Progress of DNN-Based Natural Language Processing(NLP)

Dr. Ming Zhou

(mingzhou@microsoft.com)

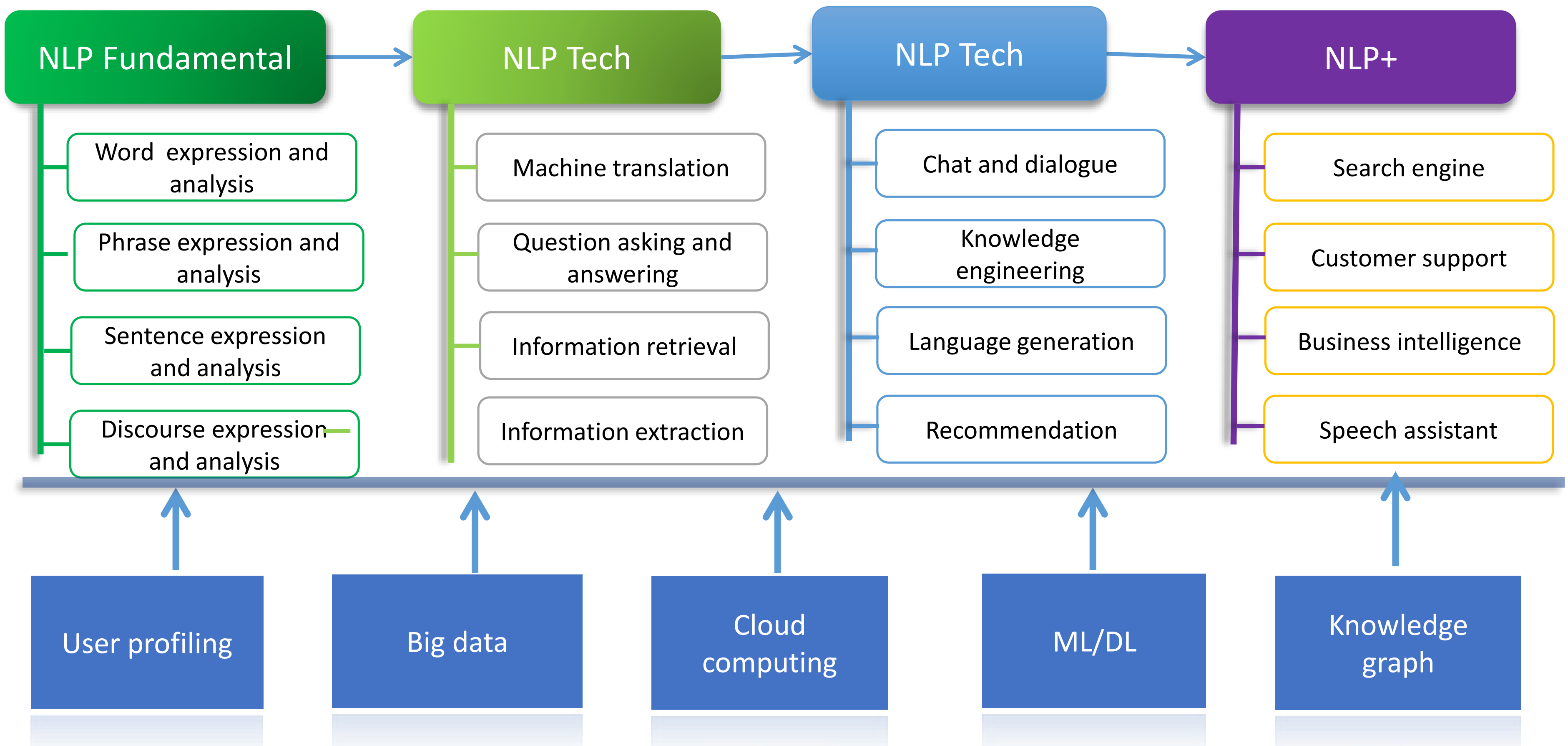
Microsoft Research Asia

GAITC NLP Forum

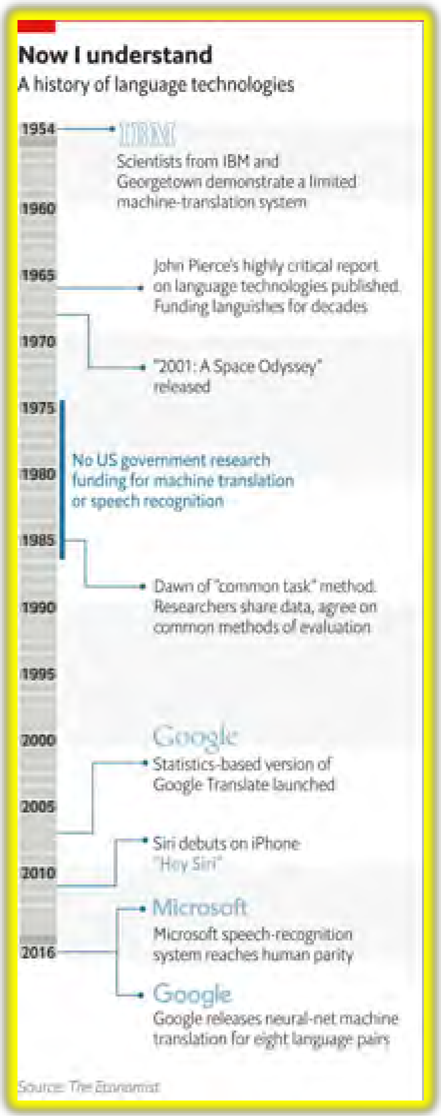
May 22, 2017 @ Beijing

Natural Language Processing（NLP）

NLP is a branch of AI, referring to the tech to analyze, understand and generate human language to facilitate human-computer interaction and human-human exchange.



NLP Evolution



- 1940 ~ 1954： Invention of computer and intelligence theory
- Leader：Chomsky， Backus， Weaver， Shannon

- 1954 ~ 1970：Formal rule system, logic theory and perceptron
- Leader：Minsky， Rosenblatt

- 1970 ~ 1980：HMM-based ASR, semantic and discourse modelling
- Leader：Frederick Jelinek， Martin Kay

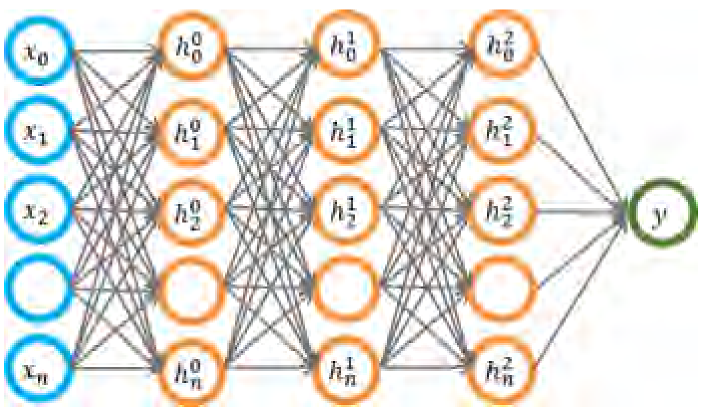
- 1980 ~ 1991：Rule base and knowledge base
- Work：WordNet (1985), HPSG (1987), CYC (1984)

- 1991 ~ 2008：statistical machine learning
- Approach：SVM, MaxEnt, PCFG, PageRank
- Application：SMT, QA and search engine

- 2008 ~ 2017：big data and DL
- Work：word embedding, NMT, chit-chat, dialogue system, reading comprehension

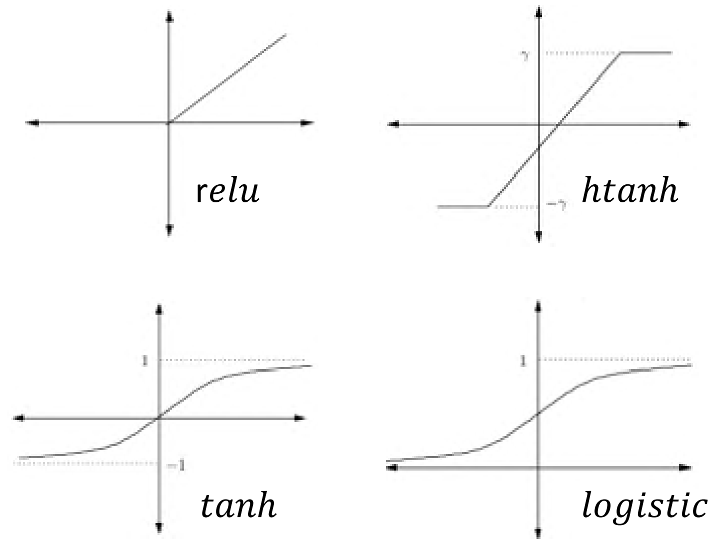
Deep Neural Network

- **D**eep **N**eural **N**etwork :
  - Involve multiple level neural networks
  - Non-Linear Learner



$h^0 = f(w^0 x)$ $\quad$ $h^1 = f(w^1 h^0)$

$h^2 = f(w^2 h^1)$ $\quad$ $y = f(w^3 h^2)$



Active functions: $y = f(x)$

DNN4NLP Progress

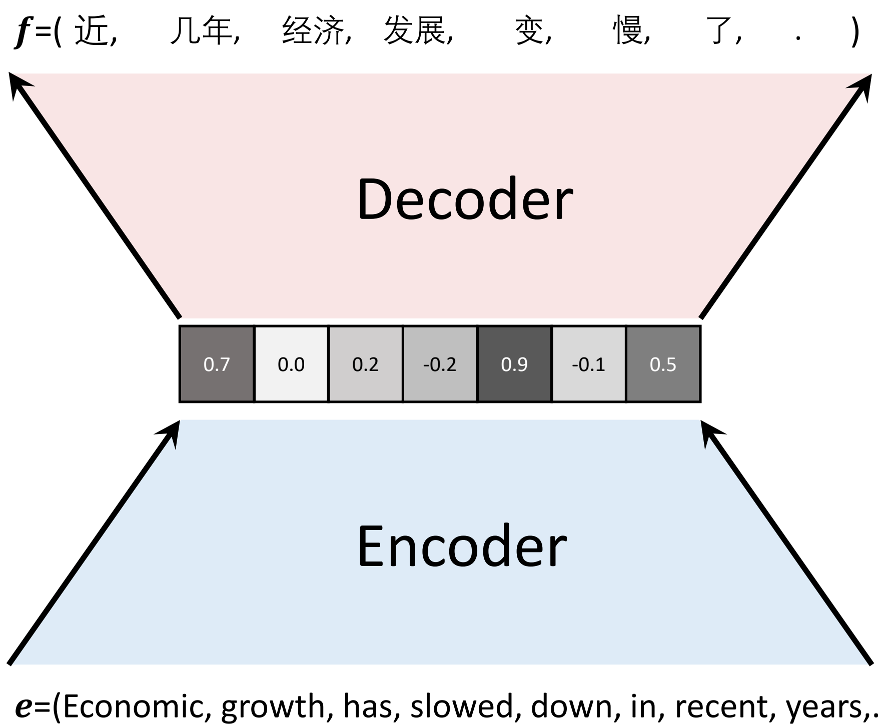
- Progress
  - Word expression with embedding
  - Sentence modelling via CNN or RNN(LSTM/GRU) for similarity estimation and sequential mapping
  - Successful applications such as NMT, chatbot, etc.
- Still exploring
  - Learning from unlabeled data (GANs, Dual Learning)
  - Learning from knowledge
  - Learning from user/environment (RL)
  - Discourse and context modelling
  - Personalized system (via user profiling)

