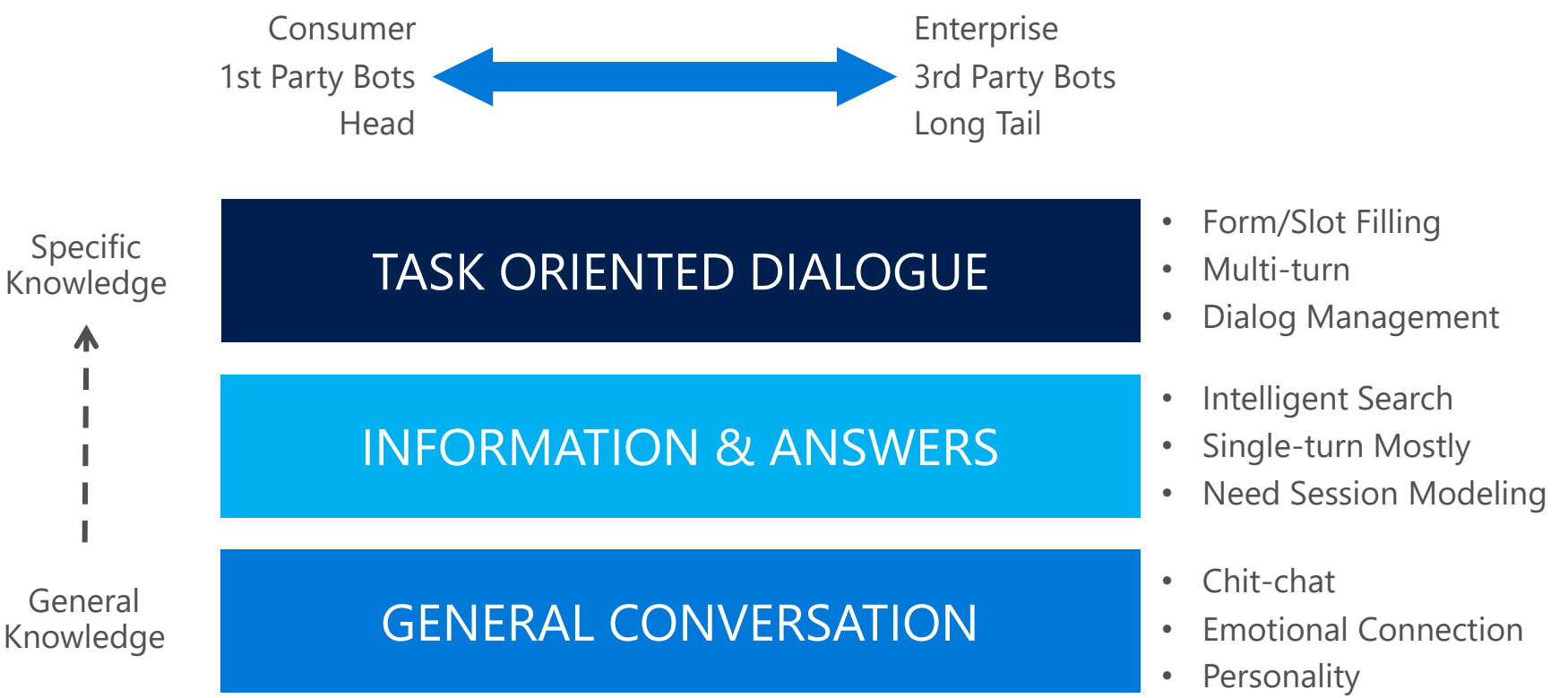
Natural Language Conversation Engine: Chit-chat, QnA and Dialogue

