Deep Learning for Natural Language Processing: Past, Present, and Future

Hang Li
Huawei Noah's Ark Lab

Key Messages of Talk

- Deep Learning Made Big Breakthroughs in Natural Language Processing
- Until Present: Purely Neural Processing
- Present and Future: Hybrid Approach Combining Neural Processing and Symbolic Processing
- Noah's Ark Lab Developed State-of-Art Models for Machine Translation, Natural Language Dialogue, Question Answering, and Image Retrieval

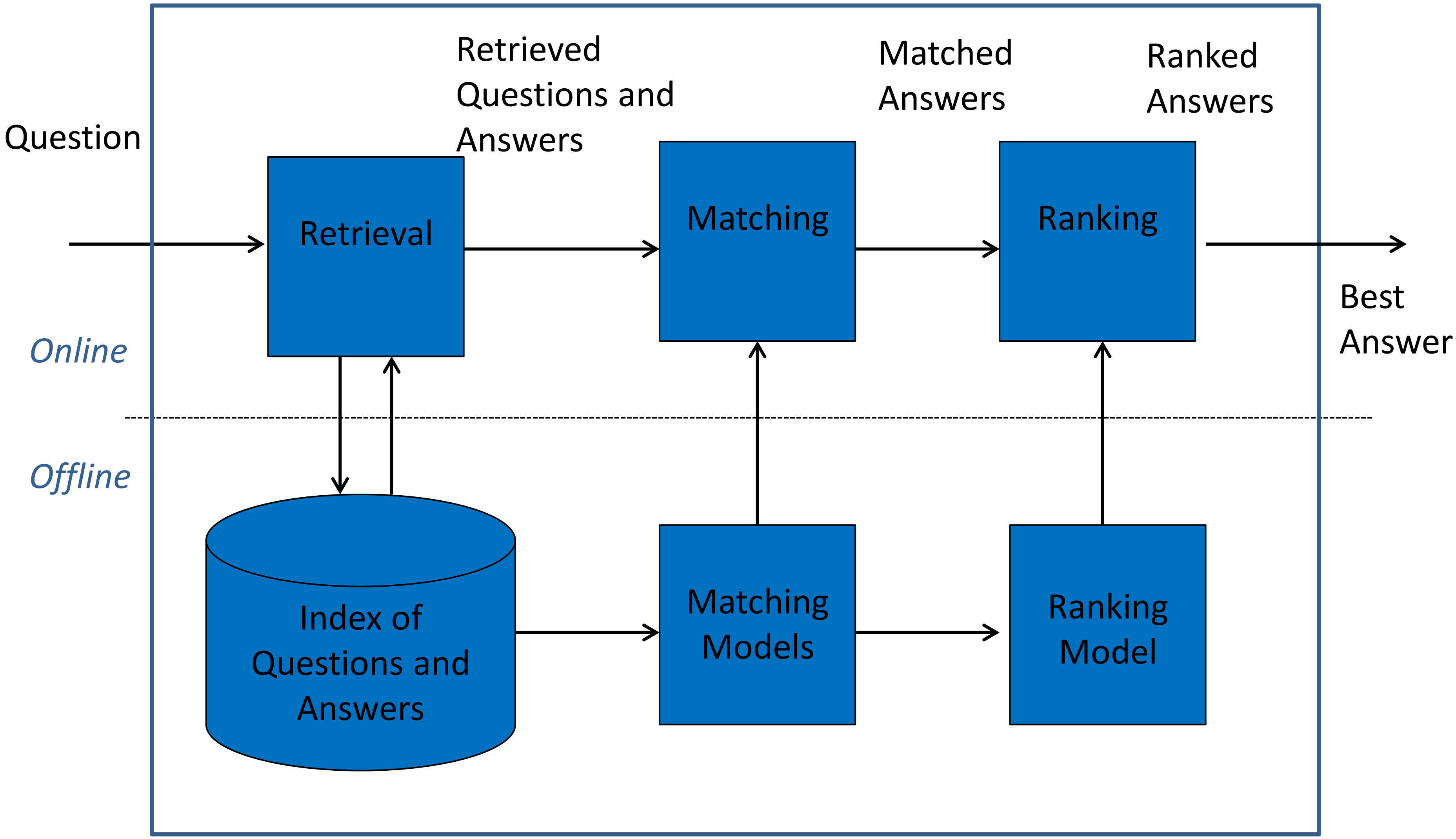
Until Present: Purely Neural Processing

Tasks and Our Work

- Question answering
  - DeepMatch CNN (convolutional neural network)
- Image retrieval
  - Multimodal CNN
- Machine translation
  - Coverage vector for NMT (neural machine translation)
  - Context gates for NMT
- Natural language dialogue
  - Neural Responding Machine

Question Answering
- DeepMatch CNN

Retrieval based Question Answering System



Deep Match Model CNN

- Represent and match two sentences simultaneously
- Two dimensional model
- State of art model for matching in question answering

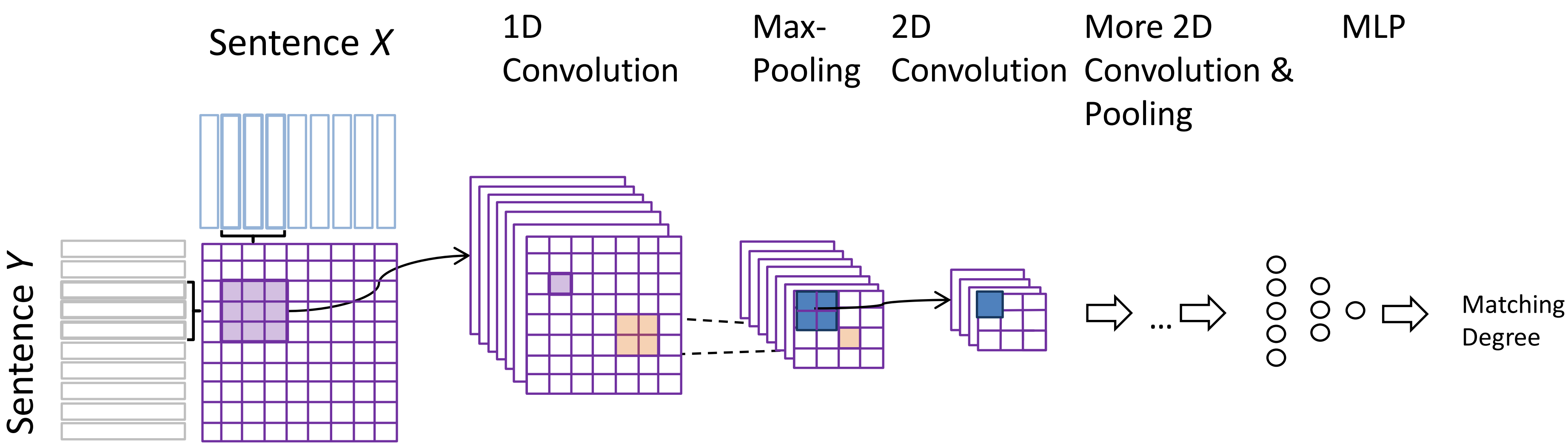
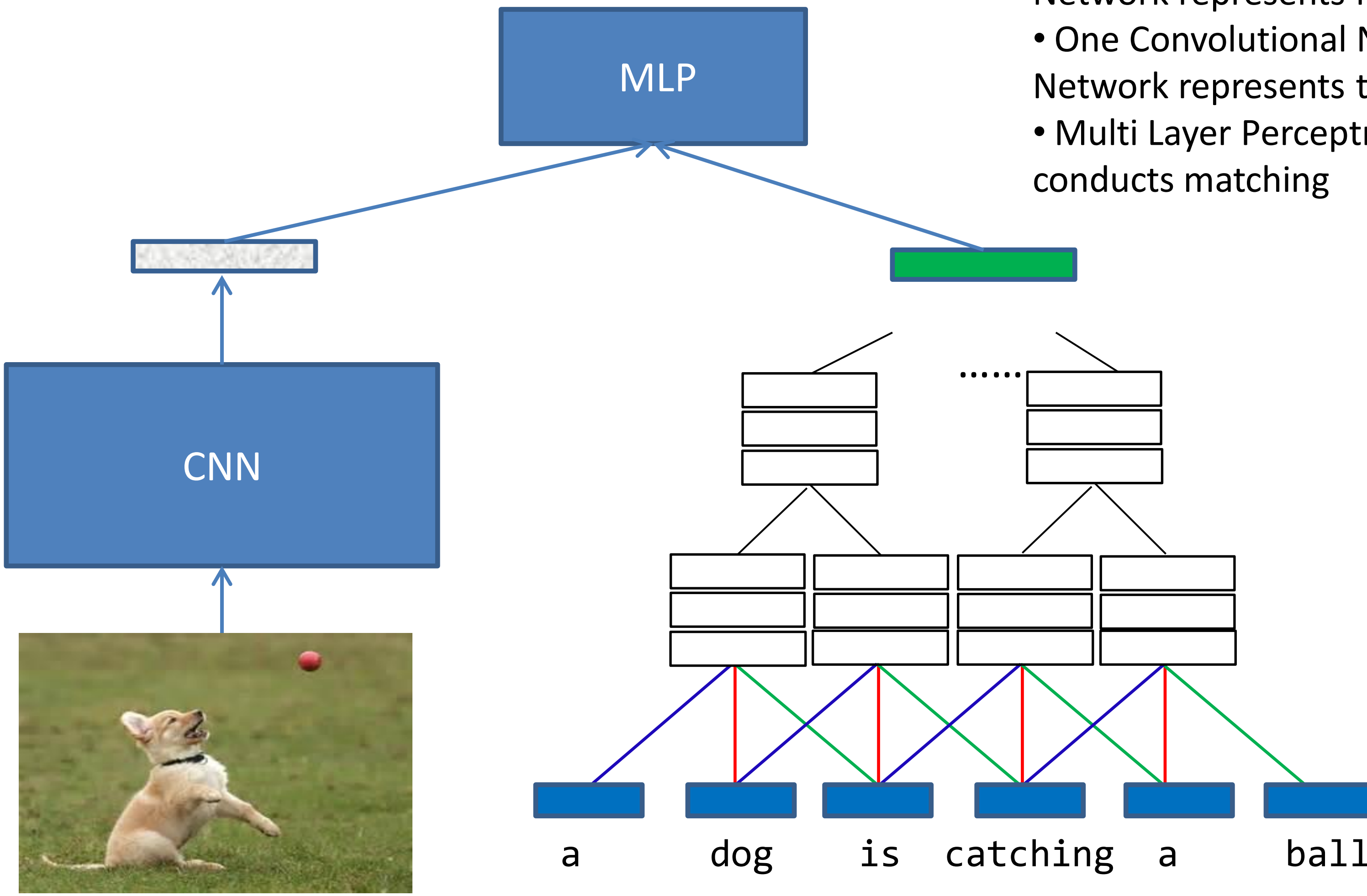