In this talk, I will focus

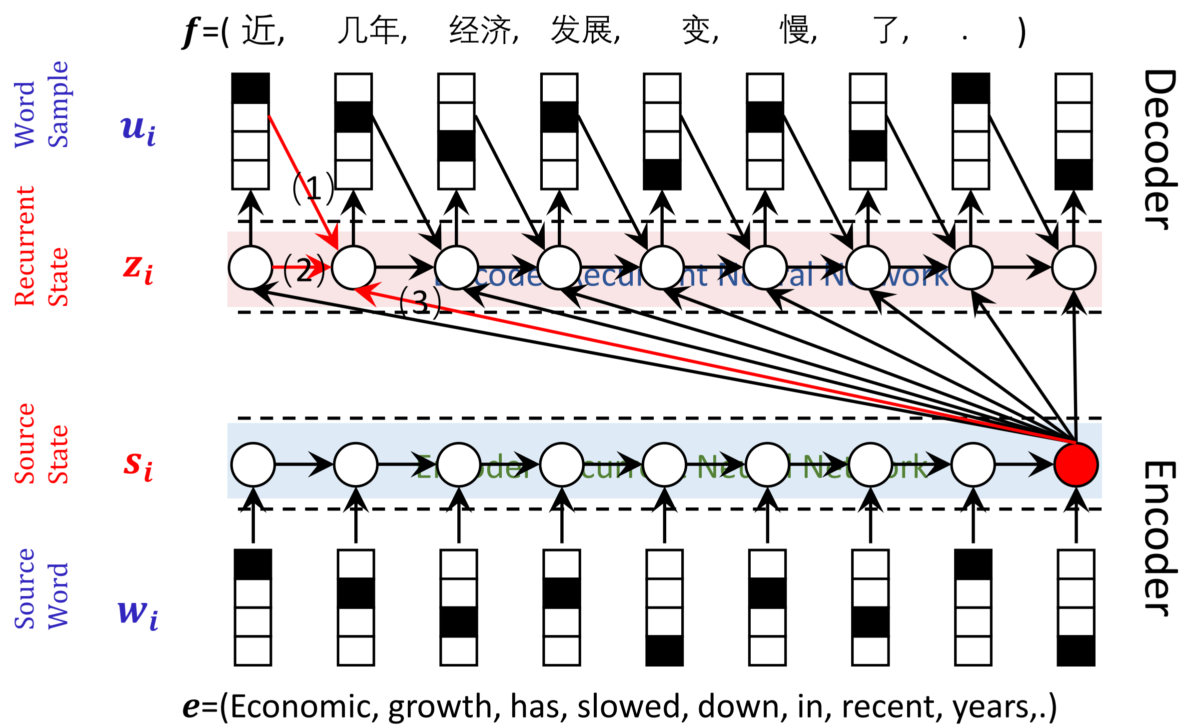
- NMT
- Chatbot
- Reading Comprehension

NMT

Encoder-Decoder for NMT with RNN

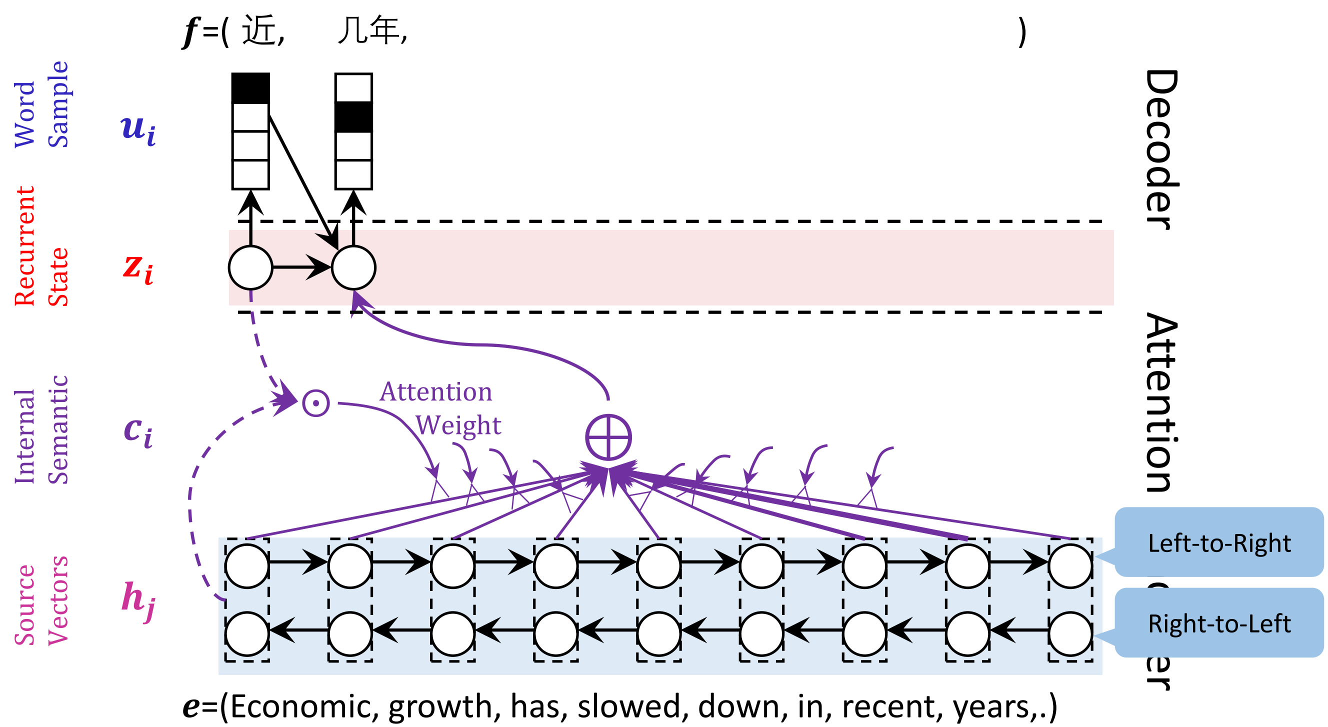


Encoder-Decoder for NMT



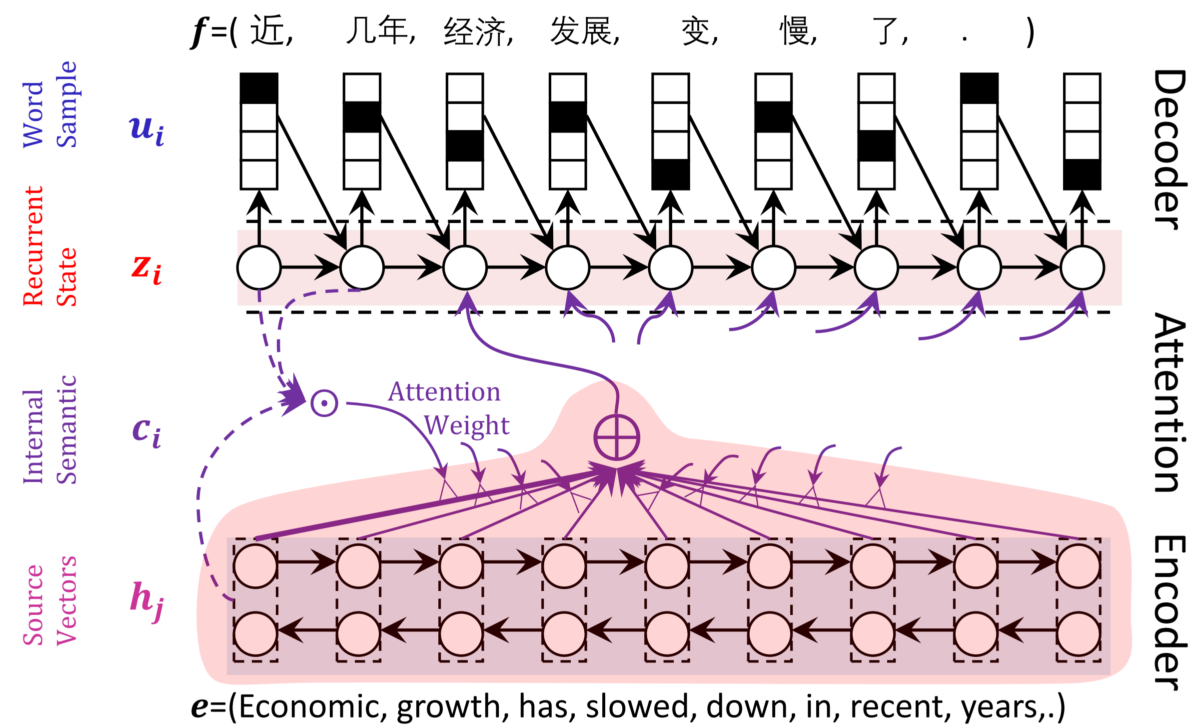
Sutskever et al., NIPS, 2014

Attention based Encoder-Decoder



Bahdanau et al., ICLR, 2015

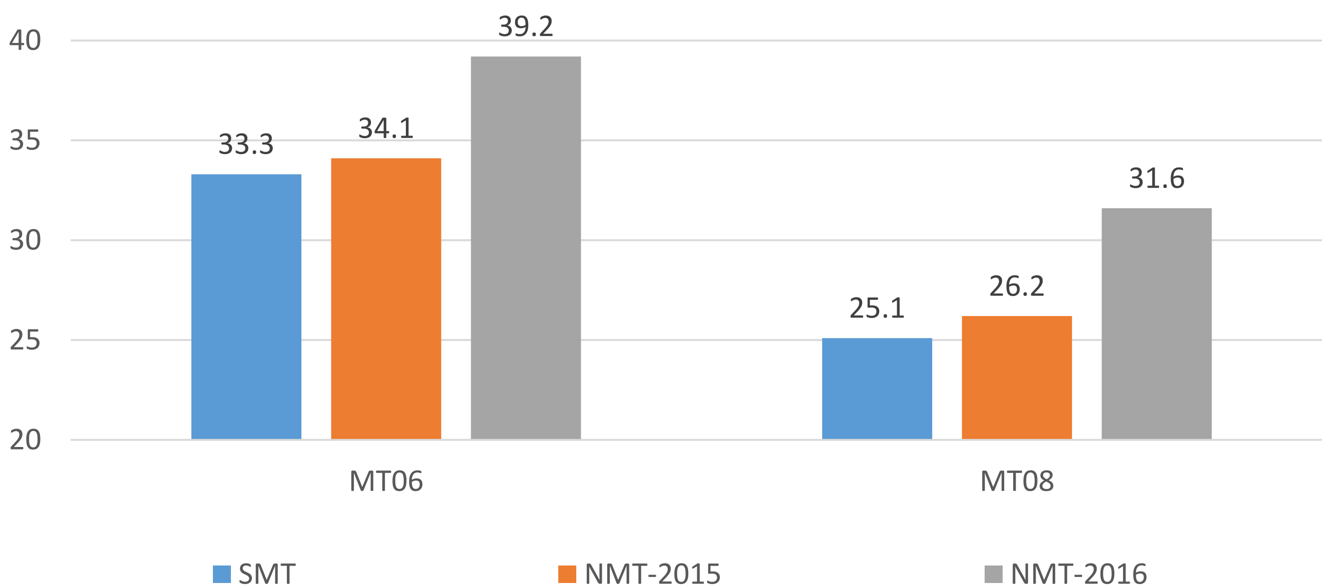
Attention based Encoder-Decoder



Bahdanau et al., ICLR, 2015

NMT Progress

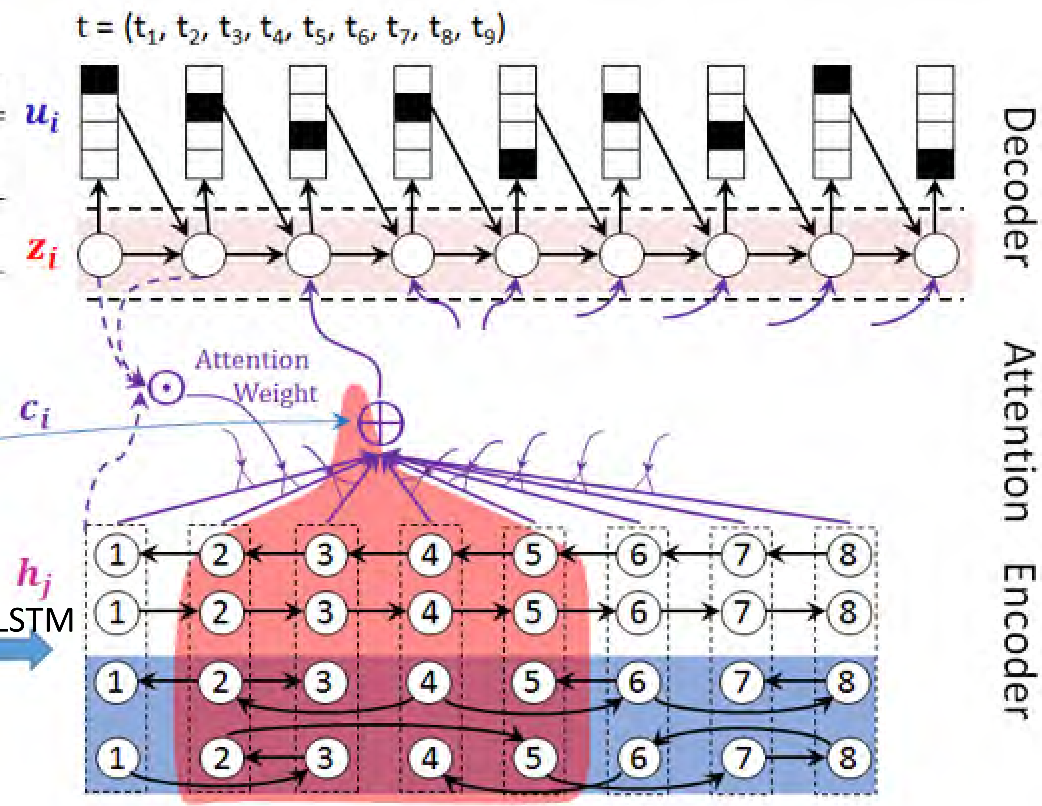
- 4+ BLEU points improvements over SMT
  - Main stream research (dominating papers in ACL)
  - Productization (MS, Baidu, Google)

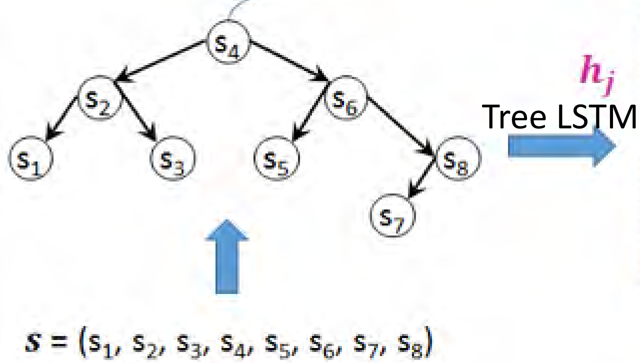


Fusing with Linguistic Knowledge

- Decoding from structure to string (tree-to-sequence NMT)

	K	RIBES	BLEU
Proposed model	500	71.8	20.0
Proposed model	2000	72.6	20.5
ANMT (Luong et al., 2015a)	500	70.7	18.5
ANMT (Luong et al., 2015a)	2000	71.5	19.4

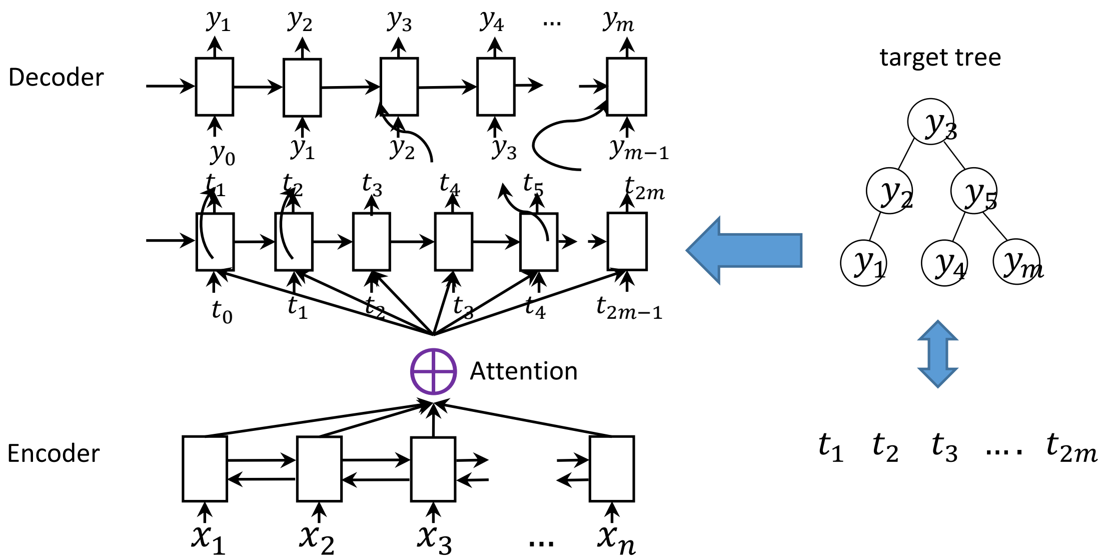




Akiko Eriguchi , et al, *Tree-to-Sequence Attentional Neural Machine Translation, ACL 2016*

Fusing with Linguistic Knowledge