Ming Zhou
Microsoft Research Asia
编程之美2017 @Tsinghua University

Microsoft

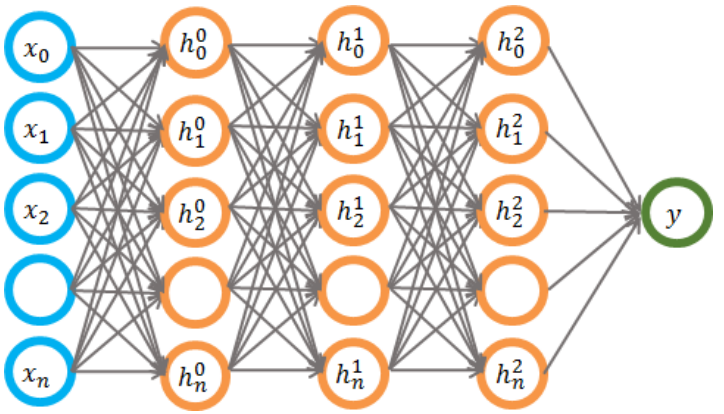
Conversation Engine Architecture



DNN4NLP Fundamentals

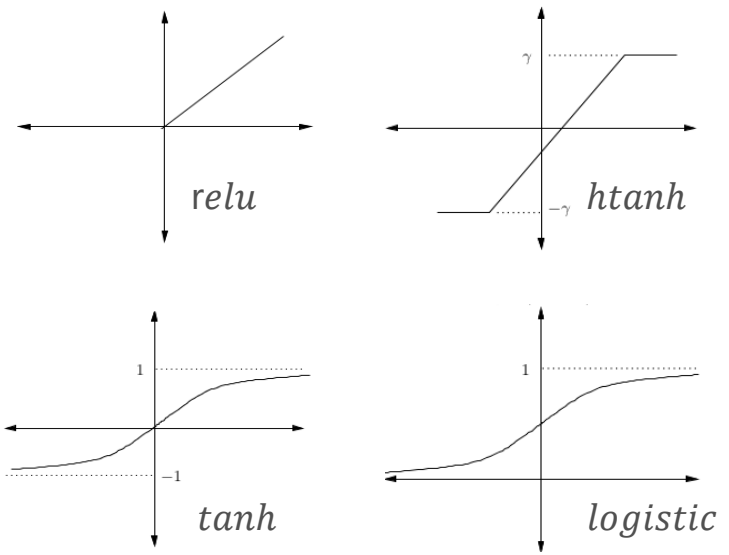
Deep Neural Network

- Deep Neural Network :
  - Involve multiple level neural networks
  - Non-Linear Learner



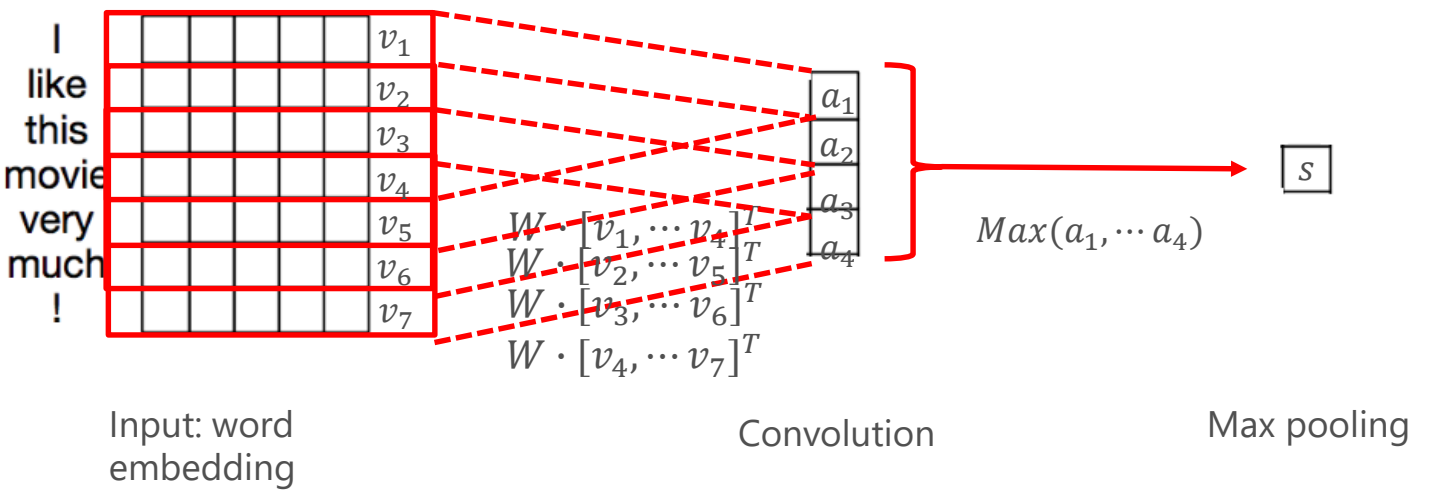
$$h^0 = f(w^0 x) \qquad h^1 = f(w^1 h^0)$$