Image Retrieval
- Multimodal CNN

Demo

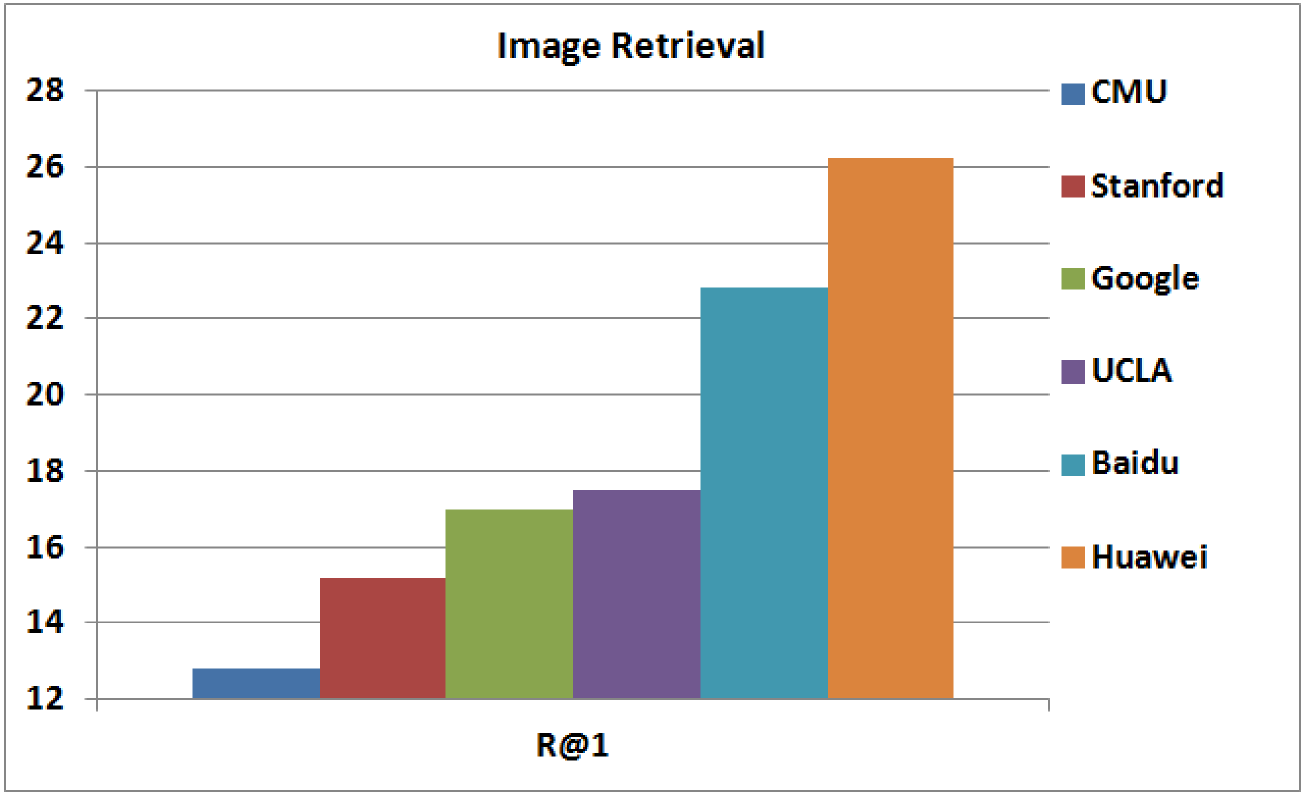
Multimodal CNN

- One Convolutional Neural Network represents image
- One Convolutional Neural Network represents text
- Multi Layer Perceptron conducts matching



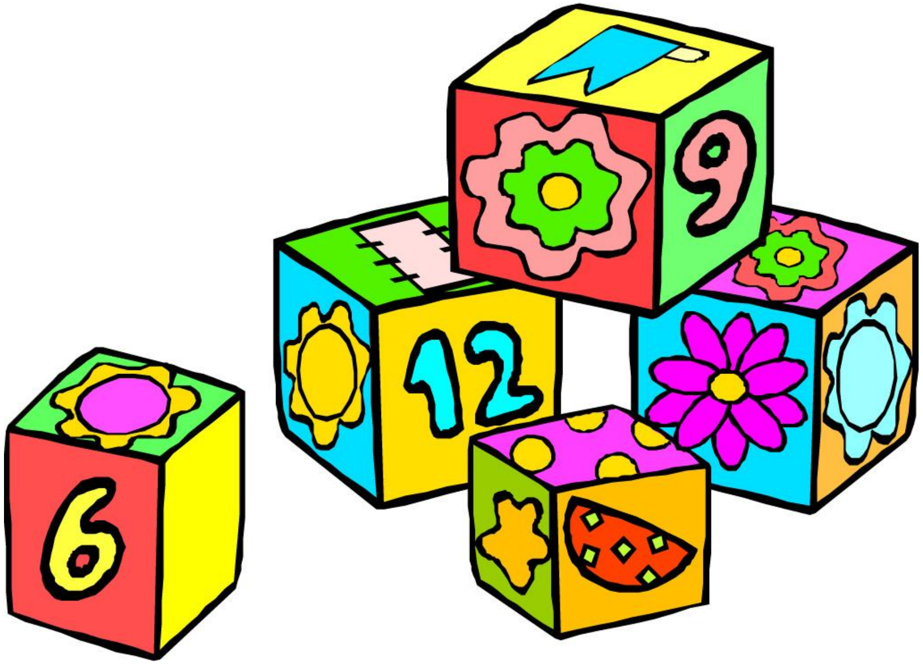
Experimental Results

- Experiment
  - Trained with 30K Flickr data
  - Outperforming other state-of-the-art models