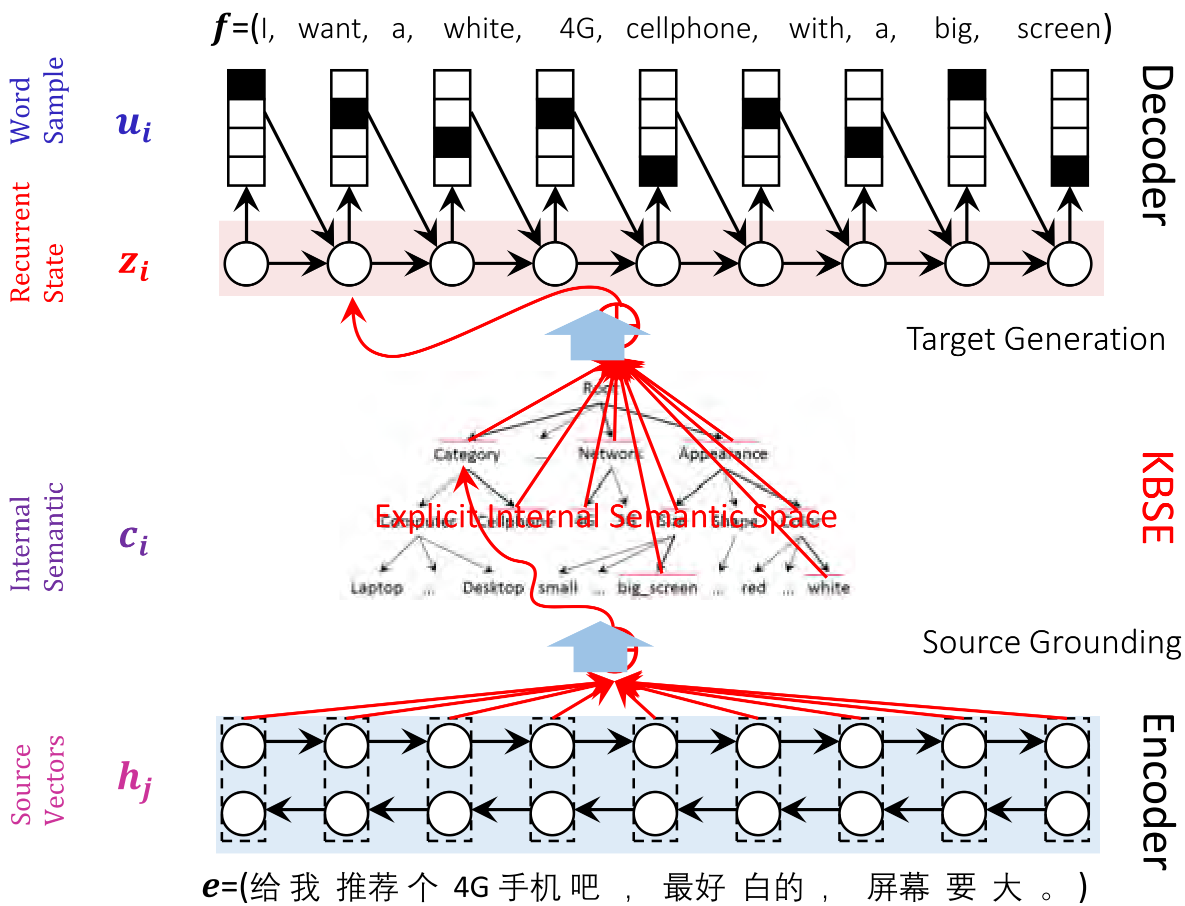
- Decoding from string to structure (sequence-to-tree NMT)



Settings	NIST 2005	NIST 2006	NIST 2008	NIST 2012	Average
HPSMT	35.34	33.56	26.06	27.47	30.61
RNNsearch	38.07	38.95	31.61	28.95	34.39
SD-NMT\K	38.83	39.23	31.92	29.72	34.93
SD-NMT	**39.38**	**41.81**	**33.06**	**31.43**	36.42

Shuangzhi Wu , et al, *Tree-to-Sequence Attentional Neural Machine Translation, ACL 2017*

Fusing with Domain Knowledge



Chen Shi, et al, Knowledge-based semantic embedding for machine translation, ACL 2016

Remaining Challenges

- Use of monolingual data
- OOV
- Linguistic rules at phrase and sentence levels
- Discourse level translation

Chatbot

- IR-based
- Generation-based

An Example of Conversation

Clerk: Good morning! Sir.

Me: Good morning!

Clerk: How are you today?

Me: Good. It is a good weather today.

Clerk: Yes. What I can help you?

Me: I want to buy instant noodles.

Clerk: What brand?

Me: Kangshifu （康师傅）

Clerk: how many boxes do you want?

Me: 3, please.

Clerk: I see.

Me: how much is the price?

Clerk: 3 Yuans

Me: Thanks.

Clerk: How do you pay? Cash or wechat?

