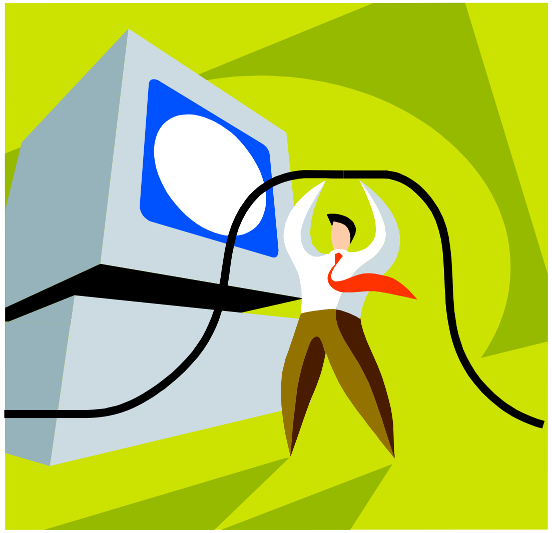
我想买三星手机。 I want to buy a Samsung phone

还是支持一下国产的吧。 Let us support our national brands

vs. Accuracy of translation approach = 26%

Accuracy of retrieval based approach = 70%

Demo

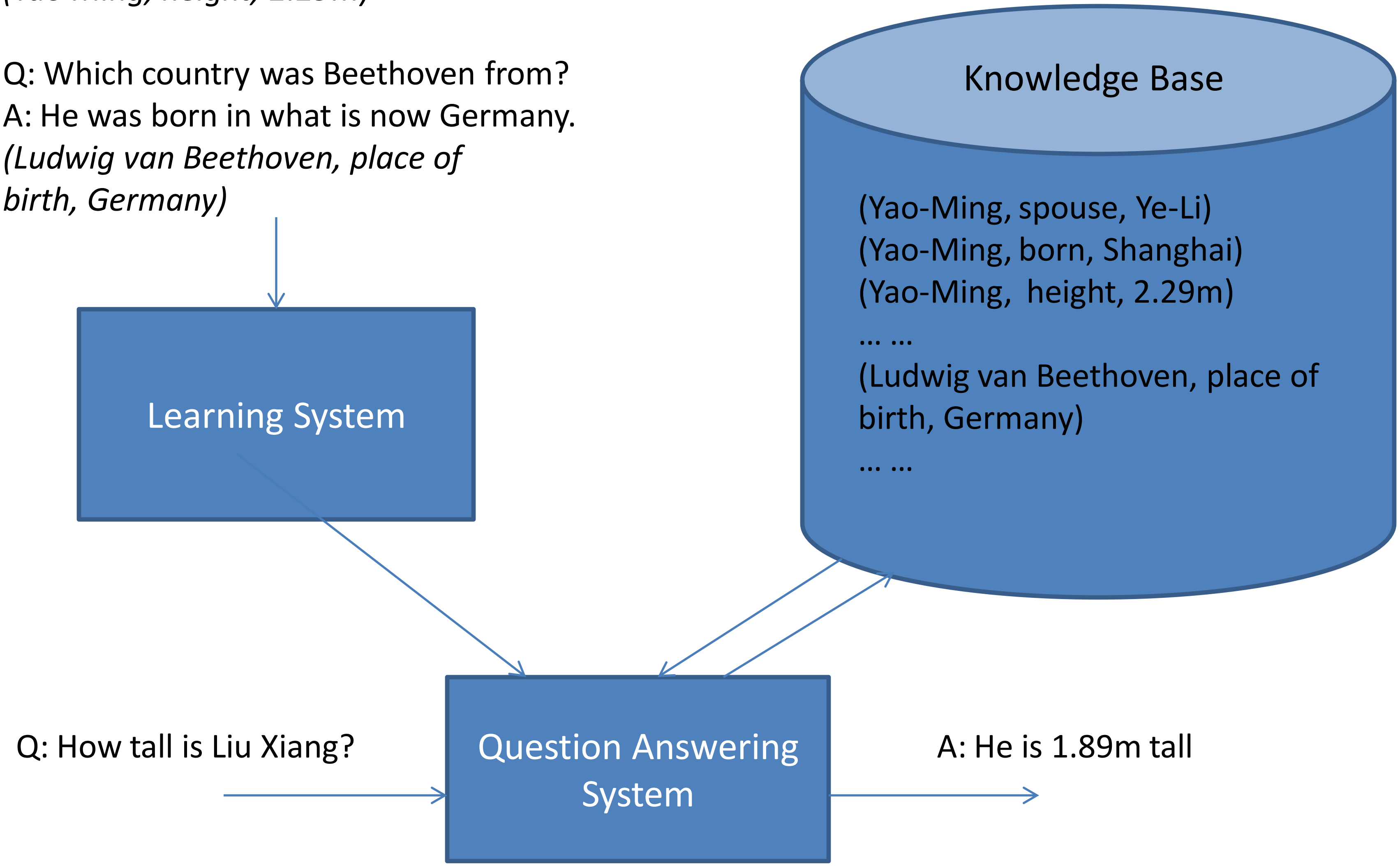


Question Answering from Knowledge Base

Question Answering from Knowledge Base

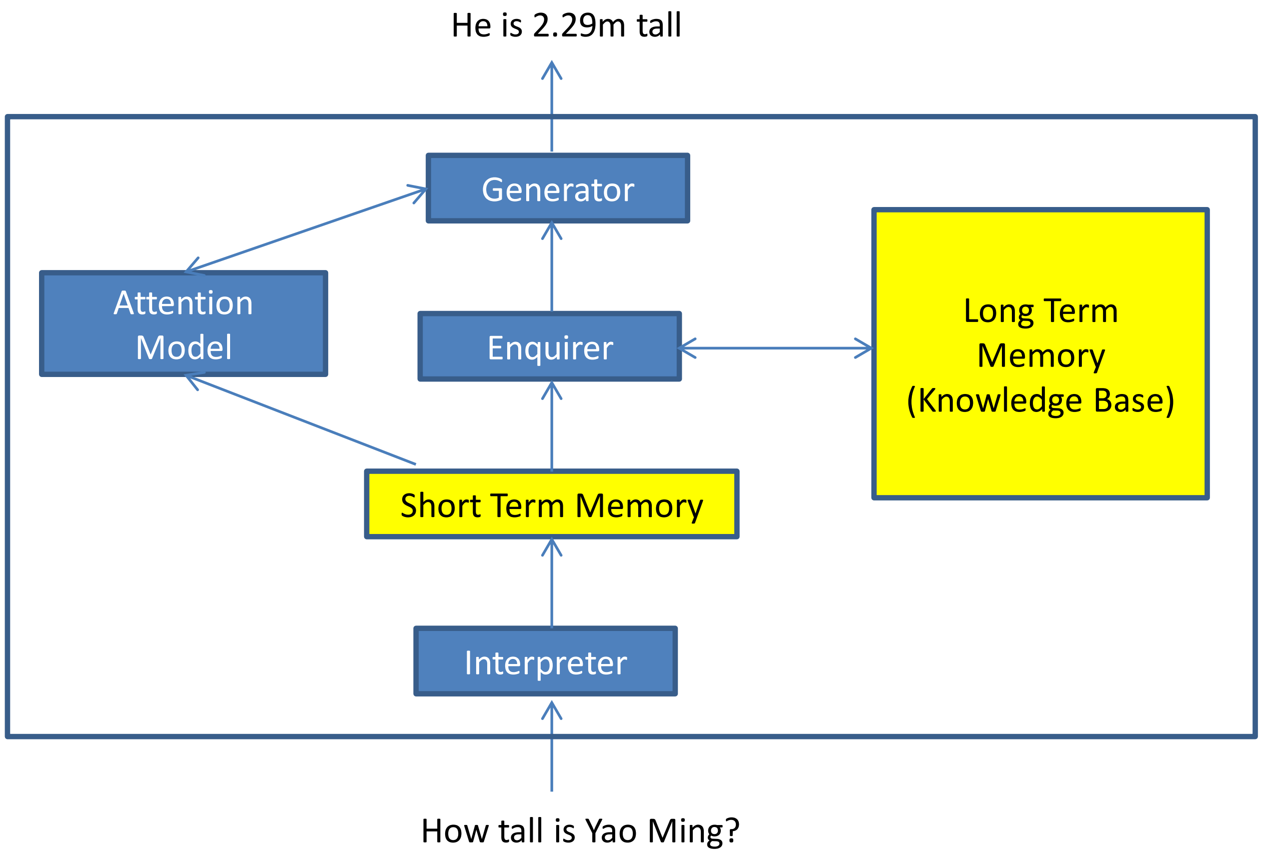
Q: How tall is Yao Ming?
A: He is 2.29m tall and is visible from space.
(Yao Ming, height, 2.29m)