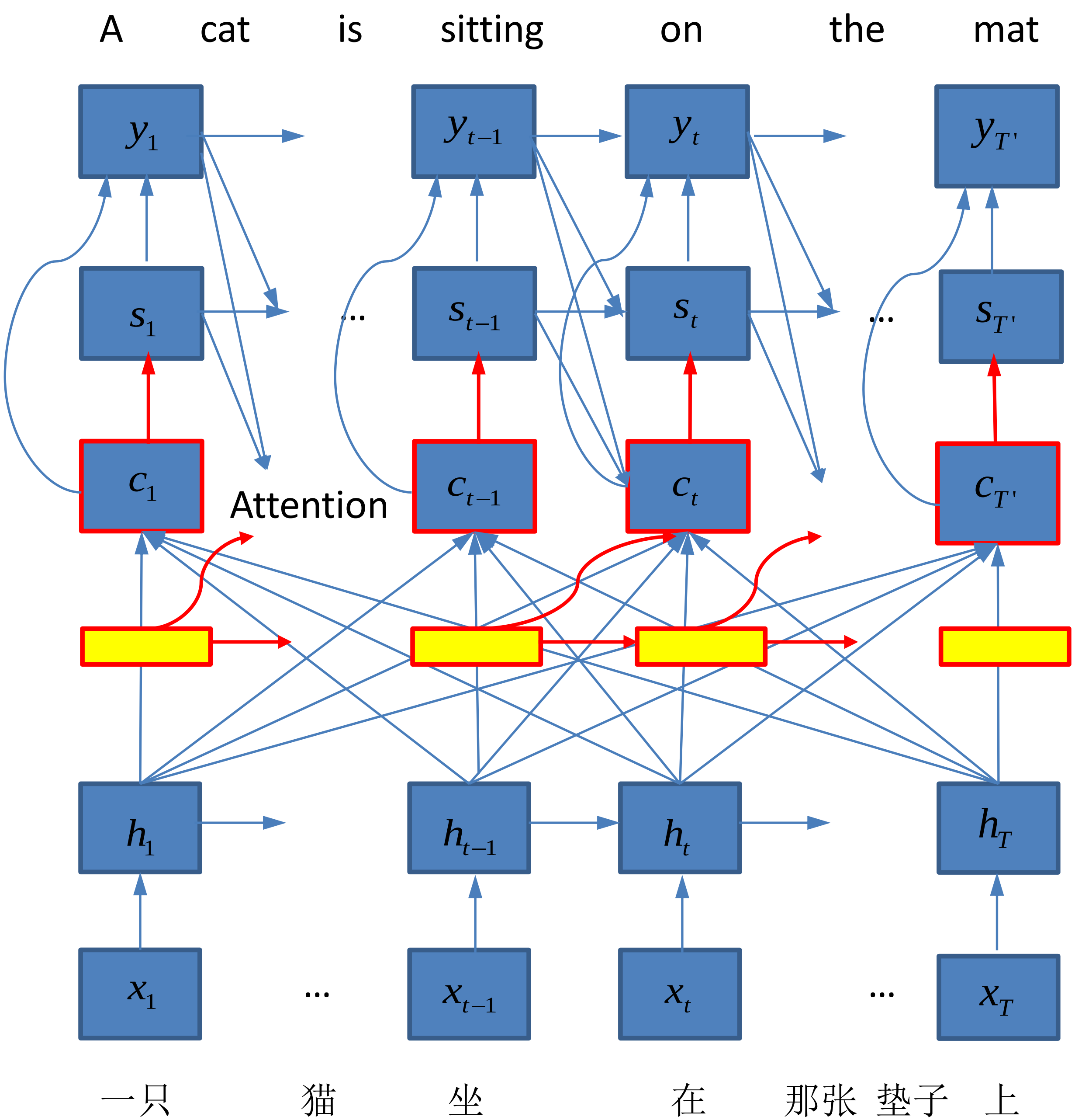


Neural Machine Translation

Demo



Neural Machine Translation



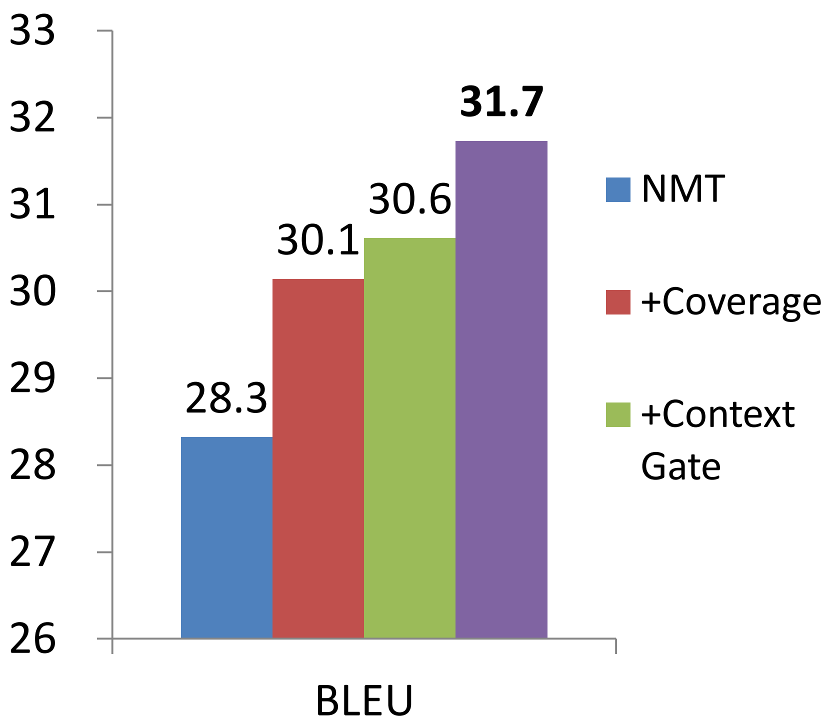
Decoder

- Using coverage vectors to avoid over-translation and under-translation
- Using context gates to dynamically control the impact of attention

Encoder

Experimental Results

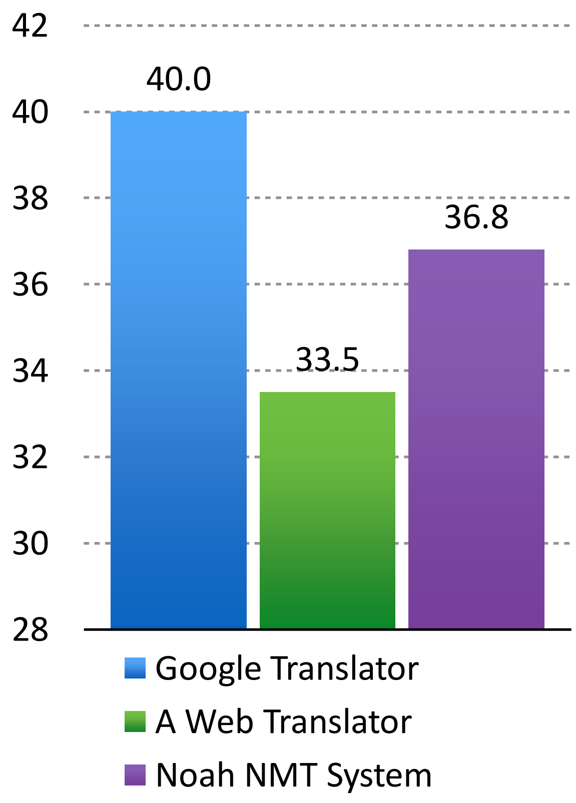
- Experiment
  - Trained with 1.25 million LDC data (Chinese-English)



C-E Translation	有一些恐怖袭击会愈演愈烈。
NMT	There will be some terrorist attacks.
+Both	Some terrorist attacks *will become more and more intense*.

Experimental Result

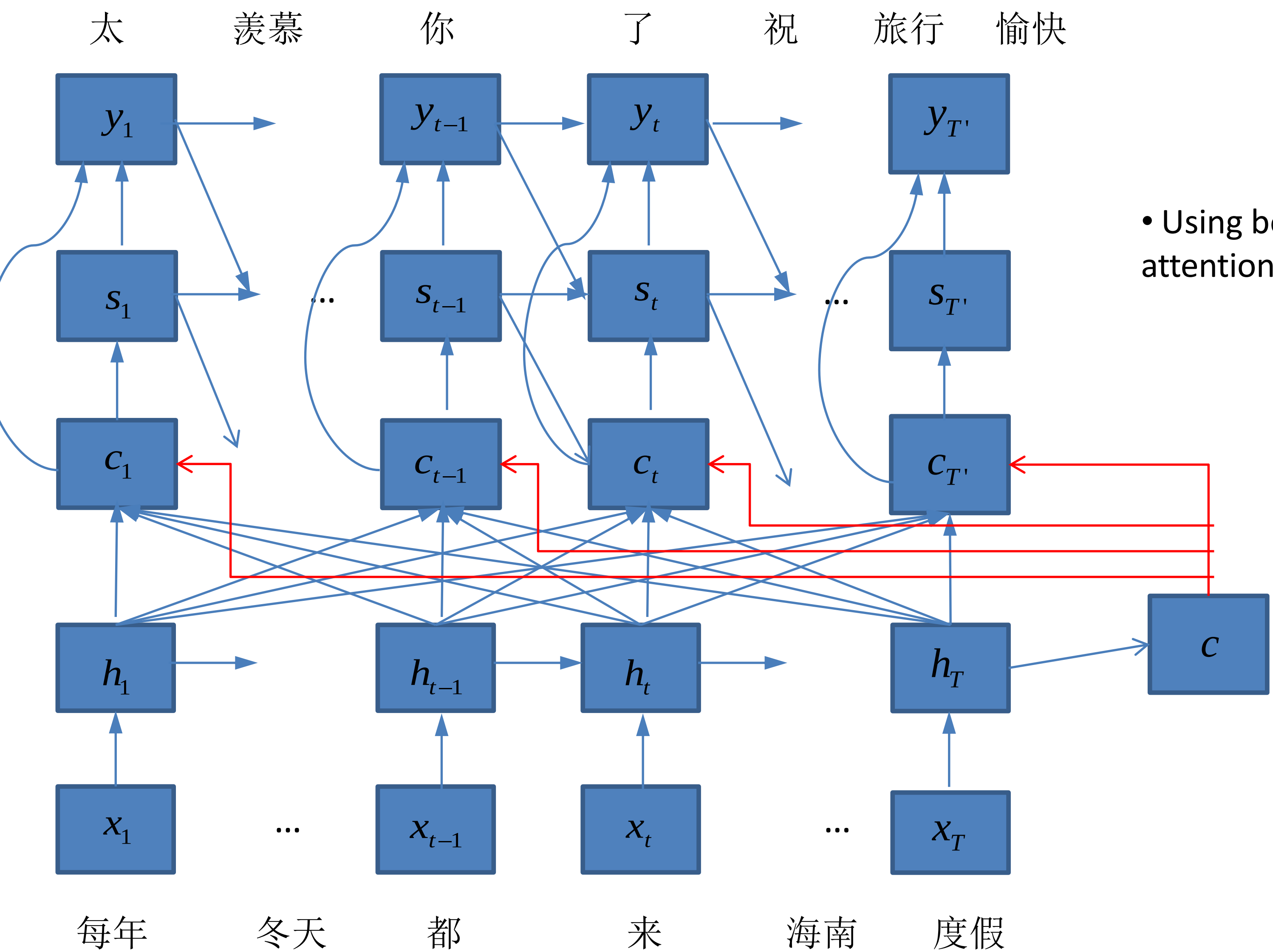
- Google NMT system works better, apparently due to its larger training data and more powerful computing architecture
- Google NMT system also employs coverage mechanism



Natural Language Dialogue
- Neural Responding Machine

Demo

Neural Responding Machine



- Using both local and global attention mechanisms

Experimental Results

- Experiment
  - Trained with 4.4 million Weibo data (Chinese)
  - 95% of responses are natural, 76% of responses are appropriate as replies

Message	Response
Occupy Central is finally over	Will Lujiazui (finance district in Shanghai) be the next?
I want to buy a Samsung phone	Let us support our national brand.