Me: Wechat

Clerk: please scan here

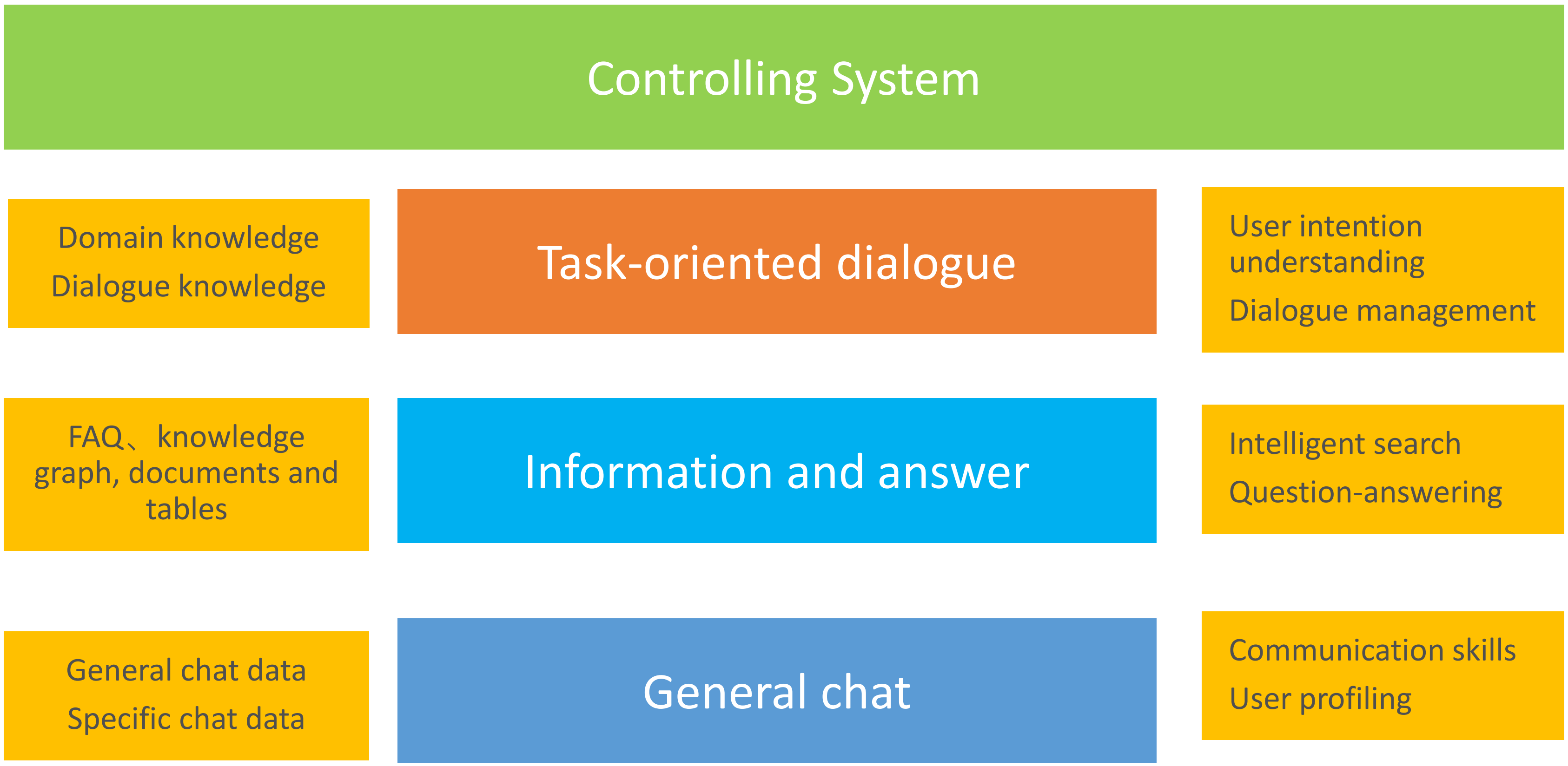
Me: Thanks

Chatbot (chit-chat)

Information & Answer

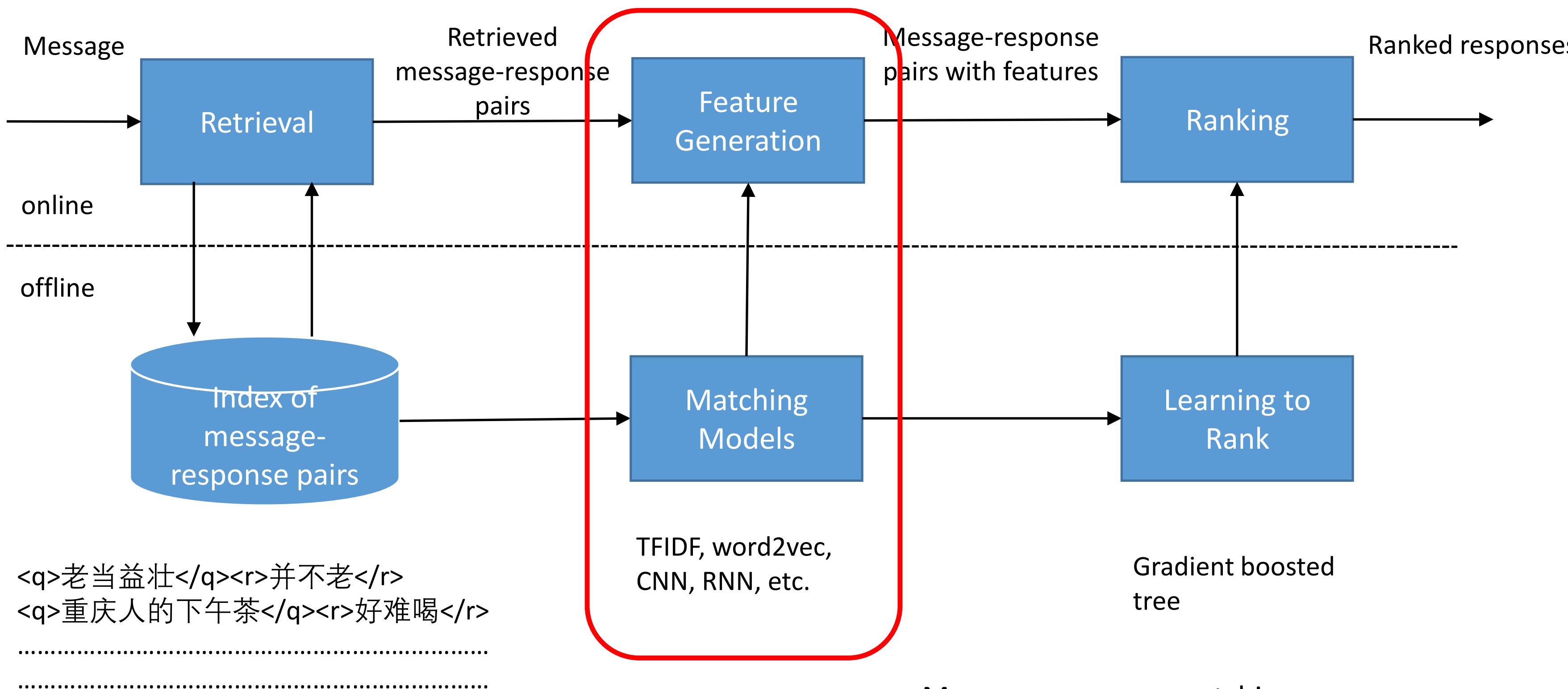
Task-oriented dialogue

Architecture of Conversation System



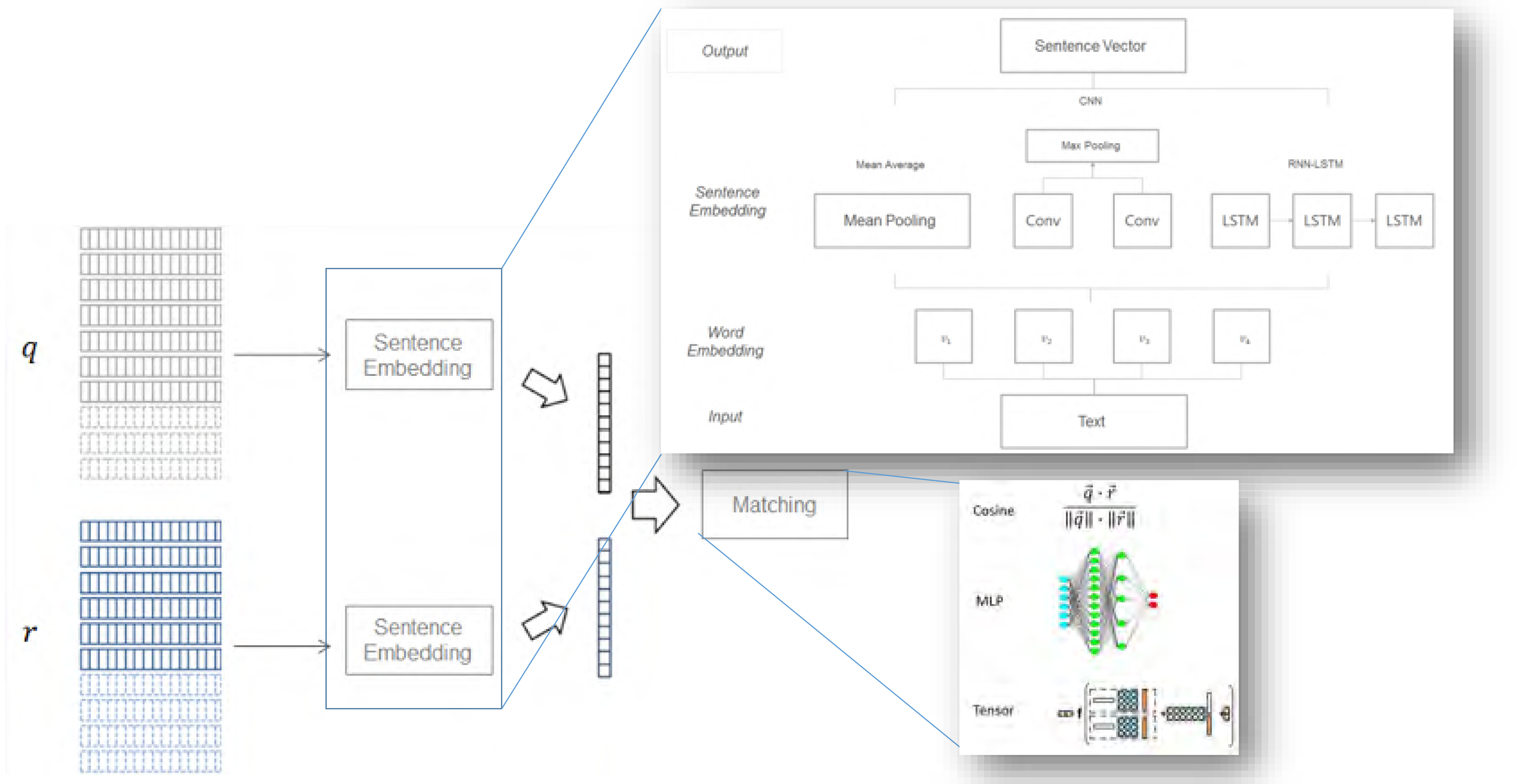
Retrieval-based Model

Architecture of Retrieval-based Chatbot (Single-Turn)