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



GenQA

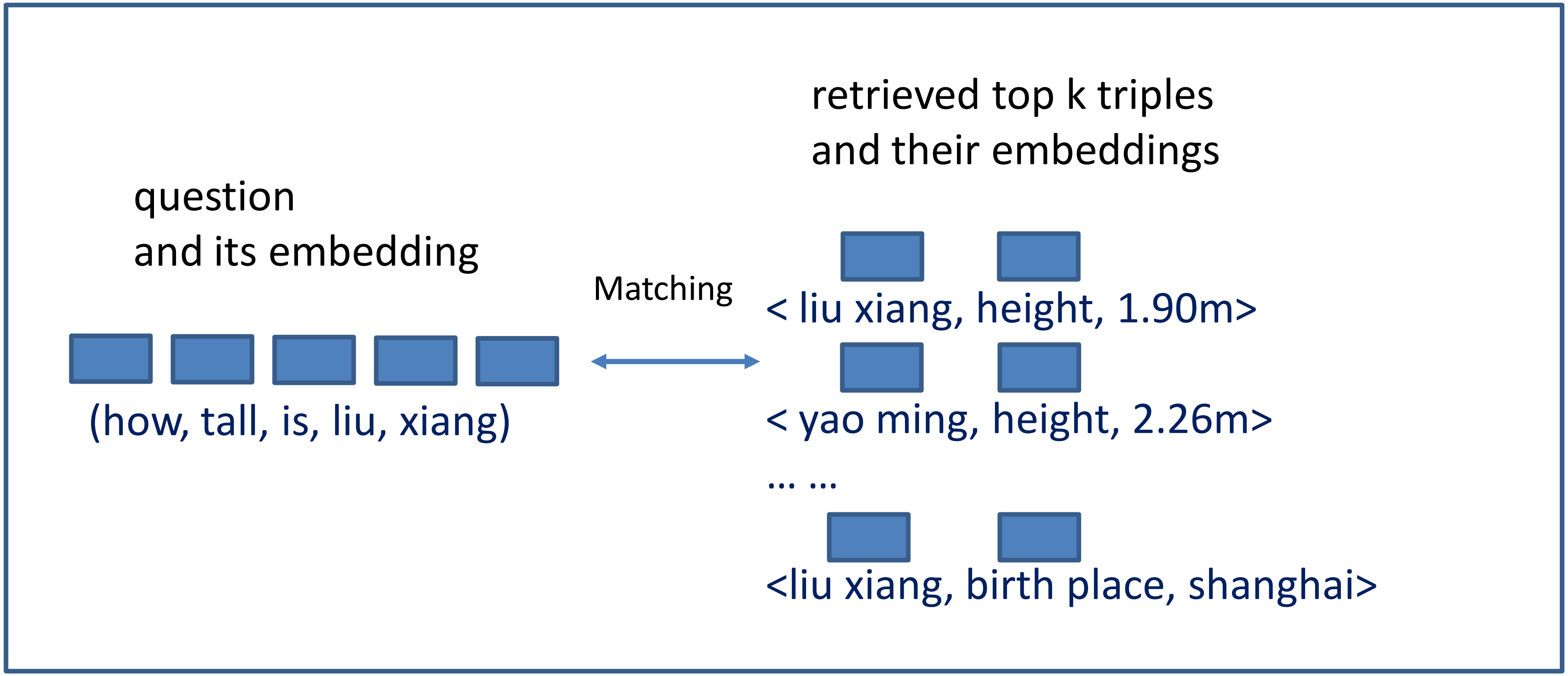
- Interpreter: creates representation of question using RNN
- Enquirer: retrieves top k triples with highest matching scores using CNN model
- Generator: generates answer based on question and retrieved triples using attention-based RNN
- Attention model: controls generation of answer



Key idea:
- Generation of answer based on question and retrieved result
- Combination of neural processing and symbolic processing