Present and Future: Hybrid Approach Combining Neural Processing and Symbolic Processing

Tasks and Our Work

- Question answering
  – GenQA
  – Neural Enquirer
- Machine translation
  – NMT advised by SMT
- Natural language dialogue
  – CopyNet

Question Answering
- Neural Enquirer

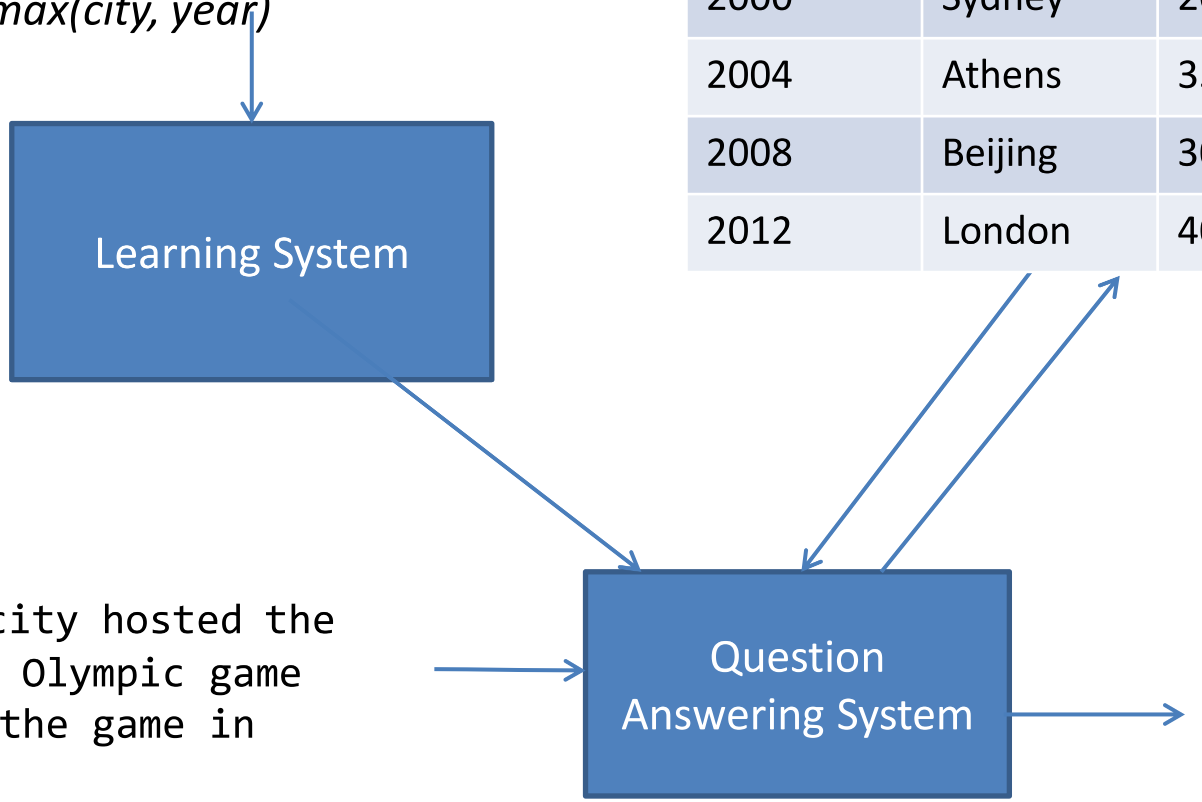
Question Answering from Relational Database

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



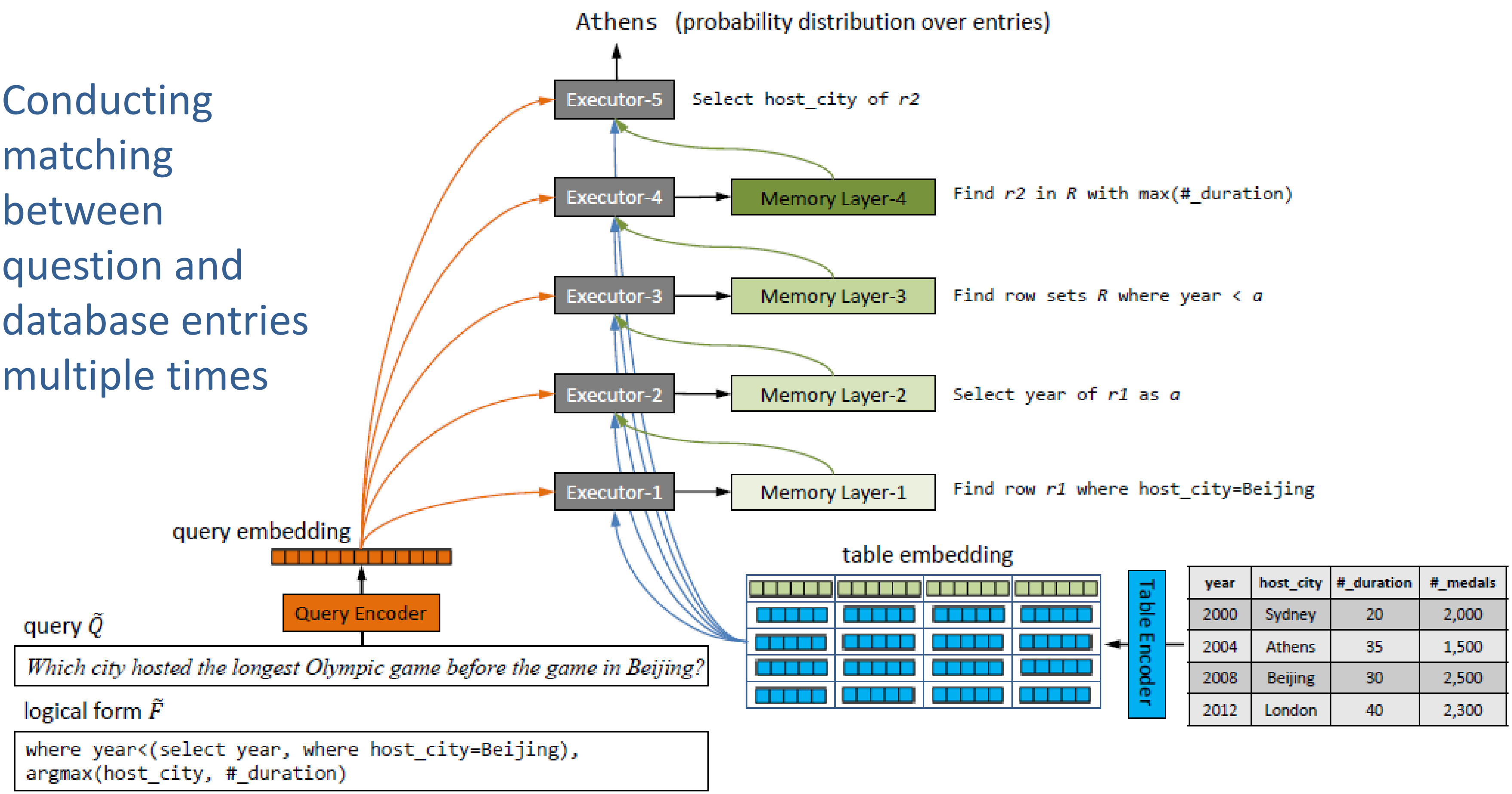
Q: Which city hosted the longest Olympic game before the game in Beijing?