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

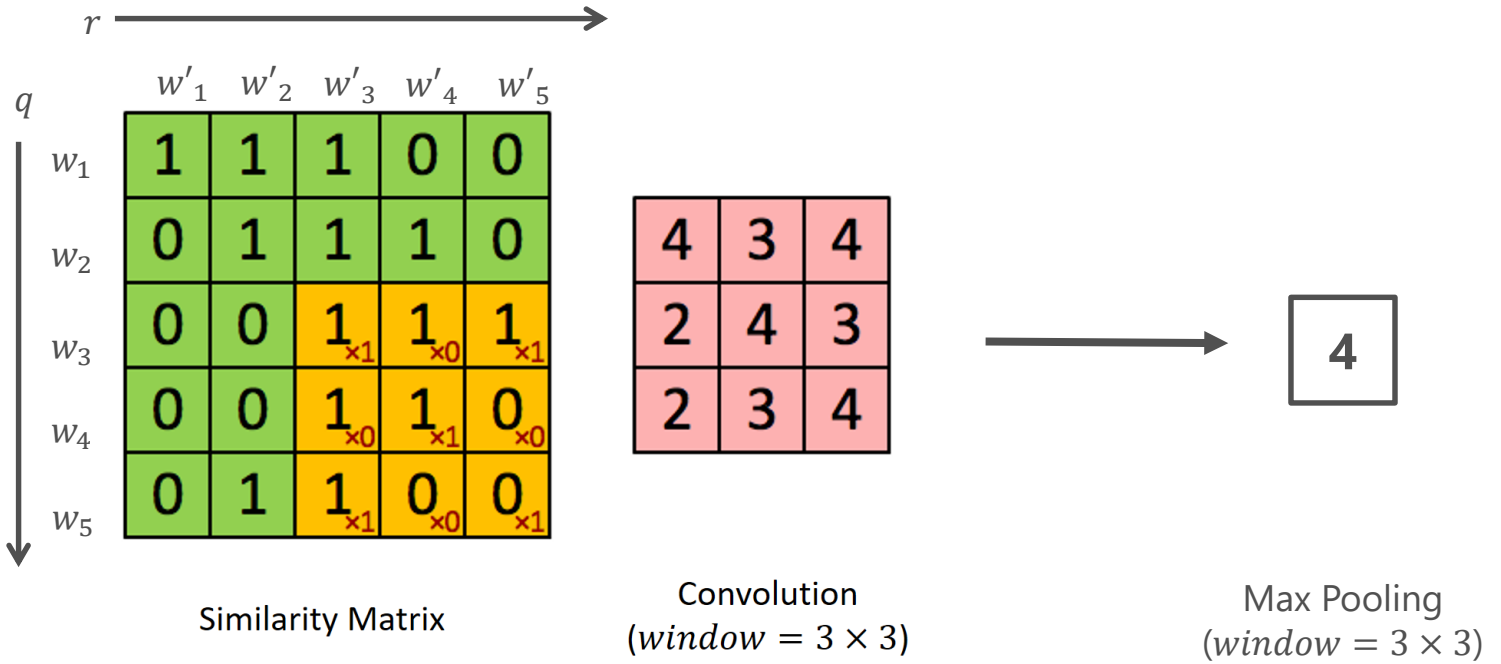


Active functions: $y = f(x)$

CNN (1D)