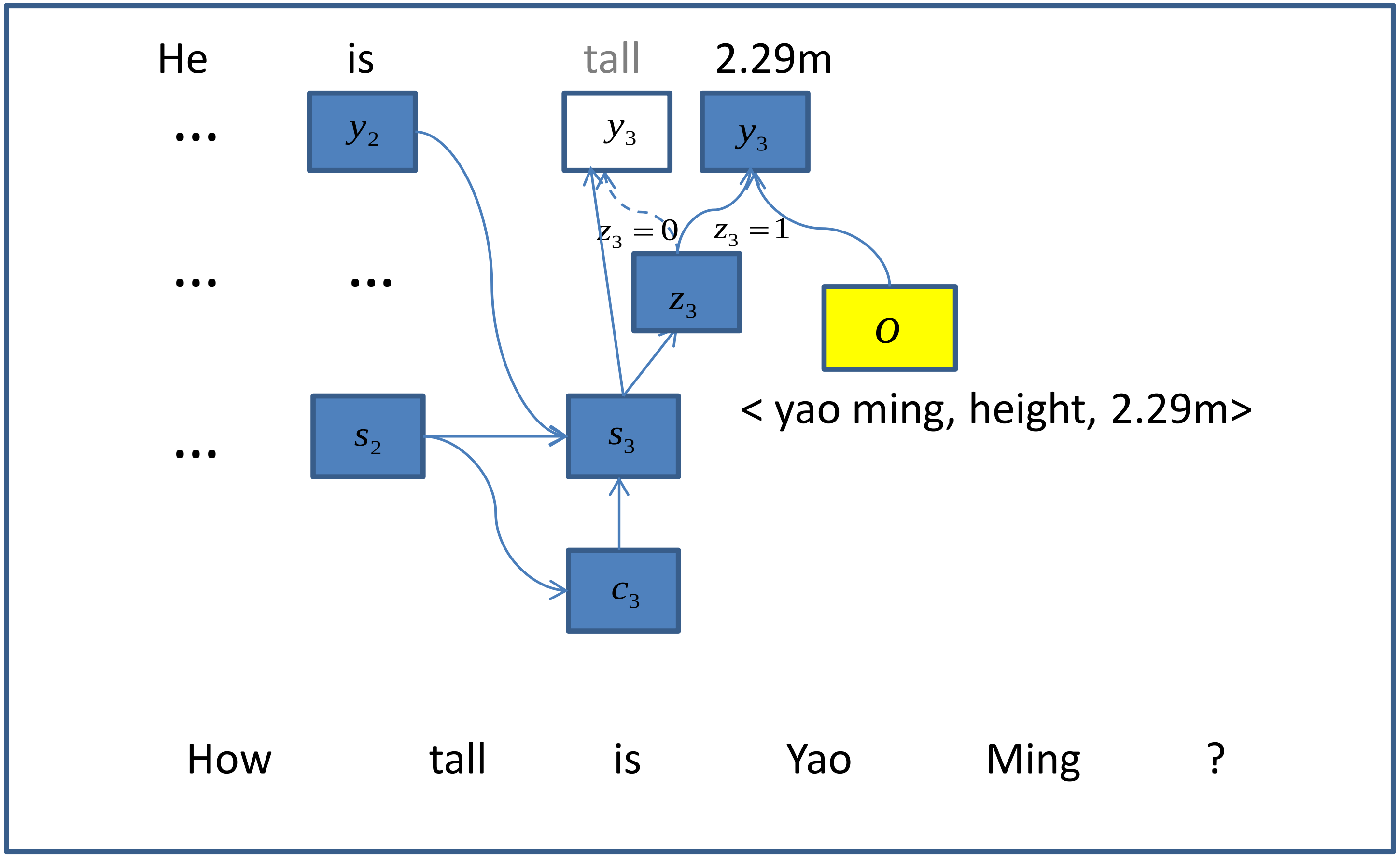
Yin et al. 2015

Enquirer: Retrieval and Matching



- Retaining both symbolic representations and vector representations
- Using question words to retrieve top k triples
- Calculating matching scores between question and triples using CNN model
- Finding best matched triples

Generator: Answer Generation



- Generating answer using attention mechanism
- At each position, a classifier decides whether to generate a word or use the object of top triple