A: Athens

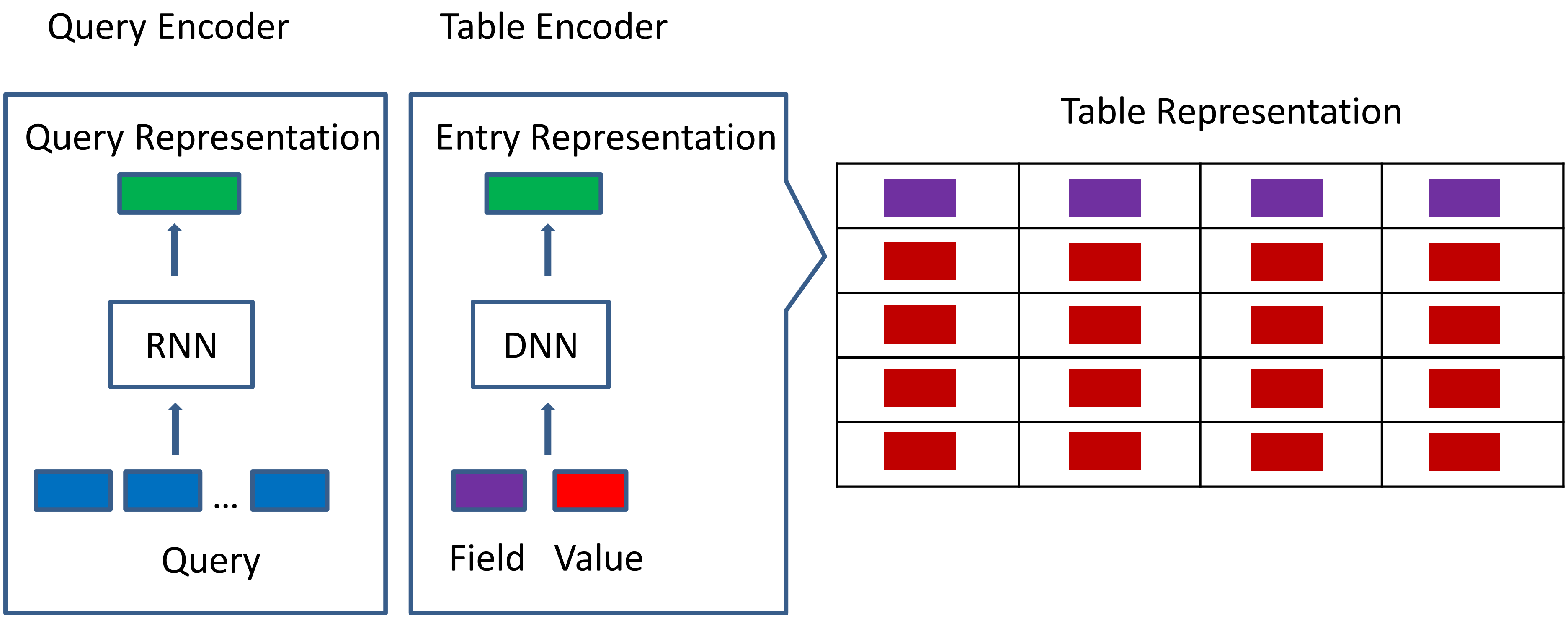
Neural Enquirer

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

Conducting matching between question and database entries multiple times

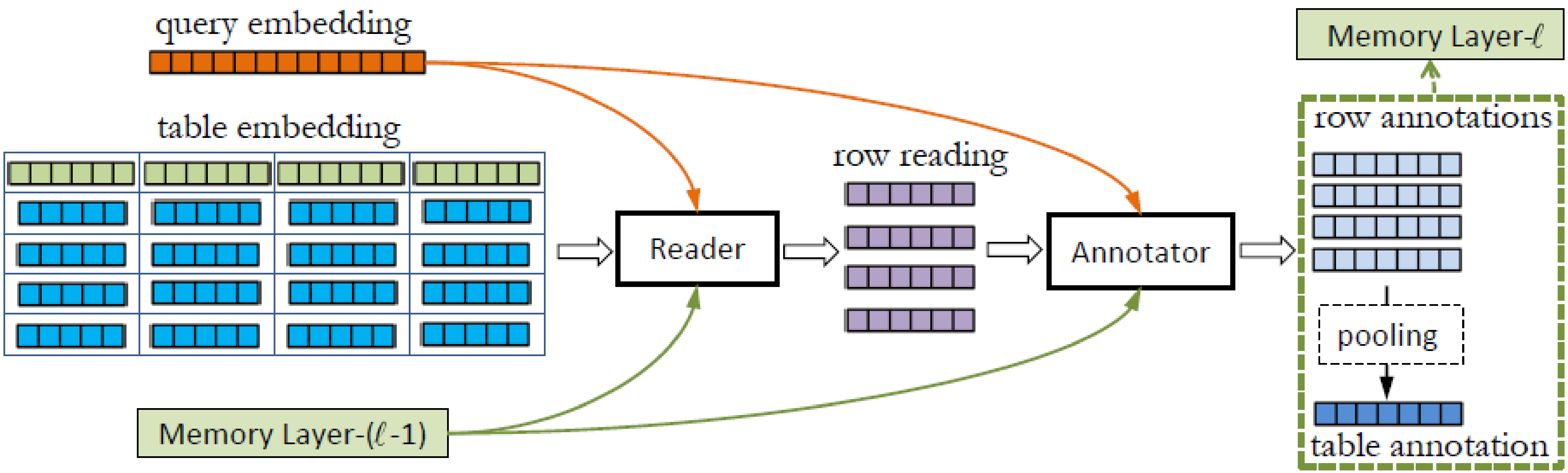


Query Encoder and Table Encoder



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

Executors



- 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

- 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

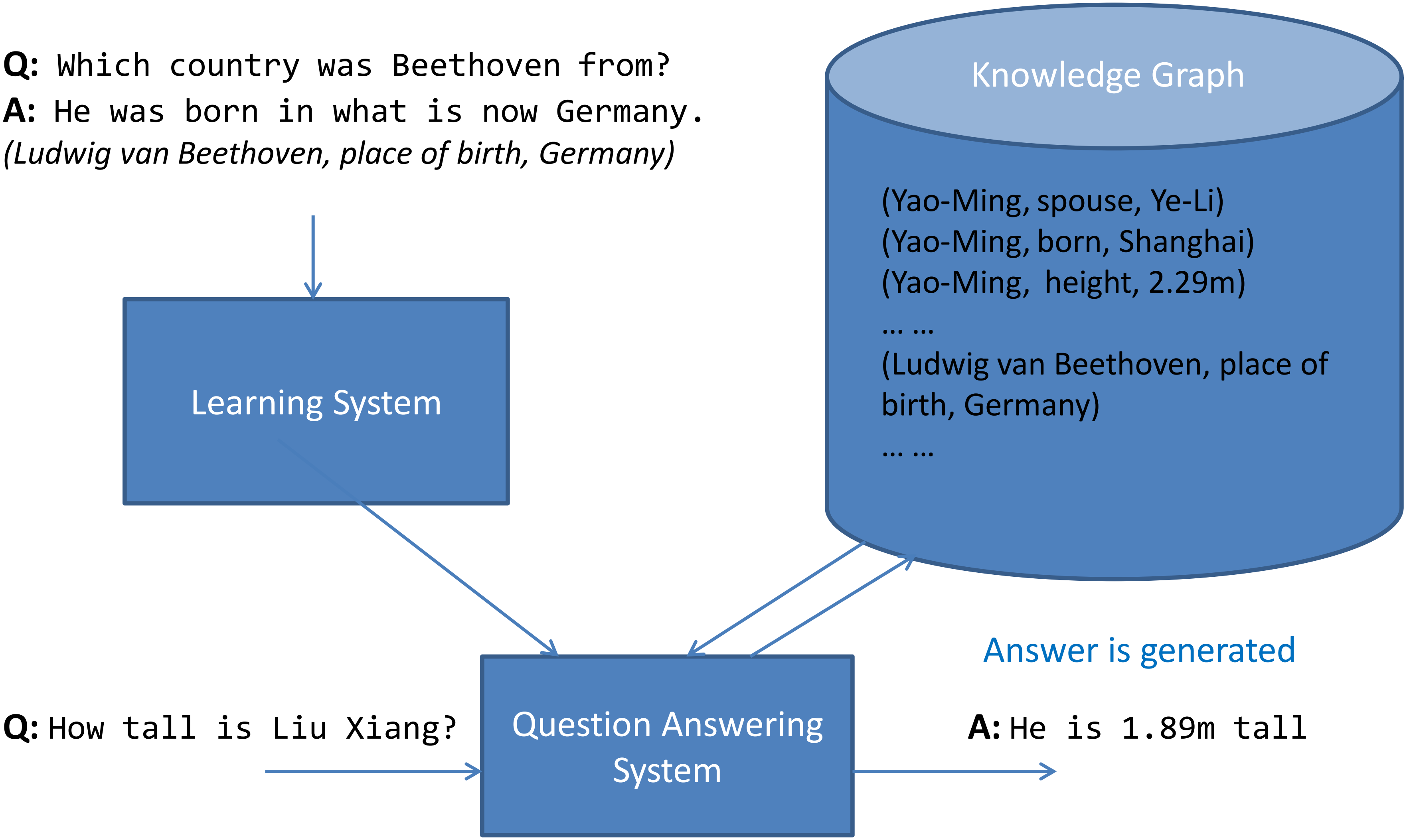
25K Data			100K Data		
Semantic Parser	End-to-End	Step-by-Step	Semantic Parser	End-to-End	Step-by-Step
65.2%	84.0%	96.4%	NA	90.6%	99.9%

Question Answering
- GenQA

Question Answering from Knowledge Graph

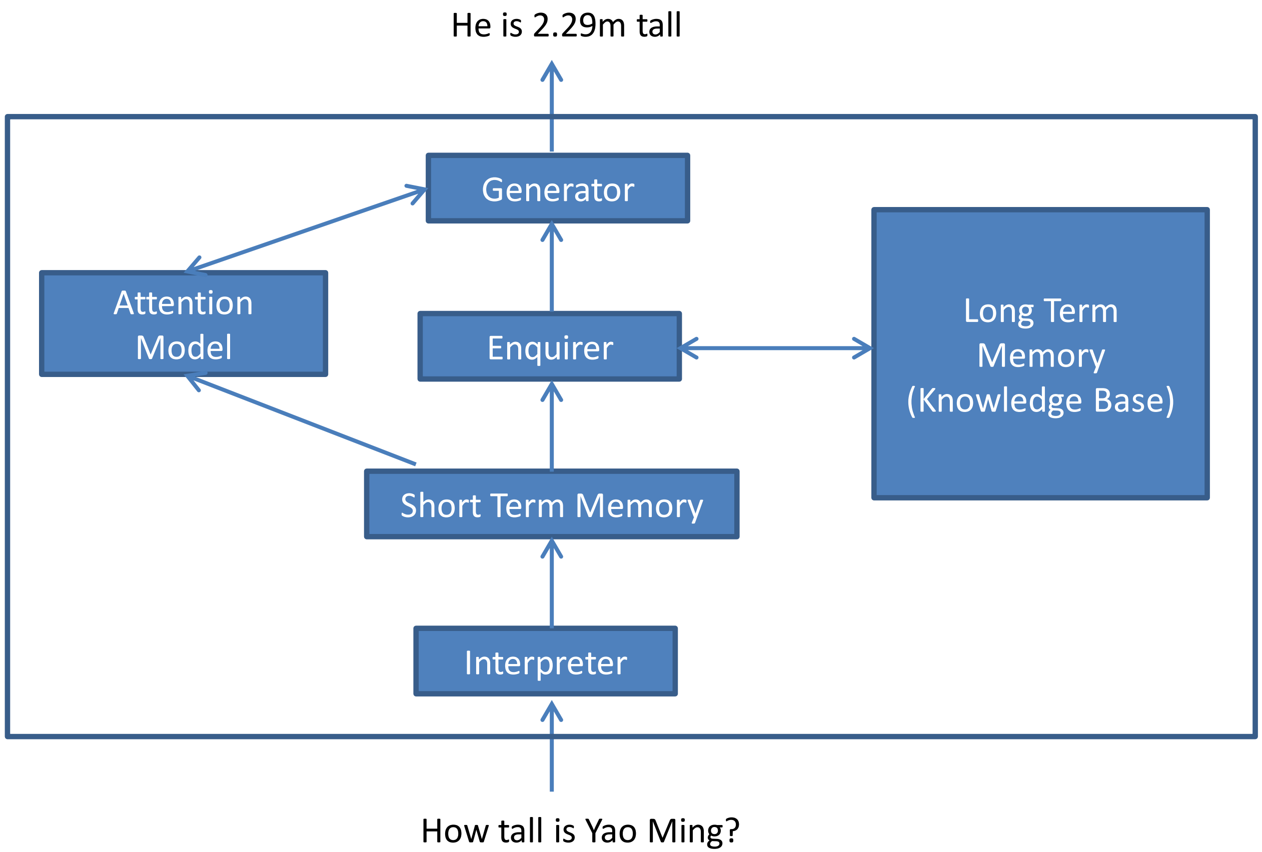
Q: `How tall is Yao Ming?`
A: `He is 2.29m tall and is visible from space.`
(Yao Ming, height, 2.29m)