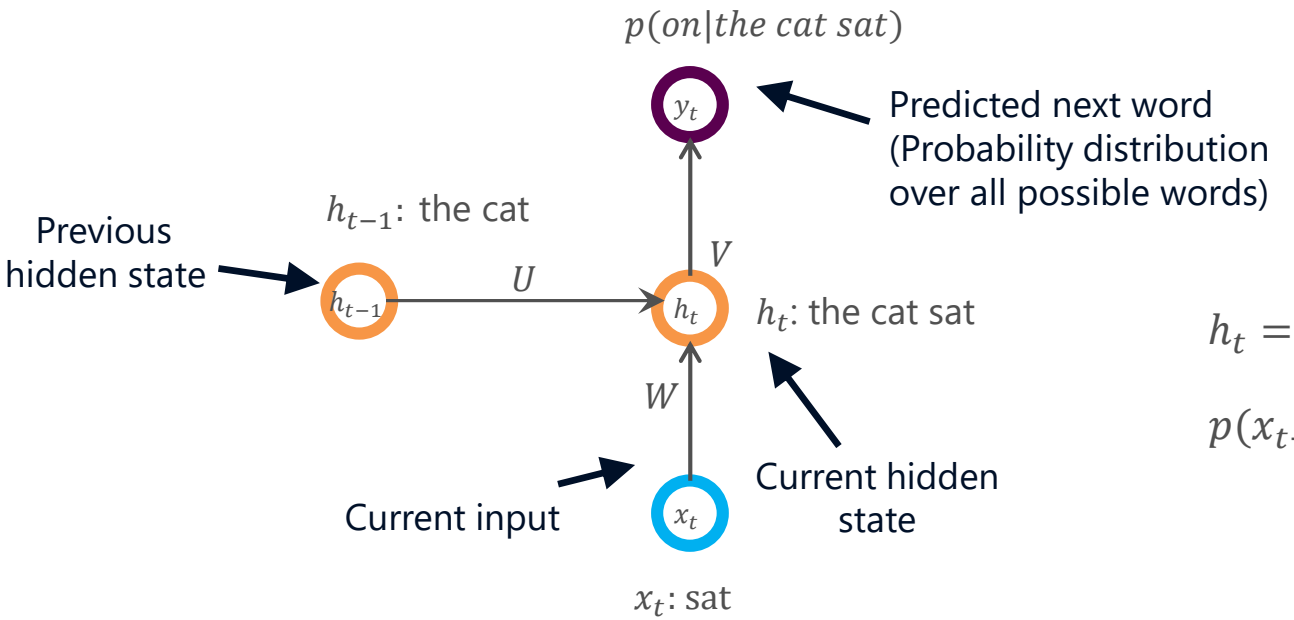

I
like
this
movie
very
much
!
v_1
v_2
v_3
v_4
v_5
v_6
v_7
$W \cdot [v_1, \cdots v_4]^T$
$W \cdot [v_2, \cdots v_5]^T$
$W \cdot [v_3, \cdots v_6]^T$
$W \cdot [v_4, \cdots v_7]^T$
a_1
a_2
a_3
a_4
$Max(a_1, \cdots a_4)$
s
Input: word embedding
Convolution
Max pooling

CNN (2D)


r
q
w'_1 w'_2 w'_3 w'_4 w'_5
w_1 1 1 1 0 0
w_2 0 1 1 1 0
w_3 0 0 1 1 1
w_4 0 0 1 1 0
w_5 0 1 1 0 0
×1 ×0 ×1
×0 ×1 ×0
×1 ×0 ×1
4 3 4
2 4 3
2 3 4
4
Similarity Matrix
Convolution
(window = 3 × 3)
Max Pooling
(window = 3 × 3)

RNN(Recurrent Neural Network)

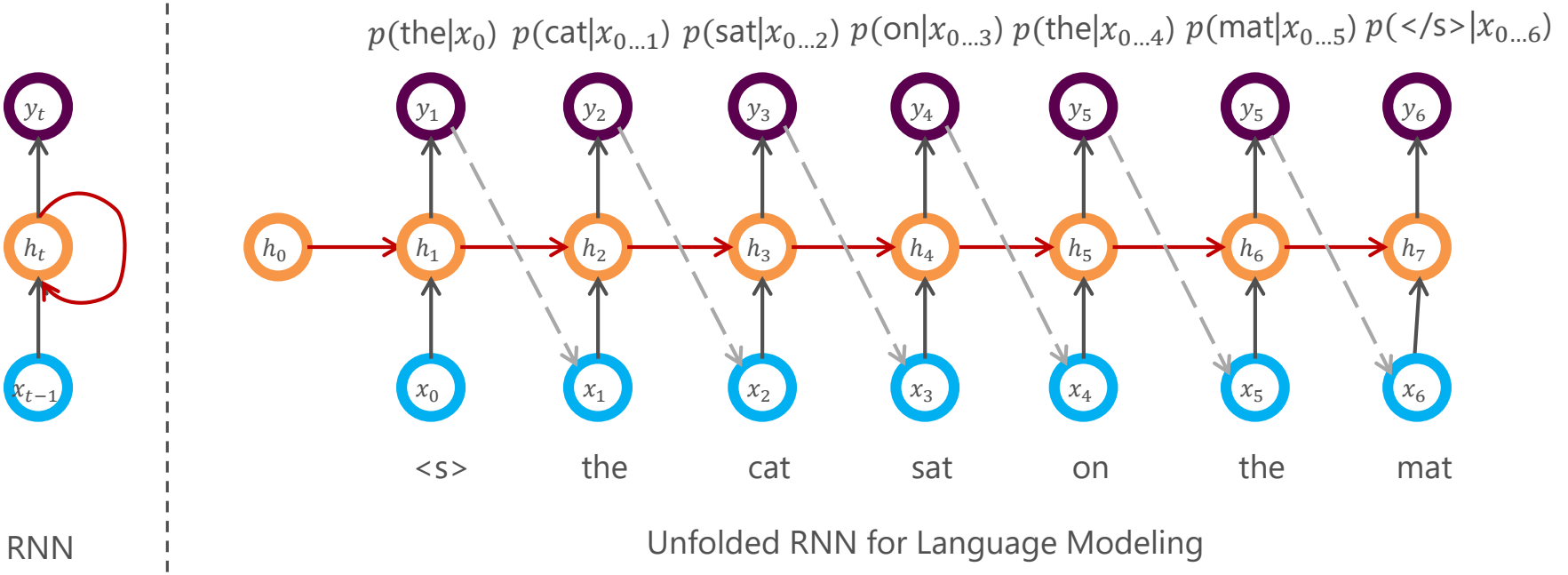
- Inputs: History h_{t-1} at time $t-1$ and input x_t at time t
- Output: History h_t at time t and probability of next word y_t