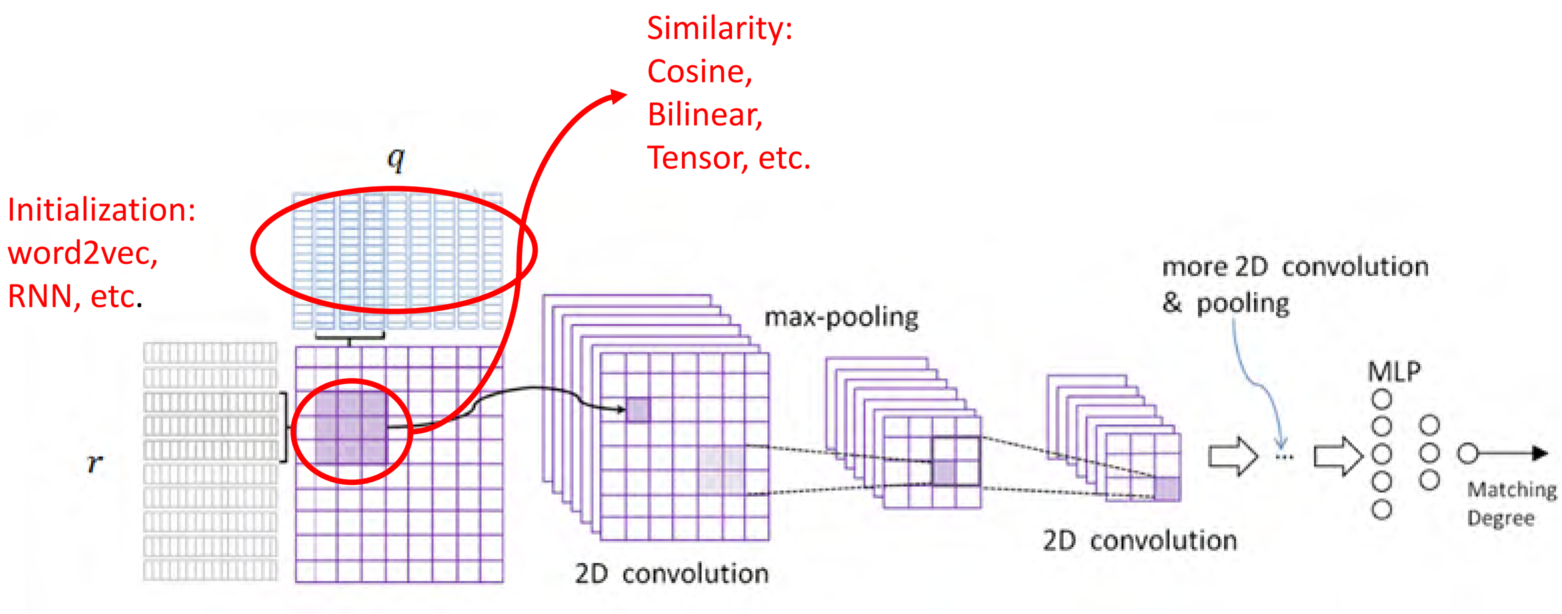
Message-response matching

Basic Models for Message-Response Matching : Architecture I



Baotian Hu et al. *Convolutional Neural Network Architectures for Matching Natural Language Sentences,* In NIPS'14
X Qiu and X Huang. *Convolutional Neural Tensor Network Architecture for Community-based Question Answering,* In IJCAI'15

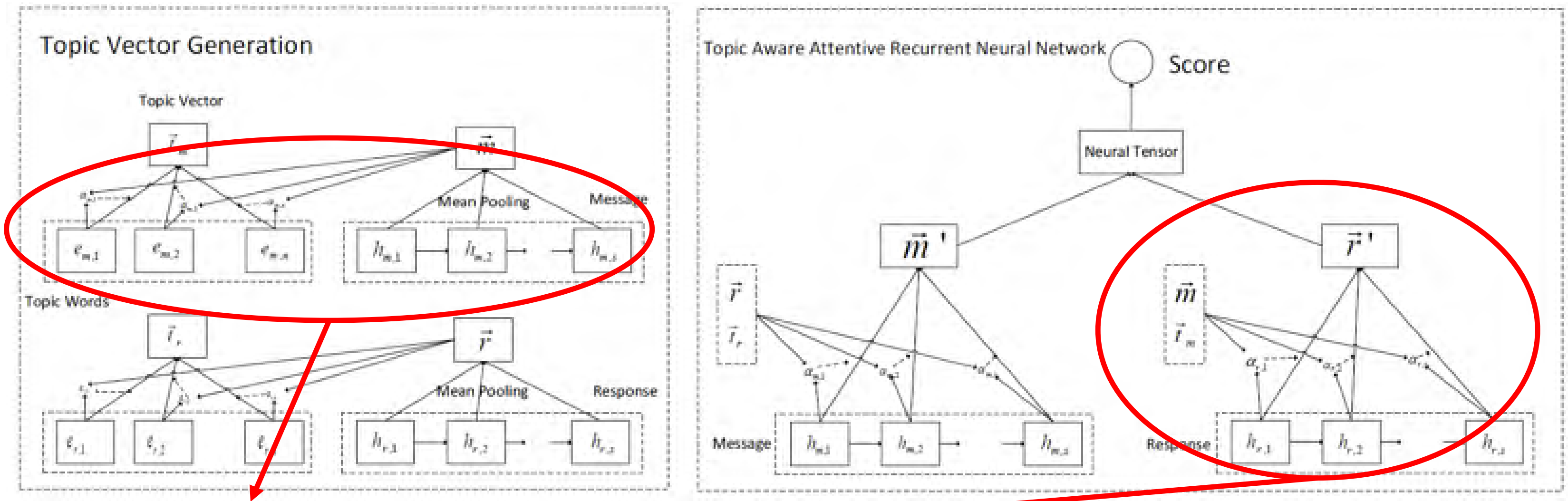
Basic Models for Message-Response Matching : Architecture II



Baotian Hu et al. *Convolutional Neural Network Architectures for Matching Natural Language Sentences,* In NIPS'14
Liang Pang et al. *Text Matching as Image Recognition*, In AAAI'16
Shengxian Wan et al. *A Deep Architecture for Semantic Matching with Multiple Positional Sentence Representations*, In AAAI'16

Fusing with External Knowledge I

- Topic Aware Attentive Recurrent Neural Network (TAARNN)



Let message/response attend to important parts in external knowledge (topics)

$$\vec{t}_m = \vec{\alpha}_m \cdot T_m$$
$$\vec{\alpha}_m \propto T_m \cdot A \cdot \vec{m}$$

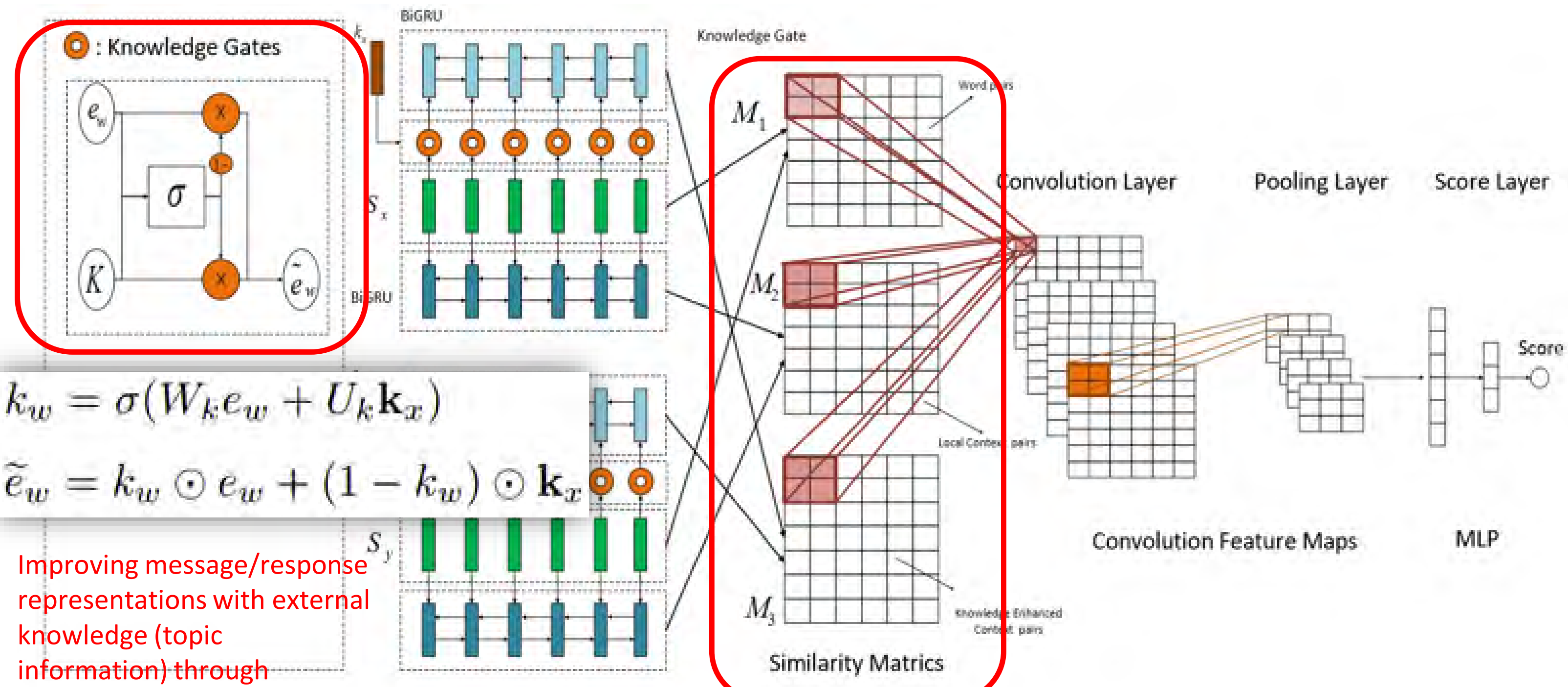
Joint attention with message/response and knowledge (topics)

$$\vec{r'} = \vec{\alpha}_r \cdot \vec{r}$$
$$\alpha_{r,i} \propto \tanh(\sum_j h_{r,i} \cdot A_2 \cdot h_{m,j} + h_{r,i} \cdot A_3 \cdot \vec{t}_m)$$

Yu Wu et al., *Response Selection with Topic Clues for Retrieval-based Chatbots*, In arxiv

Fusing with External Knowledge II

- Knowledge Enhanced Hybrid Neural Network (KEHNN)



$$k_w = \sigma(W_k e_w + U_k \mathbf{k}_x)$$

$$\tilde{e}_w = k_w \odot e_w + (1 - k_w) \odot \mathbf{k}_x$$

Improving message/response representations with external knowledge (topic information) through knowledge gates