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



GenQA

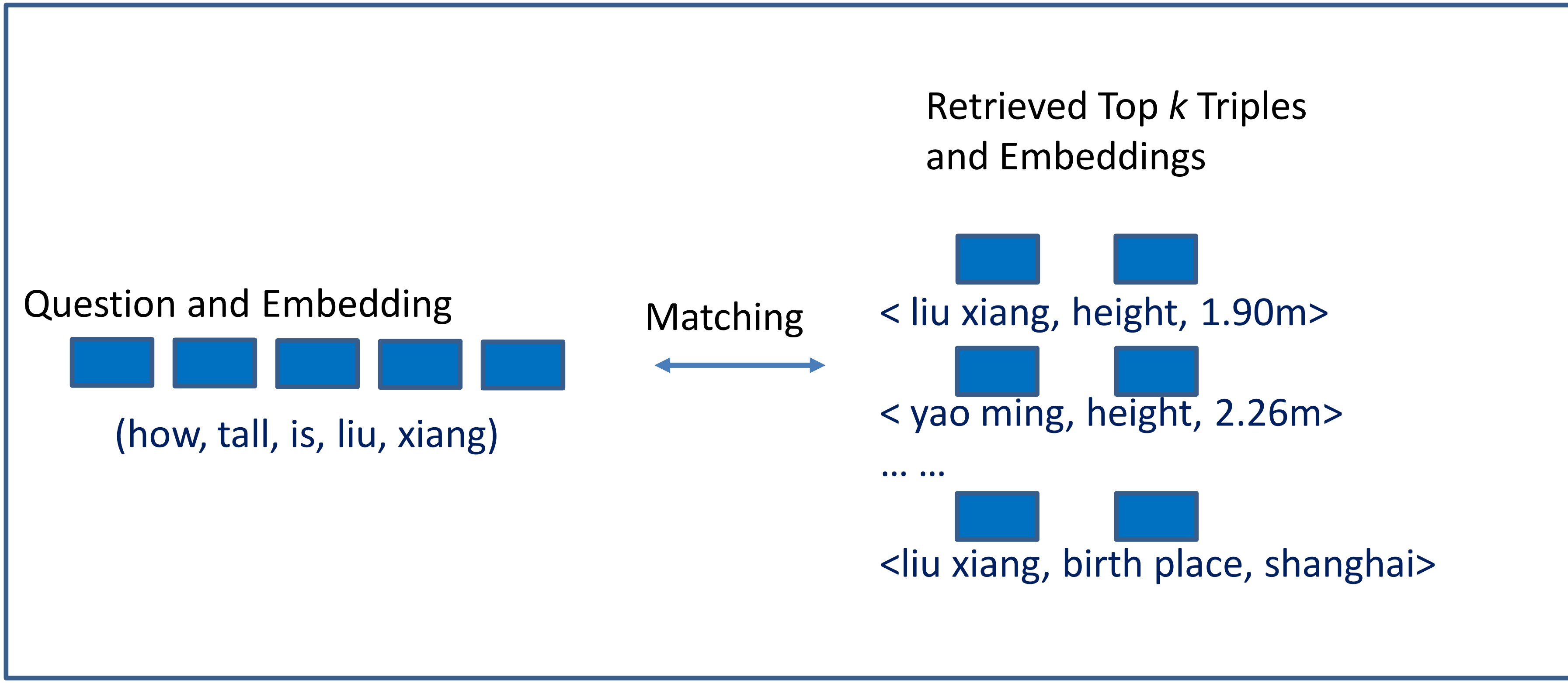
- **Interpreter:** creates representation of question using RNN
- **Enquirer:** retrieves top k triples with highest matching scores using CNN model
- **Generator:** generates answer based on question and retrieved triples using attention-based RNN
- **Attention model:** controls generation of answer



Key idea:

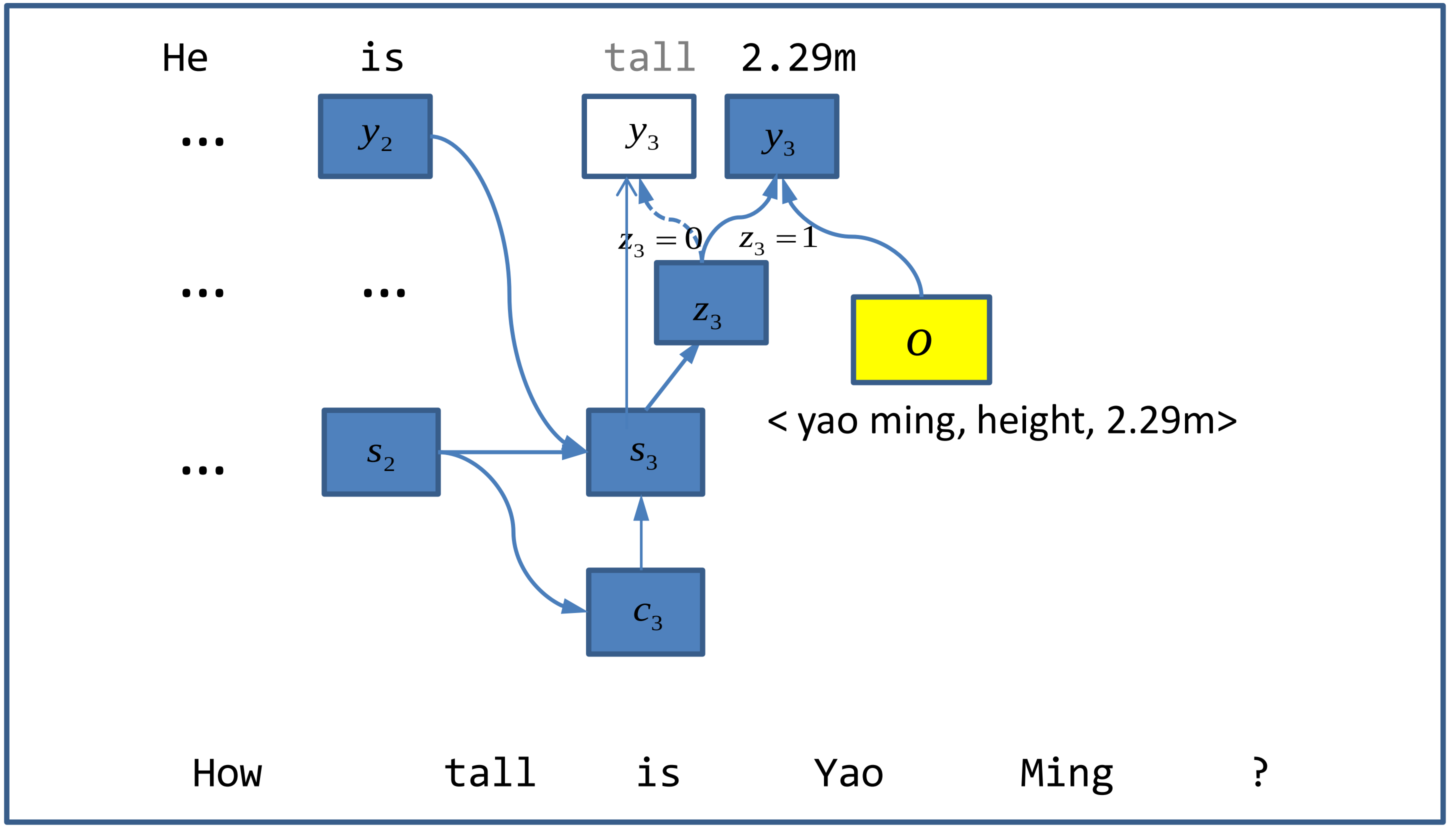
- Generation of answer based on question and retrieved result
- Combination of neural processing and symbolic processing

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 variable decides whether to generate a word or use the object of top triple