$$h_t = f(Uh_{t-1} + Wx_t)$$

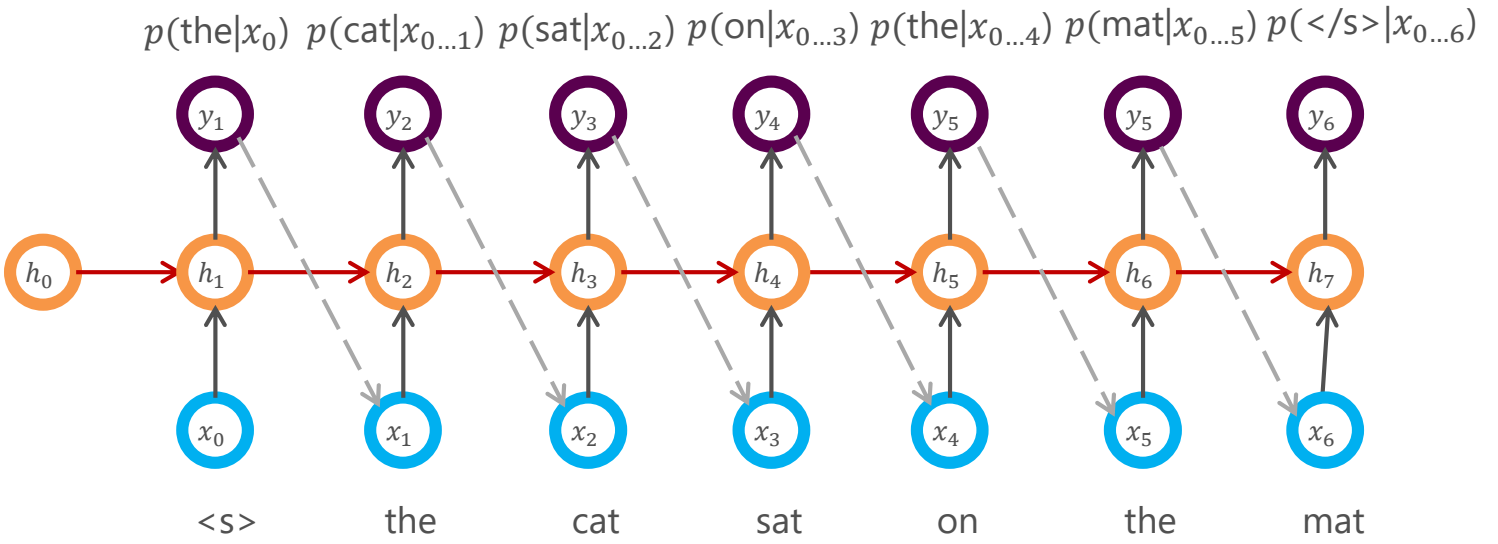
$$p(x_{t+1}) = softmax(Vh_t)$$

RNN(Recurrent Neural Network)



Sentence: The cat sat on the mat

Training of RNN



$p(the\ cat\ sat\ on\ the\ mat$ </s>$) =$
$p(\text{the}|x_0)$*$p(\text{cat}|x_{0\ldots1})$*$p(\text{sat}|x_{0..2})$*$p(\text{on}|x_{0..3})$
*$p(\text{the}|x_{0..4})$ *$p(\text{mat}|x_{0..5})$ *$p(\text{</s>}|x_{0..6})$