Experimental Results
accuracy = 52%

ID	Questions	Generated Answers
1	三国演义谁写的？ Who wrote the Romance of the Three Kingdoms?	明朝的罗贯中。 Luo Guanzhong in Ming dynasty.
2	项羽是哪国人？ Which state does Xiang Yu belong to?	楚国的皇帝。 King of the Chu State.
3	还珠格格多少集？ How many episodes does My Fair Princess have?	一共24集。 24 episodes in total.
4	泰戈尔是哪国作家？ Which country is the writer Tagore from?	泰戈尔是印度著名的作家 。 Tagore is a famous Indian writer.
5	孙燕姿今年多大啊？ How old is Stefanie Sun this year?	32岁，他出生于1978年7月23日。 Thirty-two, he was born on July 23, 1978.
6	《格列佛游记》的作者是谁？ Who is the author of Gulliver's Travels?	法国作家斯威夫特。 The French writer, Swift.
7	怪物史莱克4什么时候上映呀？ When will Shrek Forever After be released?	上映日期：梦工厂电影发行公司。 Release date: DreamWorks Pictures.
8	凤凰卫视是哪里的电视台？ Which place is phoenix television located at?	总部在刘长乐。 The headquarters is at Liu Changle.

60K triples and 700K QA pairs for training

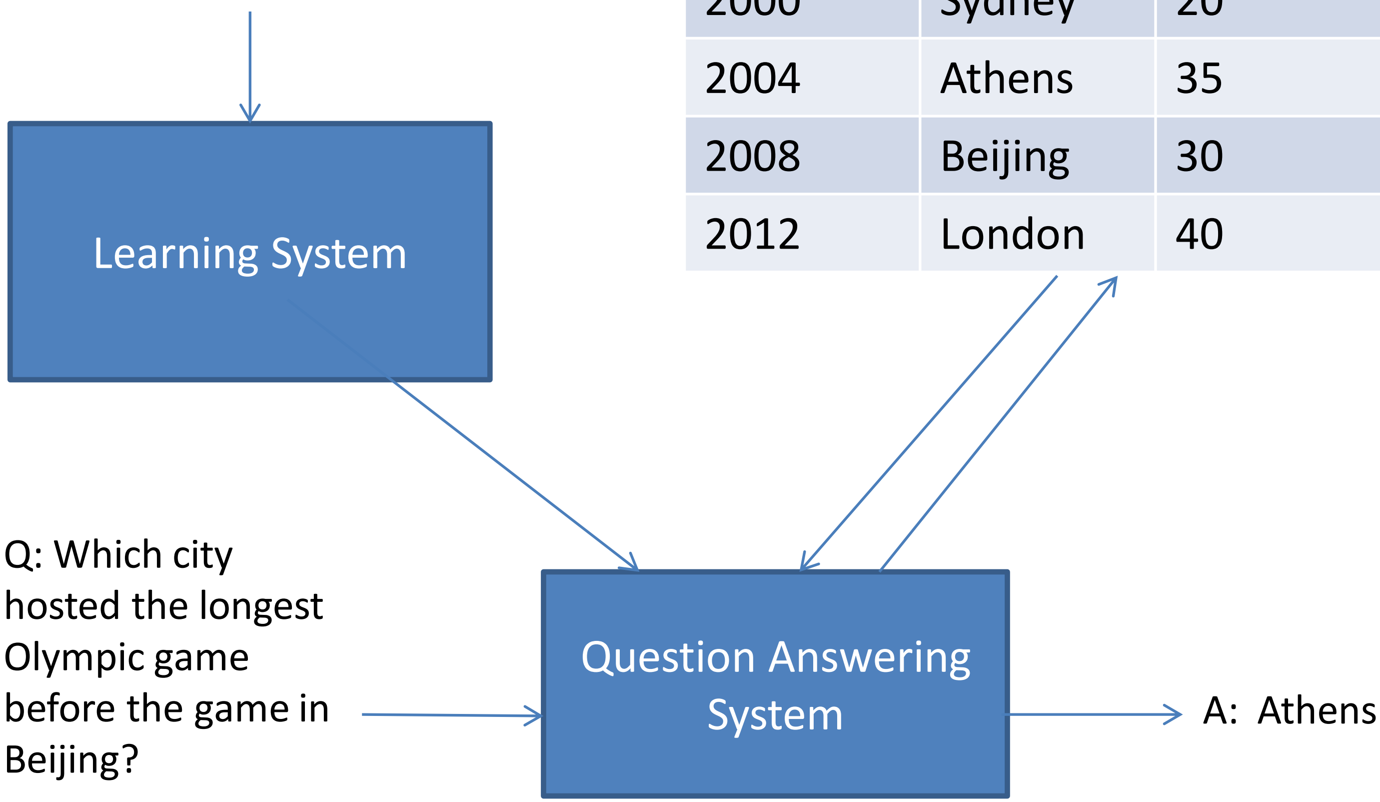
Question Answering from Relational Database