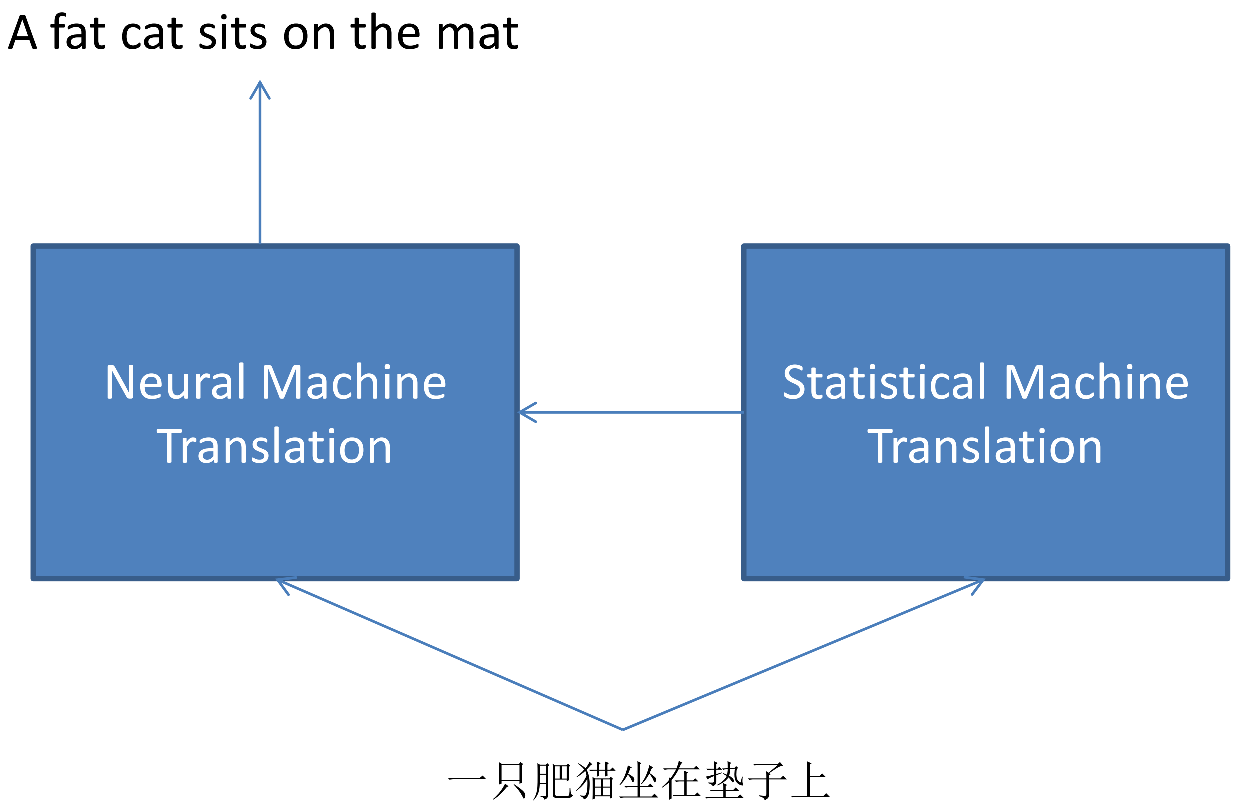
Experimental Results

- Experiment
  - Trained with 720K question-answer pairs (Chinese) associated with 1.1M triples in knowledge-base, *data is noisy*
  - Accuracy = 52%
  - Data is still noisy

Question	Answer	
Who wrote the Romance of the Three Kingdoms?	Luo Guanzhong in Ming dynasty	correct
How old is Stefanie Sun this year?	Thirty-two, he was born on July 23, 1978	wrong
When will Shrek Forever After be released?	Release date: Dreamworks Pictures	wrong

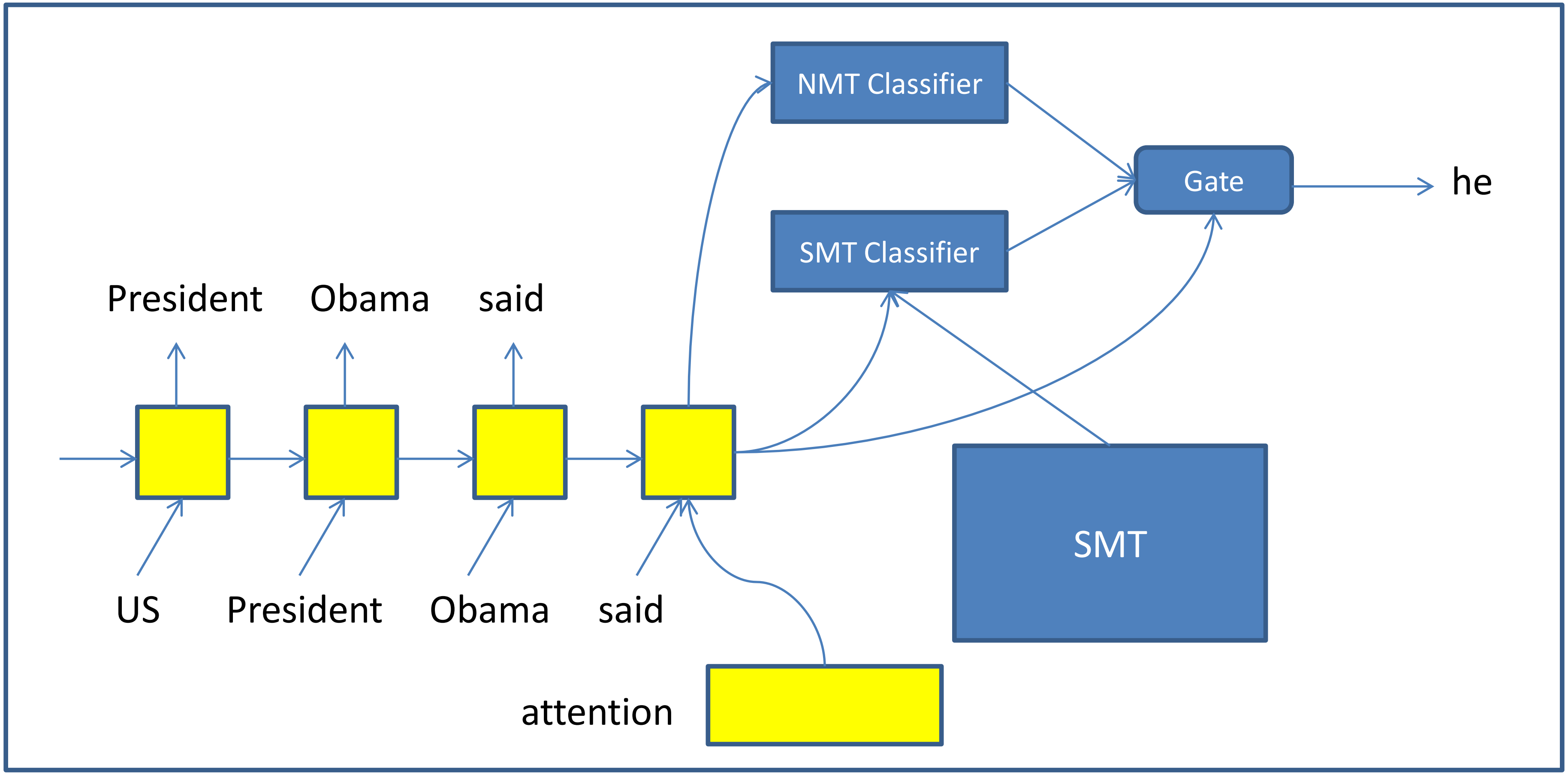
NMT advised by SMT

Neural Machine Translation advised by Statistical Machine Translation



SMT complements NMT by providing translation recommendation;
To solve (1) improper translation problem (2) UNK problem

Architecture: NMT advised by SMT



NMT Classifier generates candidates, SMT Classifier generates candidates, and Gate linearly combines the two and generates final candidate

Characteristics: NMT advised by SMT

- Framework = NMT, component = SMT
- Linearly combines candidates from NMT and candidates from SMT → generate better translation
- Use most likely candidate from SMT to replace UNK word → deal with low frequency problem
- Training:
  - First train NMT only model to initialize parameters
  - Next train NMT-SMT model to learn all parameters

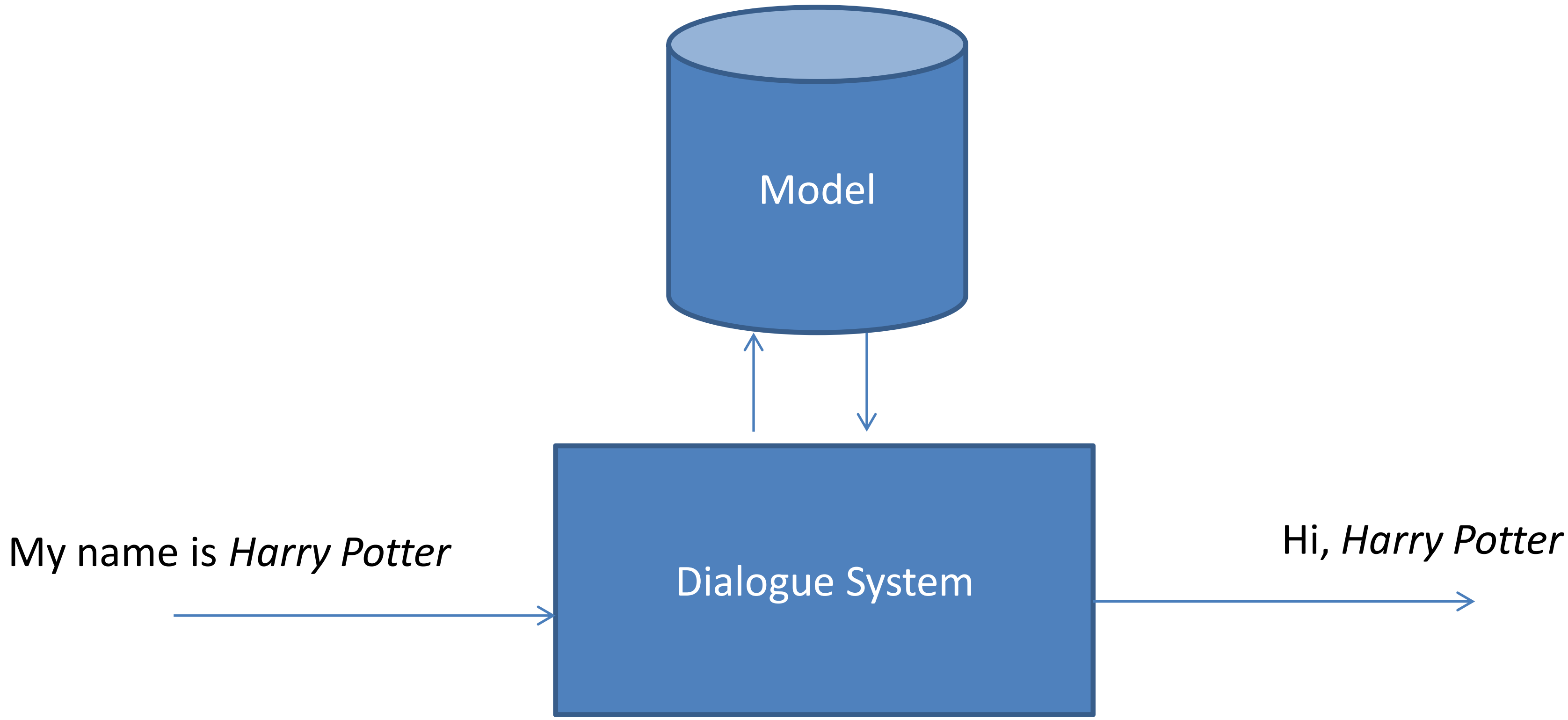
Experimental Results

- Experiment
  - Chinese to English translation
  - Trained with 1.25M sentence pairs
  - Tested with NIST datasets