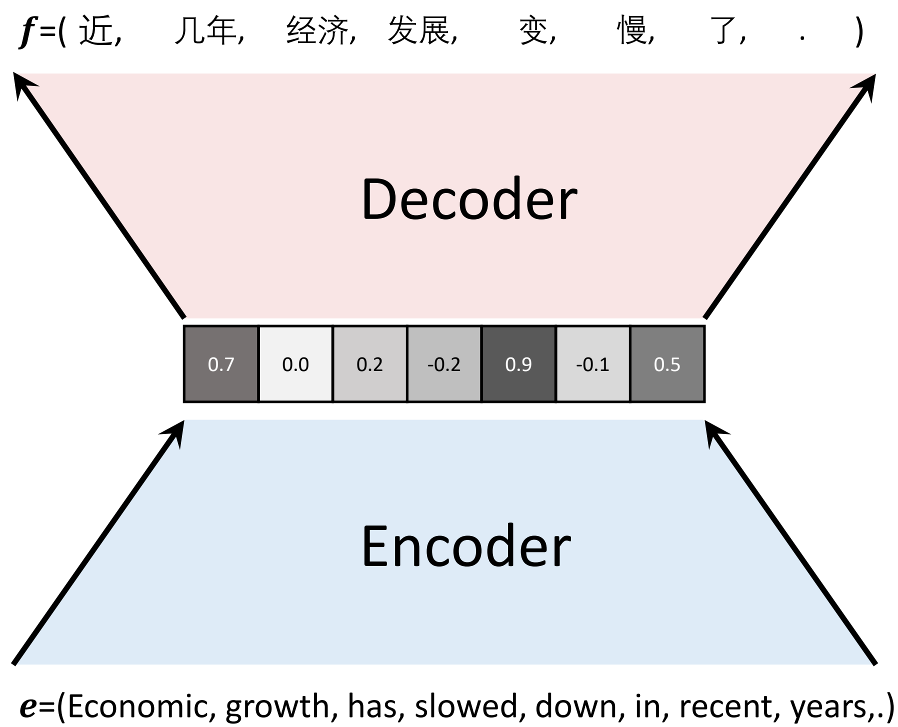
Matching with Multiple Channels

Yu Wu et al. *Knowledge Enhanced Hybrid Neural Network for Text Matching,* In arxiv

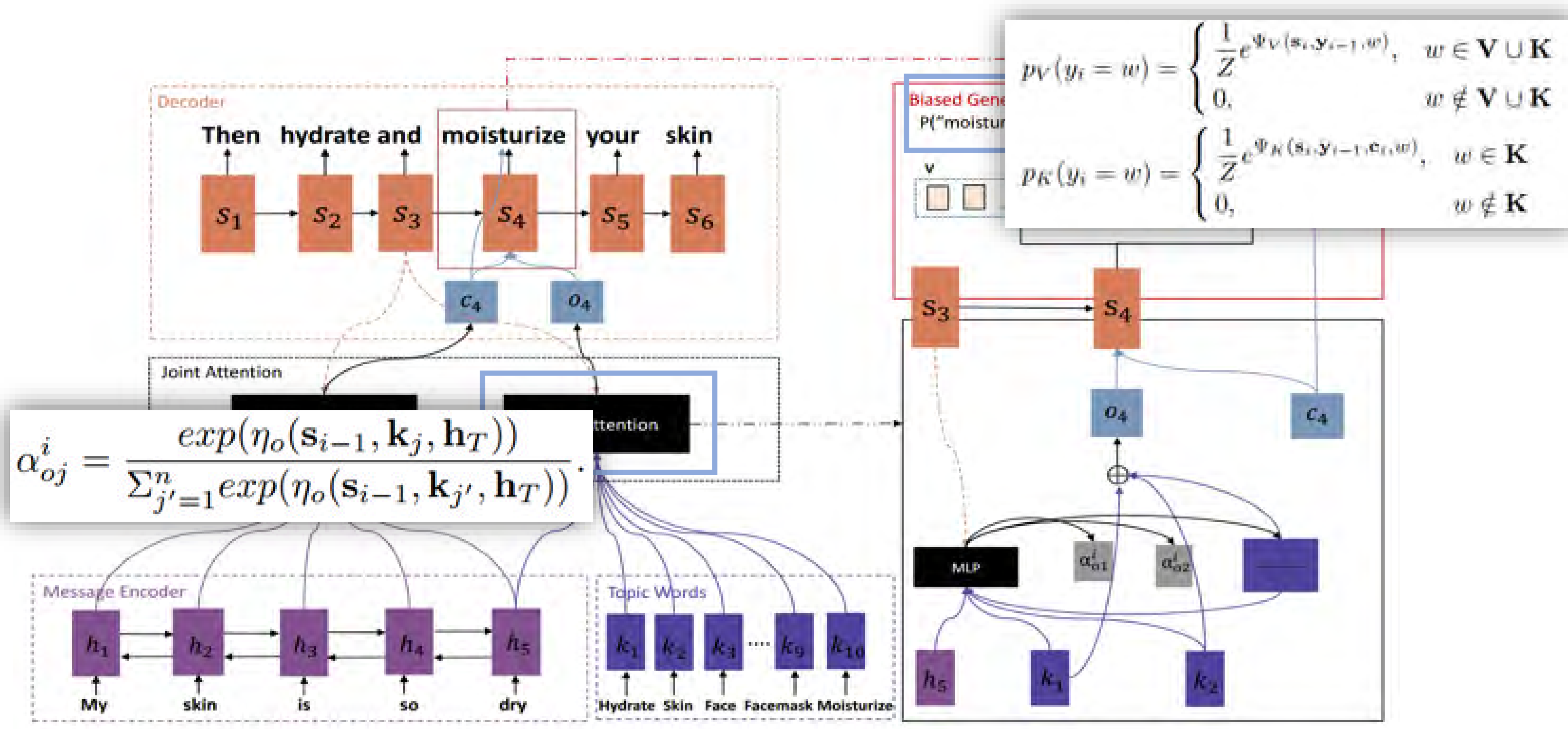
Generation-based Model

- Template-based approach(AIML)
- SMT-based approach
- Focus on neural net approaches

Encoder-Decoder for Sentence Generation



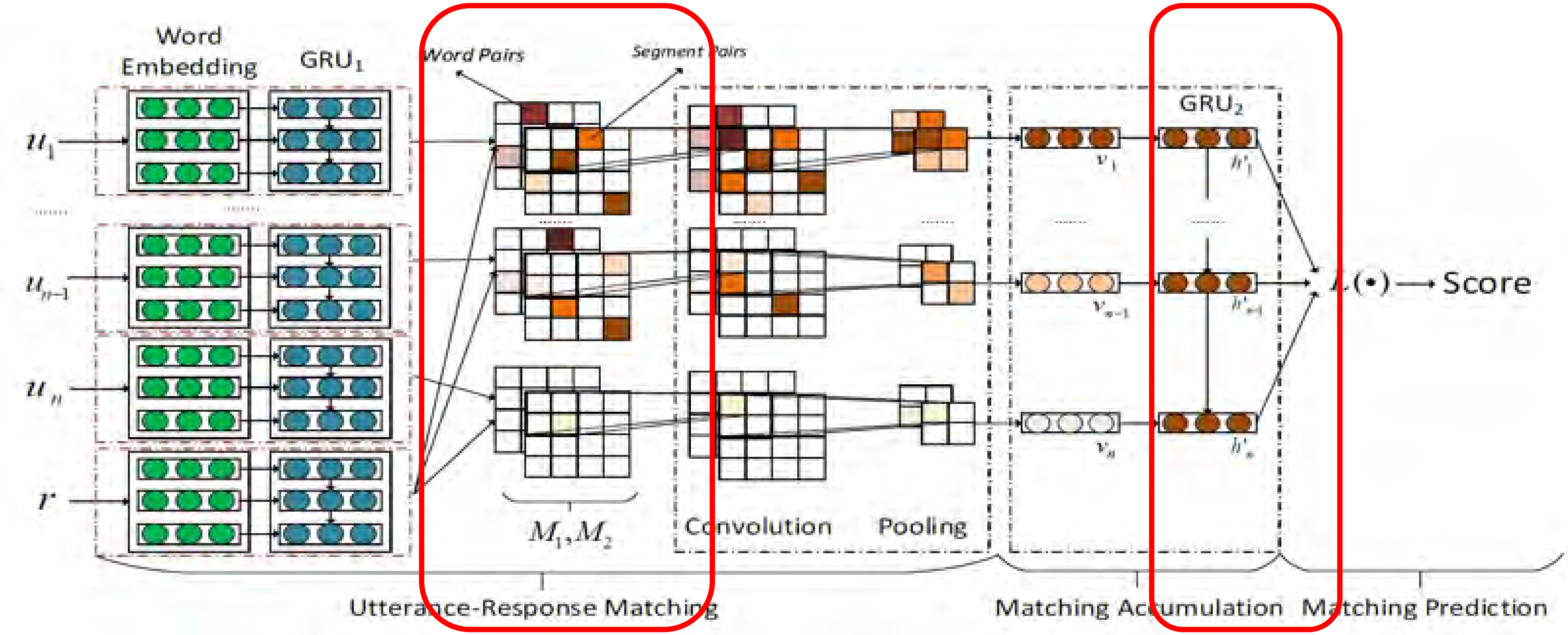
Topic-aware Neural Response Generation (TA-Seq2Seq)



Chen Xing et al. *Topic-aware Neural Response Generation,* In AAAI'17

Multi-Turn Conversation

Session-Response Matching : Sequential Matching Network

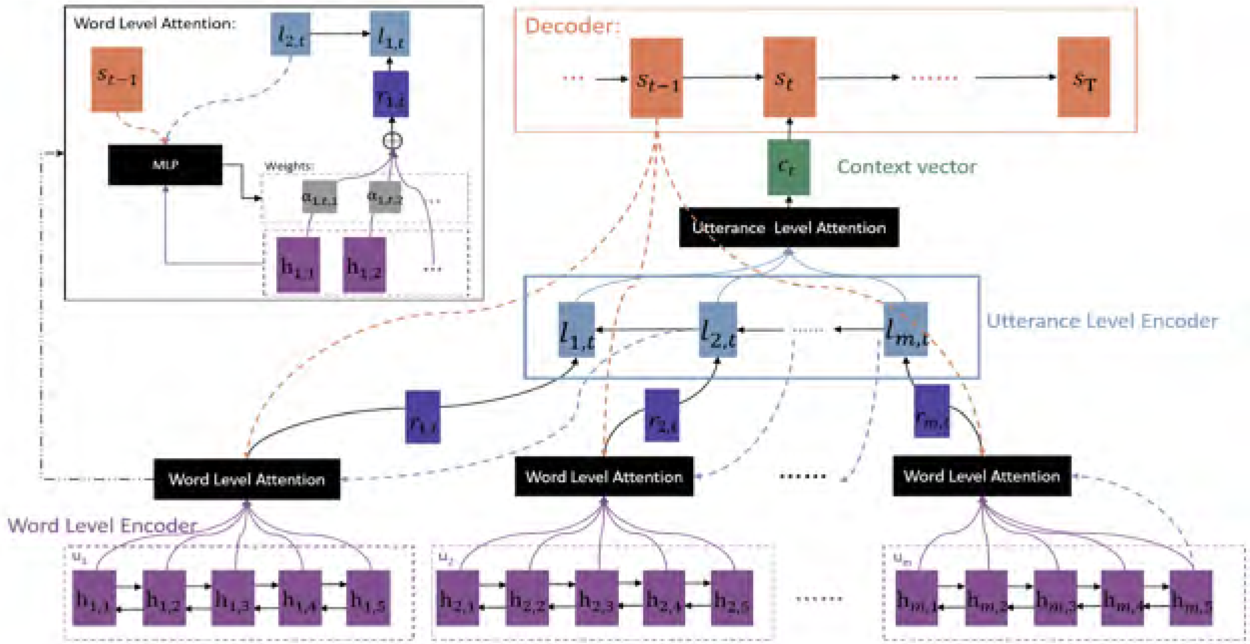


Matching each utterance with response at the beginning

Modeling utterance relationship

Yu Wu et al., *Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-based Chatbots,* To appear in ACL'17