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)

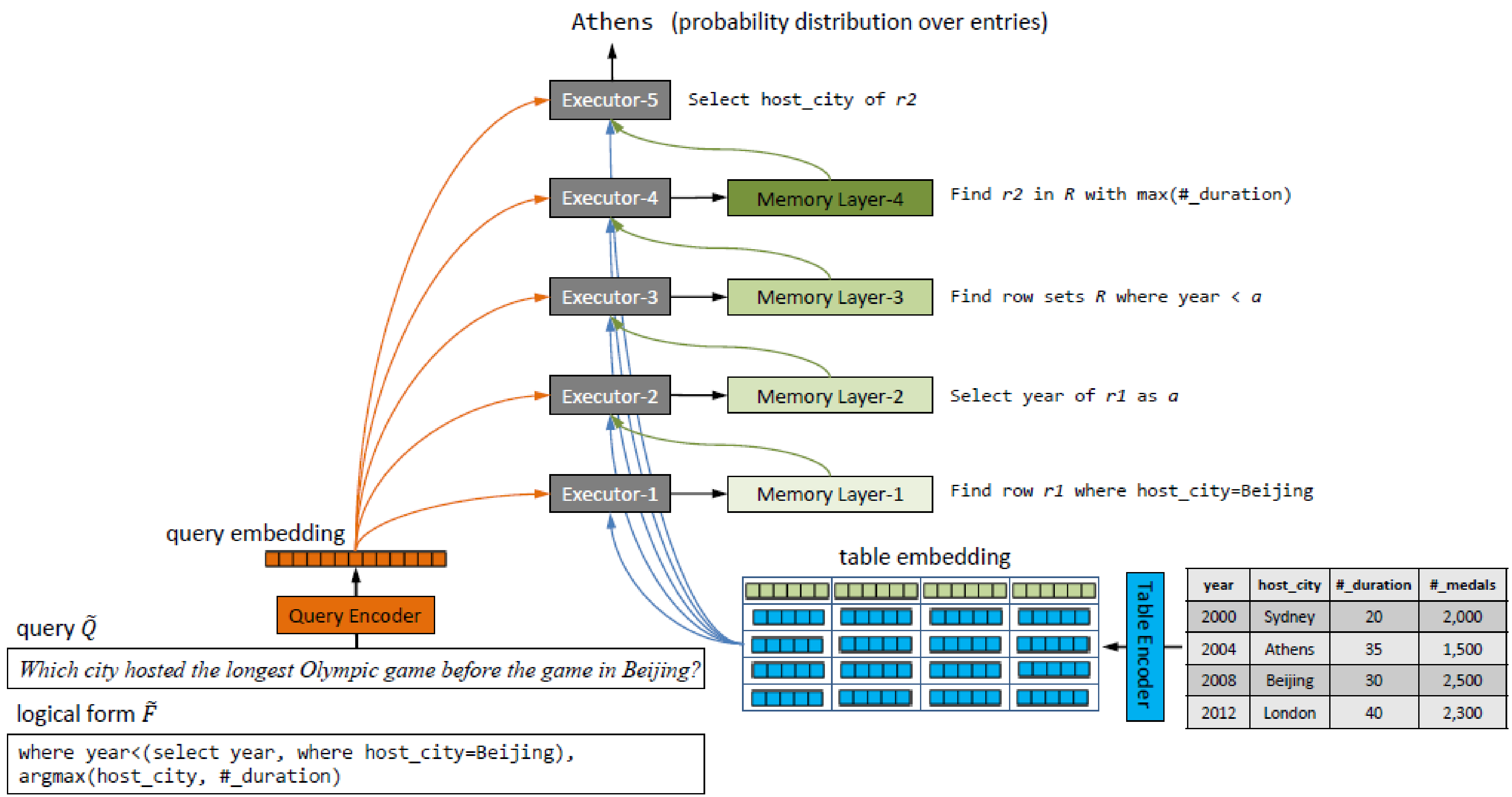
Relational Database

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