System	Moses (SMT)	Groundhog (NMT)	RNN Search (NMT)	+SMT	+SMT +UNK
BLEU	31.73	30.99	32.50	33.70	34.94

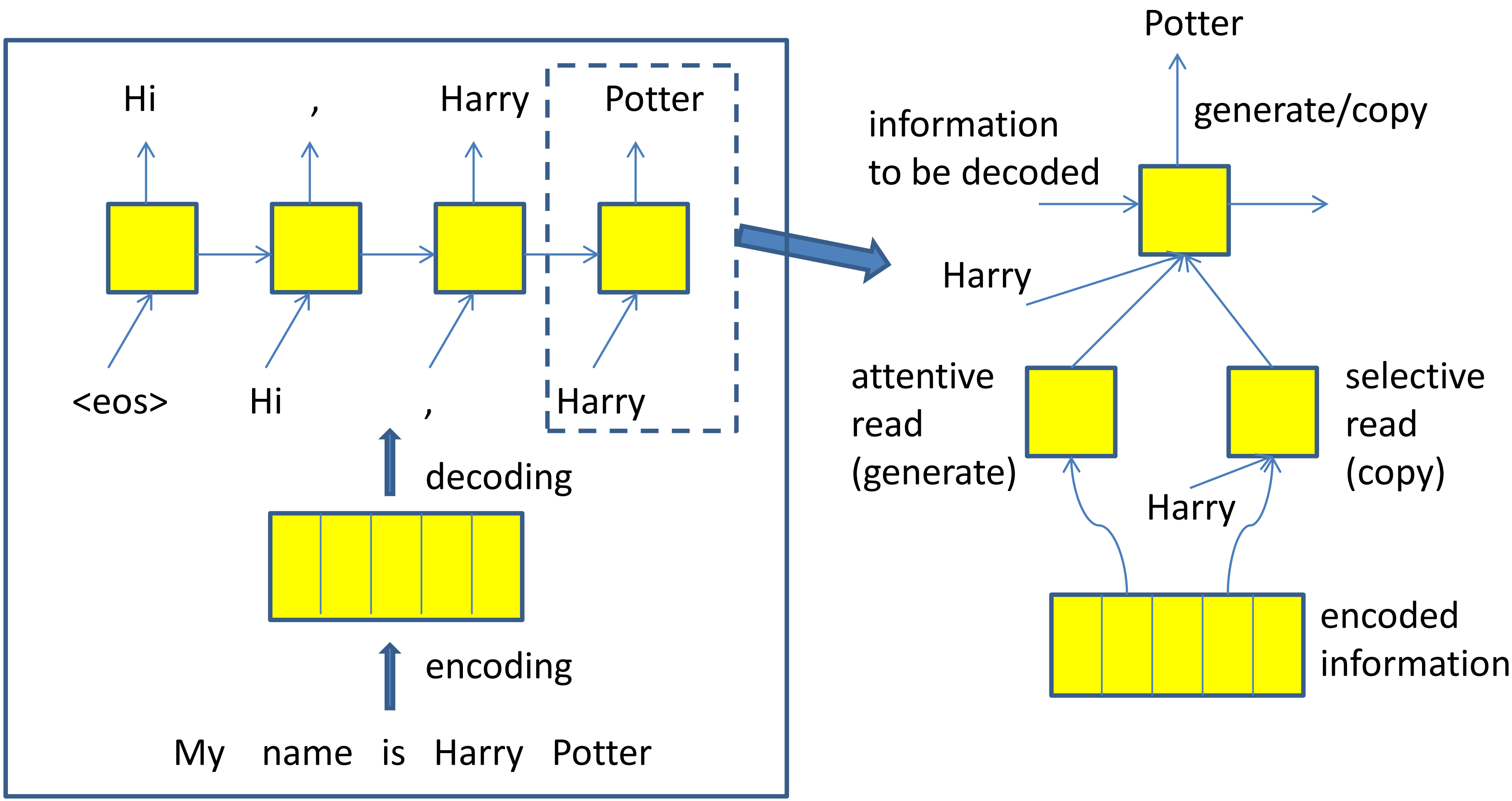
Natural Language Dialogue - CopyNet

Single Turn Dialogue with Generating and Copying Mechanism



Dialogue system can not only generate response, but also copy from given message

Architecture: CopyNet



CopyNet can either generate word based on attentive read, or copy word based on selective read

Characteristics: CopyNet

- Decoder can both generate and copy
- Mixture model of generating and copying
- Attentive read: find suitable word to influence generation of word in target sequence
- Selective read: find location of word to be copied from source sequence
- Model is fully differentiable
- Training: maximum likelihood of target sequence given source sequence

Experimental Results

- Experiment
  - Summarization of short text in Chinese
  - Trained with 2.4M text-summary pairs
  - Tested with 9.3K text-summary pairs

Model	ROUGE-1	ROUGE-2	ROUGE-L
RNN -C	29.9	17.4	27.2
RNN -W	26.8	16.1	24.1
CopyNet -C	**34.4**	**21.6**	**31.3**
CopyNet -W	**35.0**	**22.3**	**32.0**

Summary of Talk

- Deep Learning Made Big Breakthroughs in Natural Language Processing
- Until Present: Purely Neural Processing
- Present and Future: Hybrid Approach Combining Neural Processing and Symbolic Processing
- Noah's Ark Lab Developed State-of-Art Models for Question Answering, Machine Translation, Natural Language Dialogue, and Image Retrieval

References

1. Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, and Hang Li. Neural Machine Translation with Reconstruction. The 31st AAAI Conference on Artificial Intelligence (AAAI 2017).
2. Xing Wang, Zhengdong Lu, Zhaopeng Tu, Hang Li, Deyi Xiong, and Min Zhang. Neural Machine Translation Advised by Statistical Machine Translation. The 31st AAAI Conference on Artificial Intelligence (AAAI 2017).
3. Zhaopeng Tu, Yang Liu, Zhengdong Lu, Xiaohua Liu, and Hang Li. Context Gates for Neural Machine Translation. Transactions of the Association for Computational Linguistics (TACL 2017).
4. Fandong Meng, Zhengdong Lu, Hang Li, Qun Liu. Interactive Attention for Neural Machine Translation. Proceedings of the 26th International Conference on Computational Linguistics (COLING'16), 2016.
5. Mingxuan Wang, Zhengdong Lu, Hang Li, Qun Liu. Memory-enhanced Decoder for Neural Machine Translation. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP'16), 278-286, 2016.
6. Pengcheng Yin, Zhengdong Lu, Hang Li, Ben Kao, Neural Enquirer: Learning to Query Tables in Natural Language. IEEE Data Engineering Bulletin 39(3): 63-73, 2016.
7. Jiatao Gu, Zhengdong Lu, Hang Li, Victor O. K. Li. Incorporating Copying Mechanism in Sequence-to-Sequence Learning. Proceedings of the 54th Annual Meeting of Association for Computational Linguistics (ACL'16), 2016.
8. Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu and Hang Li. Modeling Coverage for Neural Machine Translation. Proceedings of the 54th Annual Meeting of Association for Computational Linguistics (ACL'16), 2016.
9. Jun Yin, Xin Jiang, Zhengdong Lu, Lifeng Shang, Hang Li, Xiaoming Li. Neural Generative Question Answering. Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16), 2972-2978, 2016.

References

10. Pengcheng Yin, Zhengdong Lu, Hang Li, Ben Kao. Neural Enquirer: Learning to Query Tables with Natural Language. Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI’16), 2308-2314, 2016.
11. Lin Ma, Zhengdong Lu, Hang Li. Learning to Answer Questions From Image Using Convolutional Neural Network. Proceedings of the Thirtieth AAAI Conference (AAAI’16), 2016.
12. Lin Ma, Zhengdong Lu, Lifeng Shang, Hang Li, Multimodal Convolutional Neural Networks for Matching Image and Sentence. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), 2623-2631, 2015.
13. Baotian Hu, Zhaopeng Tu, Zhengdong Lu, Hang Li, Qingcai Chen. Context-Dependent Translation Selection Using Convolutional Neural Network. Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP'15), 536-541, 2015.
14. Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP'15), 1577-1586, 2015.
15. Mingxuan Wang, Zhengdong Lu, Hang Li, Wenbin Jiang, Qun Liu. GenCNN: A Convolutional Architecture for Word Sequence Prediction. Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP'15), 1567-1576, 2015.
16. Fandong Meng, Zhengdong Lu, Mingxuan Wang, Hang Li, Wenbin Jiang, Qun Liu. Encoding Source Language with Convolutional Neural Network for Machine Translation. Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP'15), 20-30, 2015.

References

17. Mingxuan Wang, Zhengdong Lu, Hang Li, Qun Liu. Syntax-based Deep Matching of Short Texts. Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15), 1354-1361, 2015.
18. Baotian Hu, Zhengdong Lu, Hang Li, Qingcai Chen. Convolutional Neural Network Architectures for Matching Natural Language Sentences. Proceedings of Advances in Neural Information Processing Systems 27 (NIPS'14), 2042-2050, 2014.
19. Zhengdong Lu, Hang Li. A Deep Architecture for Matching Short Texts. Proceedings of Neural Information Processing Systems 26 (NIPS'13), 1367-1375, 2013.

Thank you!