Multi-turn Response Generation: Hierarchical Recurrent Attention Network



Chen Xing et al., *Hierarchical Recurrent Attention Network for Response Generation,* In arxiv

Visualization

Remaining Challenges

- Good public dataset
- Effective evaluation metric
- Sentiment-aware chat
- Effective memory mechanism
- Personalized chat

Reading Comprehension

SQuAD: 100,000+ Questions for Machine Comprehension of Text

The Stanford Question Answering Dataset Best Resource Paper in EMNLP 2016

Pranav Rajpurkar and **Jian Zhang** and **Konstantin Lopyrev** and **Percy Liang**
{pranavsr,zjian,klopyrev,pliang}@cs.stanford.edu
Computer Science Department
Stanford University

passag
e

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called "showers".

query What causes precipitation to fall?
answer gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Dataset	# of questions
Training	87,599
Dev	10,570
Test	~10K

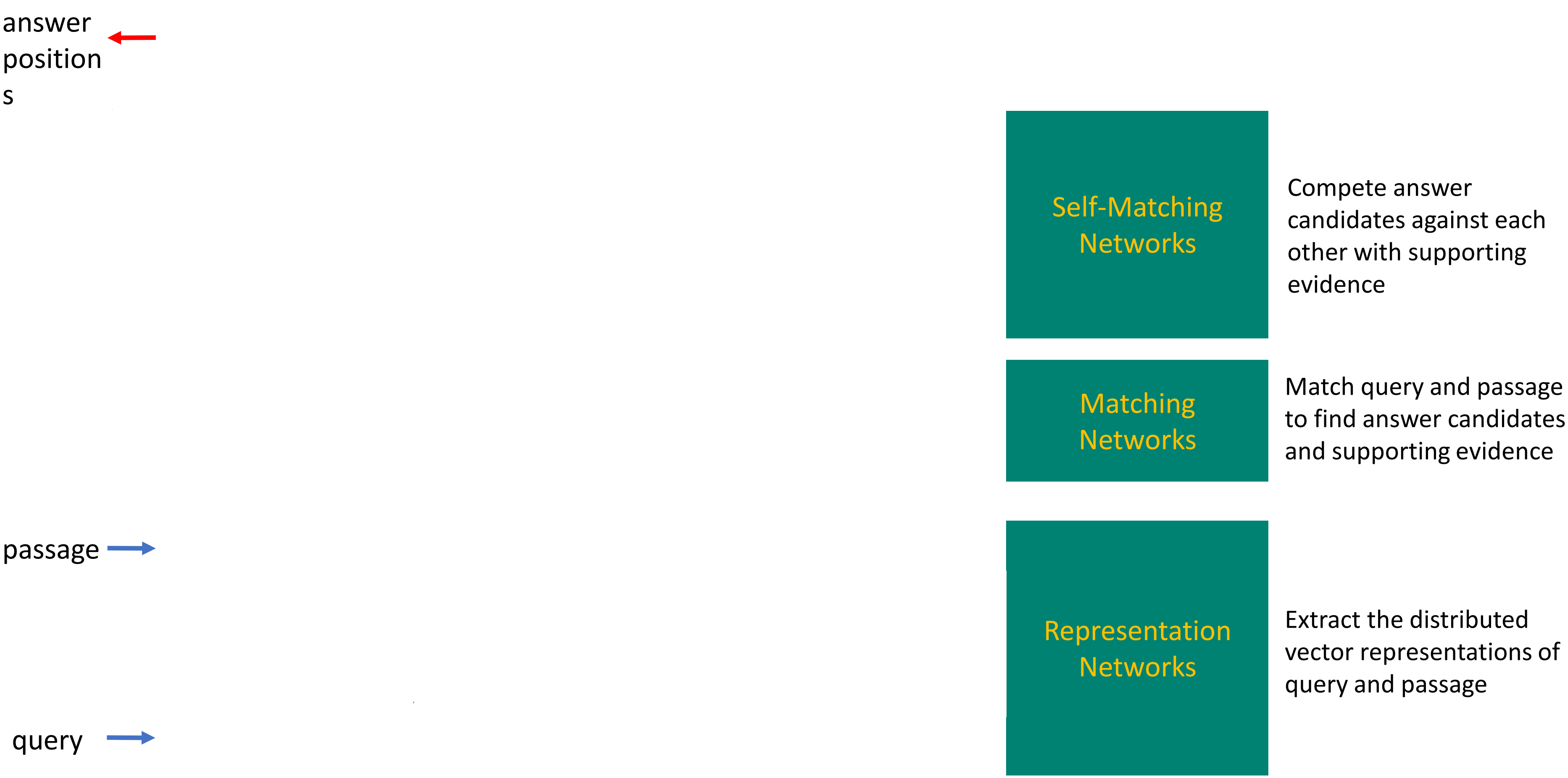
ImageNet-like evaluation: the test dataset is not public and we need to submit our system for evaluation on test set.

Rank	Model	EM	F1
1	r-net (ensemble) *Microsoft Research Asia*	75.863	82.947
2	ReasoNet (ensemble) *MSR Redmond*	73.419	81.752
2	Multi-Perspective Matching (diversity-ensemble) *IBM Research*	73.765	81.257
3	BiDAF (ensemble) *Allen Institute for AI & University of Washington*	73.314	81.089
4	Dynamic Coattention Networks (ensemble) *Salesforce Research*	71.625	80.383
5	r-net (single model) *Microsoft Research Asia*	71.258	79.66
6	Document Reader (single model) *Facebook AI Research*	69.967	78.974
7	ReasoNet (single model) *MSR Redmond*	69.107	78.895
7	FastQAExt *German Research Center for Artificial Intelligence*	70.849	78.857
8	Multi-Perspective Matching (single model) *IBM Research*	68.877	77.771
9	jNet (single model) *USTC & National Research Council Canada & York University*	68.73	77.393
10	BiDAF (single model) *Allen Institute for AI & University of Washington*	67.974	77.323
10	FastQA *German Research Center for Artificial Intelligence*	68.436	77.07
11	Match-LSTM with Ans-Ptr (Boundary+Ensemble) *Singapore Management University*	67.901	77.022
12	Iterative Co-attention Network *Fudan University*	67.502	76.786
13	Dynamic Coattention Networks (single model) *Salesforce Research*	66.233	75.896
13	RaSoR *Google NY, Tel-Aviv University*	67.387	75.543
14	Match-LSTM with Bi-Ans-Ptr (Boundary) *Singapore Management University*	64.744	73.743
15	Attentive CNN context with LSTM *NLPR, CASIA*	63.306	73.463
16	Fine-Grained Gating *Carnegie Mellon University*	62.446	73.327
16	Dynamic Chunk Reader *IBM*	62.499	70.956
17	Match-LSTM with Ans-Ptr (Boundary) *Singapore Management University*	60.474	70.695
18	Match-LSTM with Ans-Ptr (Sentence) *Singapore Management University*	54.505	67.748

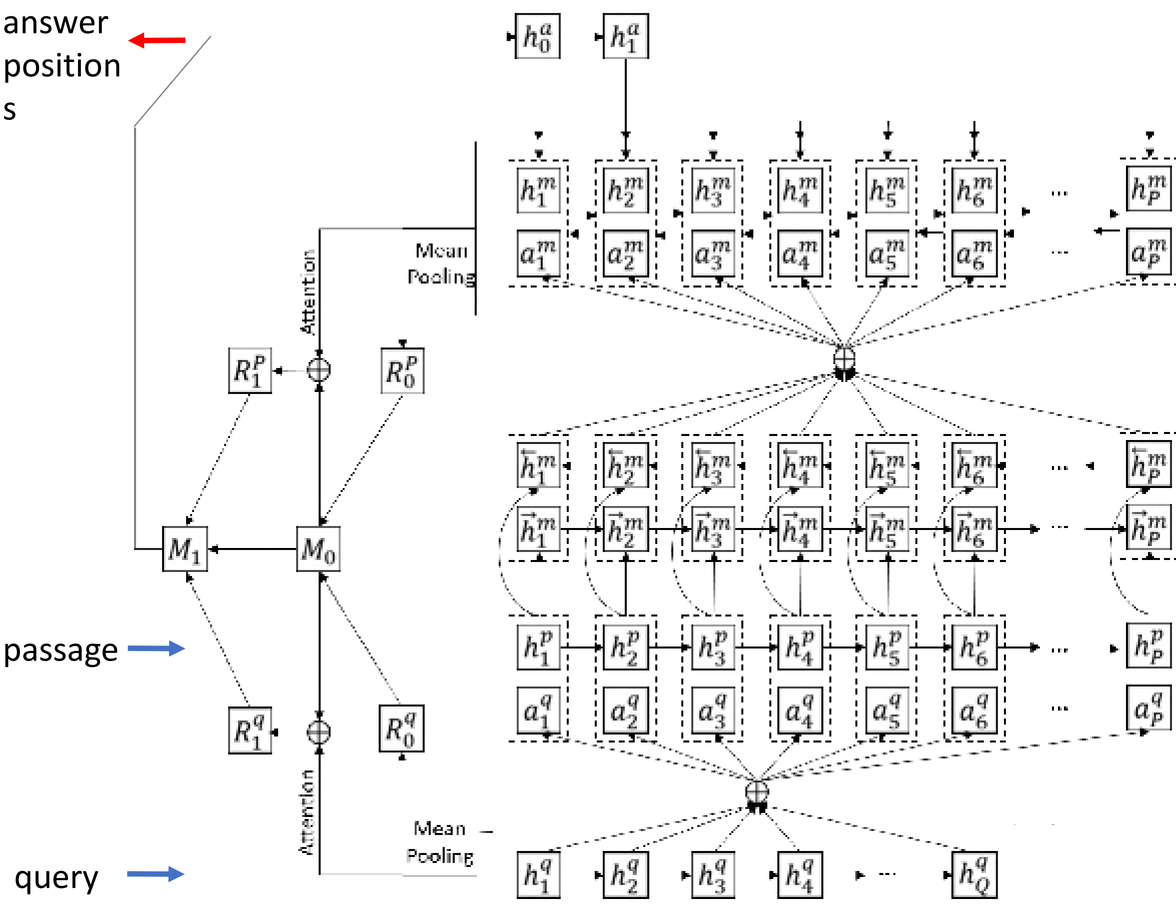
R-NET (MSRA)) is the top system (as of 2017-2-23) for both single model & ensemble model (on both dev and test set)