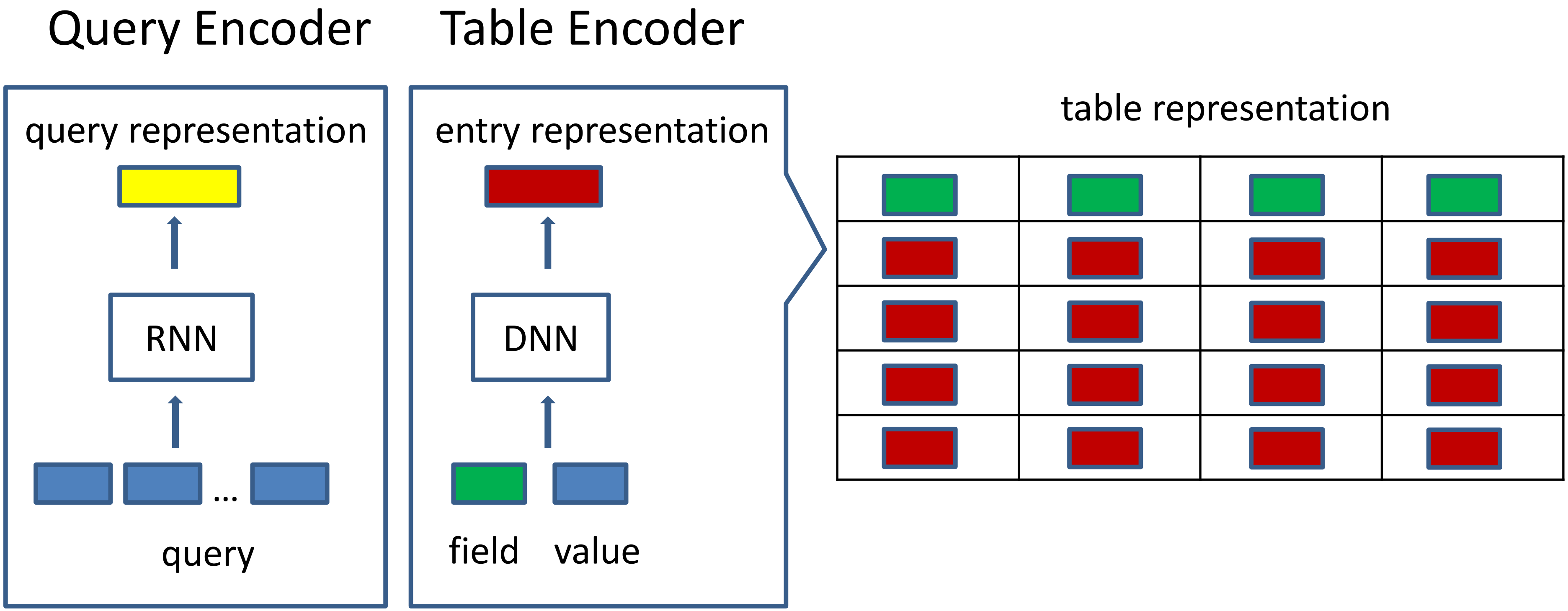
Neural Enquirer

- Query Encoder: encoding query
- Table Encoder: encoding entries in table
- Five Executors: executing query against table