Rank	Model	EM	F1
1 14 days ago	r-net (ensemble) *Microsoft Research Asia*	76.922	84.006
2 17 days ago	ReasoNet (ensemble) *MSR Redmond*	75.034	82.552
3 a month ago	BiDAF (ensemble) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603	73.744	81.525
3 2 months ago	Multi-Perspective Matching (diversity-ensemble) *IBM Research* https://arxiv.org/abs/1612.04211	73.765	81.257
4 16 days ago	r-net (single model) *Microsoft Research Asia*	72.401	80.751
5 3 months ago	Dynamic Coattention Networks (ensemble) *Salesforce Research* https://arxiv.org/abs/1611.01604	71.625	80.383
6 15 hours ago	jNet (single model) *USTC & National Research Council Canada & York University* https://arxiv.org/abs/1703.04617	70.607	79.821
7 10 days ago	Ruminate Reader (single model) *New York University*	70.586	79.492
8 18 days ago	ReasoNet (single model) *MSR Redmond*	70.555	79.364
8 12 days ago	Document Reader (single model) *Facebook AI Research*	70.733	79.353
8 3 months ago	FastQAExt *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816	70.849	78.857
9 2 months ago	Multi-Perspective Matching (single model) *IBM Research* https://arxiv.org/abs/1612.04211	68.877	77.771
9 11 days ago	RaSoR (single model) *Google NY, Tel-Aviv University* https://arxiv.org/abs/1611.01436	69.642	77.696
10 4 months ago	BiDAF (single model) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603	67.974	77.323
10 3 months ago	FastQA *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816	68.436	77.07
11 5 months ago	Match-LSTM with Ans-Ptr (Boundary+Ensemble) *Singapore Management University* https://arxiv.org/abs/1608.07905	67.901	77.022
12 2 months ago	Iterative Co-attention Network *Fudan University*	67.502	76.786
13 5 months ago	Dynamic Coattention Networks (single model) *Salesforce Research* https://arxiv.org/abs/1611.01604	66.233	75.896
14 5 months ago	Match-LSTM with Bi-Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905	64.744	73.743
15 a month ago	Attentive CNN context with LSTM *NLPR, CASIA*	63.306	73.463
16 5 months ago	Fine-Grained Gating *Carnegie Mellon University* https://arxiv.org/abs/1611.01724	62.446	73.327
16 6 months ago	Dynamic Chunk Reader *IBM* https://arxiv.org/abs/1610.09996	62.499	70.956
17 7 months ago	Match-LSTM with Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905	60.474	70.695
18 7 months ago	Match-LSTM with Ans-Ptr (Sentence) *Singapore Management University* https://arxiv.org/abs/1608.07905	54.505	67.748
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221

Screenshot on 2017-3-27



Wenhui Wang et al, Gated Self-Matching Networks for Reading Comprehension and Question Answering, ACL 2017



Pointer Networks

Determine the start and end position of the answer in passage

Slef-Matching Networks

Compete answer candidates against each other with supporting evidence

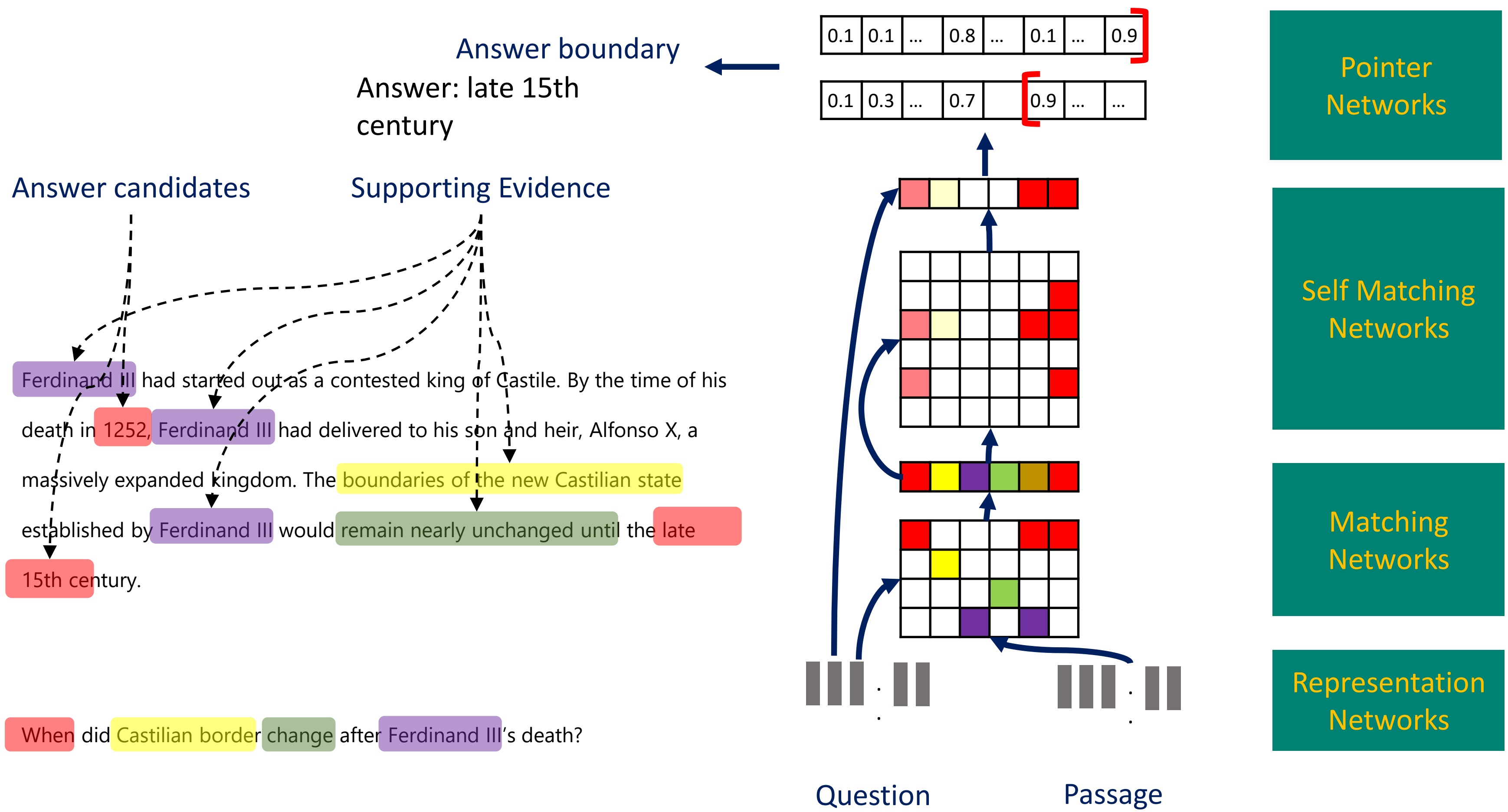
Matching Networks

Match query and passage to find answer candidates and supporting evidence

Representation Networks

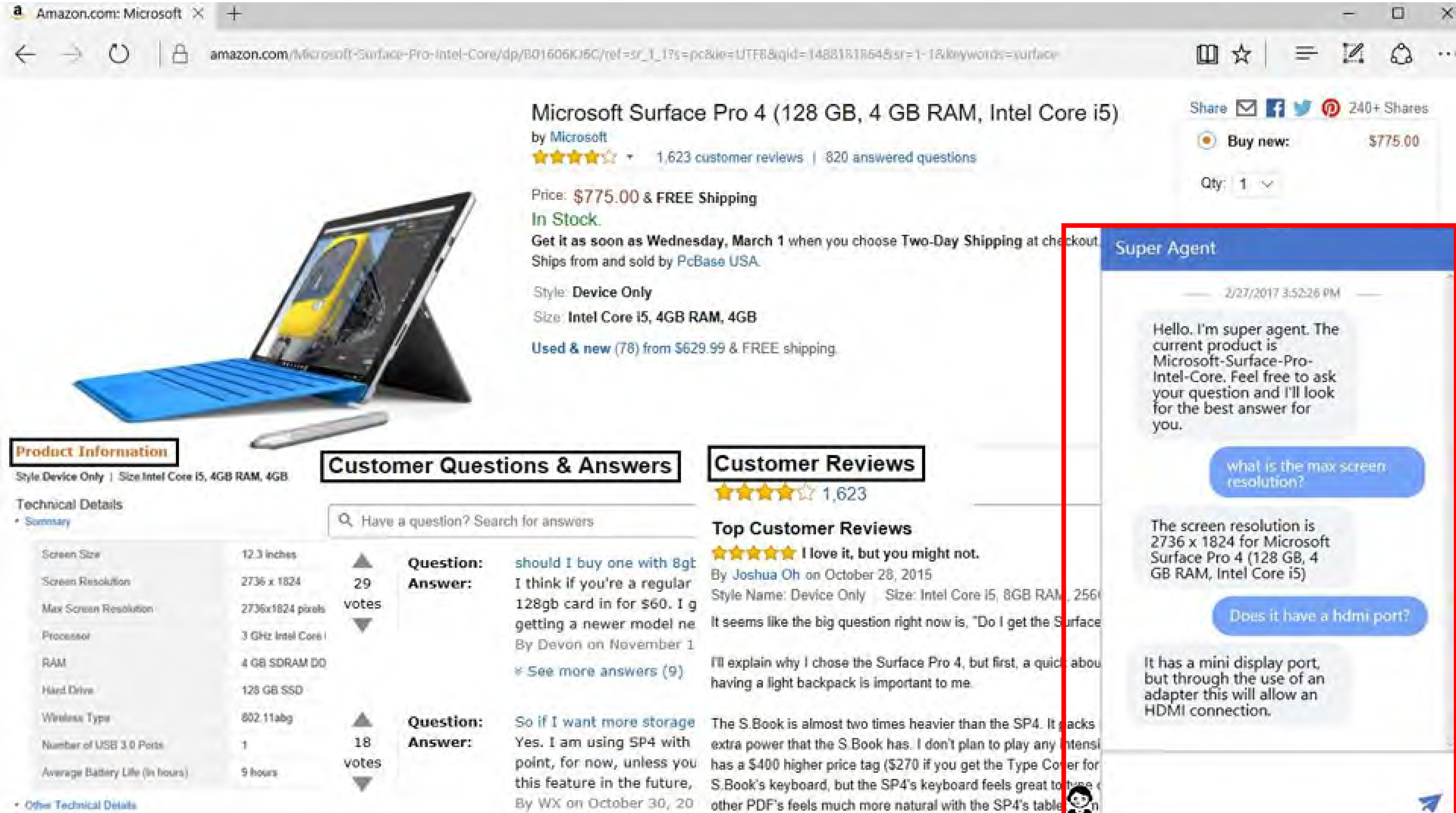
Extract the distributed vector representations of query and passage

Wenhui Wang et al, Gatd Self-Matching Networks for Reading Comprehension and Question Answering, ACL 2017



Wenhui Wang et al, Gatd Self-Matching Networks for Reading Comprehension and Question Answering, ACL 2017

Customer Support with Reading Comprehension



Lei Cui et al, SuperAgent: A Customer Service Chatbot for E-commerce, to appear at ACL 2017

Remaining Challenges

- Difficulty level setting and corresponding dataset
- From answer extraction to answer inference
- How to use knowledge and common sense

NLP in Future 5-10 Years

- Spoken Translator popularly used
- Natural conversation(chit-chat, QnA and task-oriented dialogue) reaches satisfactory quality
- Improves the productivity of customer support
- Generation of poetry, song, novel and news
- Deeply used in various verticals such as education, bank, healthcare, law, etc.

Future Direction

- Explore the explainable learning to understand the mechanism of AI and NLP
- Fusing knowledge and data to improve the efficiency of learning
- Domain adaptation via transfer learning
- Self-evolvement via reinforcement learning
- Leverage user profiling for personalized service