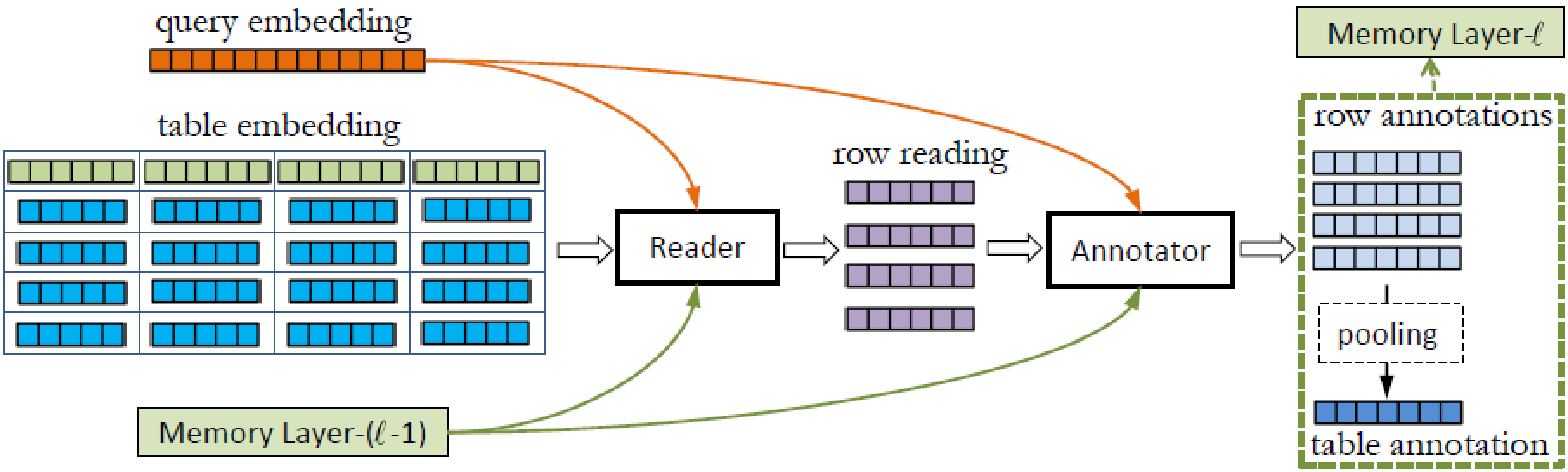
Yin et al. 2015

Query Encoder and Table Encoder



- Creating query embedding using RNN
- Creating entry table embedding for each entry using DNN

Executors



Select #_participants where city = beijing

- Five layers, except last layer, each layer has reader, annotator, and memory
- Reader fetches important representation for each row, e.g., city=beijing
- Annotator encodes result representation for each row, e.g., row where city=beijing

Experimental Results

	Mixtured-25K				Mixtured-100K		
	Sempre	N2N	SbS	N2N - OOV	N2N	SbS	N2N - OOV
Select_Where	93.8%	96.2%	99.7%	90.3%	99.3%	100.0%	97.6%
Superlative	97.8%	98.9%	99.5%	98.2%	99.9%	100.0%	99.7%
Where_Superlative	34.8%	80.4%	94.3%	79.1%	98.5%	99.8%	98.0%
Nest	34.4%	60.5%	92.1%	57.7%	64.7%	99.7%	63.9%
Overall Acc.	65.2%	84.0%	96.4%	81.3%	90.6%	99.9%	89.8%

- Experiments on synthetic data
- Outperforms Semantic Parser

Talk Outline

- Introduction to Huawei Noah's Ark Lab
- Deep Learning – New Opportunities for Information Retrieval
- Three Useful Deep Learning Tools
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Knowledge Base
  - Question Answering from Database
- Discussions and Concluding Remarks

Discussions

- Key is to combine symbolic processing and neural processing
- Advantage of symbolic processing: direct, effective, and easy to control
- Advantage of neural processing: flexible, robust, and completely data-driven
- Challenge: difficult to make the combination

Summary

- Matching is key for Information Retrieval
- Deep Learning provides new opportunities for IR
- Can learn better representations for matching
- Information Retrieval Tasks
  - Image Retrieval
  - Retrieval-based Question Answering
  - Generation-based Question Answering
  - Question Answering from Knowledge Base
  - Question Answering from Database Key question: how to combine symbolic processing and neural processing
- Future relies on combination of symbolic processing and neural processing

References

- Baotian Hu, Zhengdong Lu, Hang Li, Qingcai Chen. Convolutional Neural Network Architectures for Matching Natural Language Sentences. NIPS'14, 2042-2050, 2014.
- Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. ACL-IJCNLP'15, 2015.
- Lin Ma, Zhengodng Lu, Lifeng Shang, Hang Li . Multimodal Convolutional Neural Networks for Matching Image and Sentence, ICCV’15, 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.

Thank you!