$Loss = -\log(p(the\ cat\ sat\ on\ the\ mat$ </s>$))$
$= -\sum_i \log(p(x_i|x_{0..i-1}))$

SGD Training

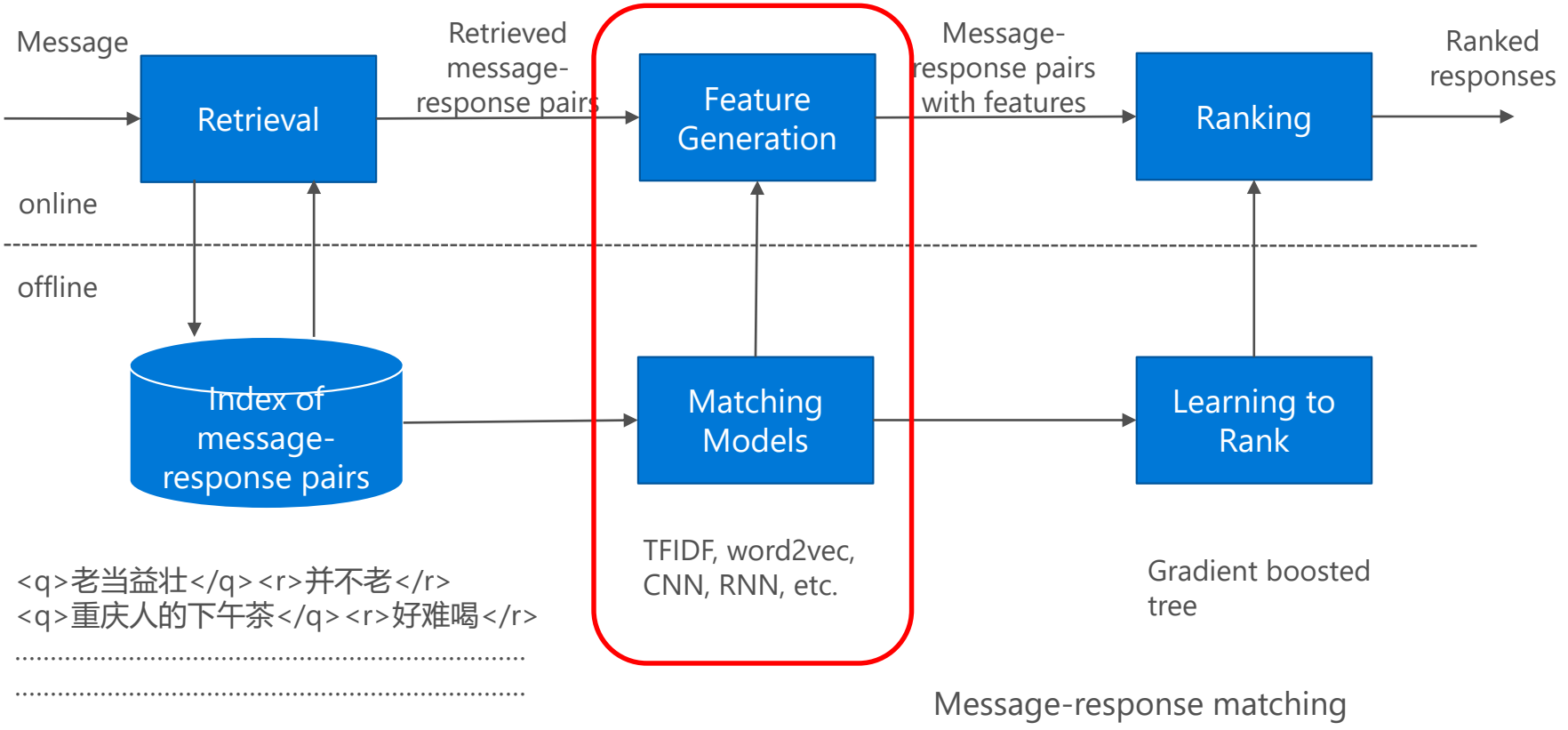
1. Sample a sentence S from corpus;
2. Build unfolded RNN for S;
3. Run forward to compute $Loss$;
4. Run backward to compute gradient;
5. Update parameters;
6. Repeat until $\mathbf{perplexity} < \varepsilon$ or epoch $> n$;

Skip the details of LSTM/GRU here

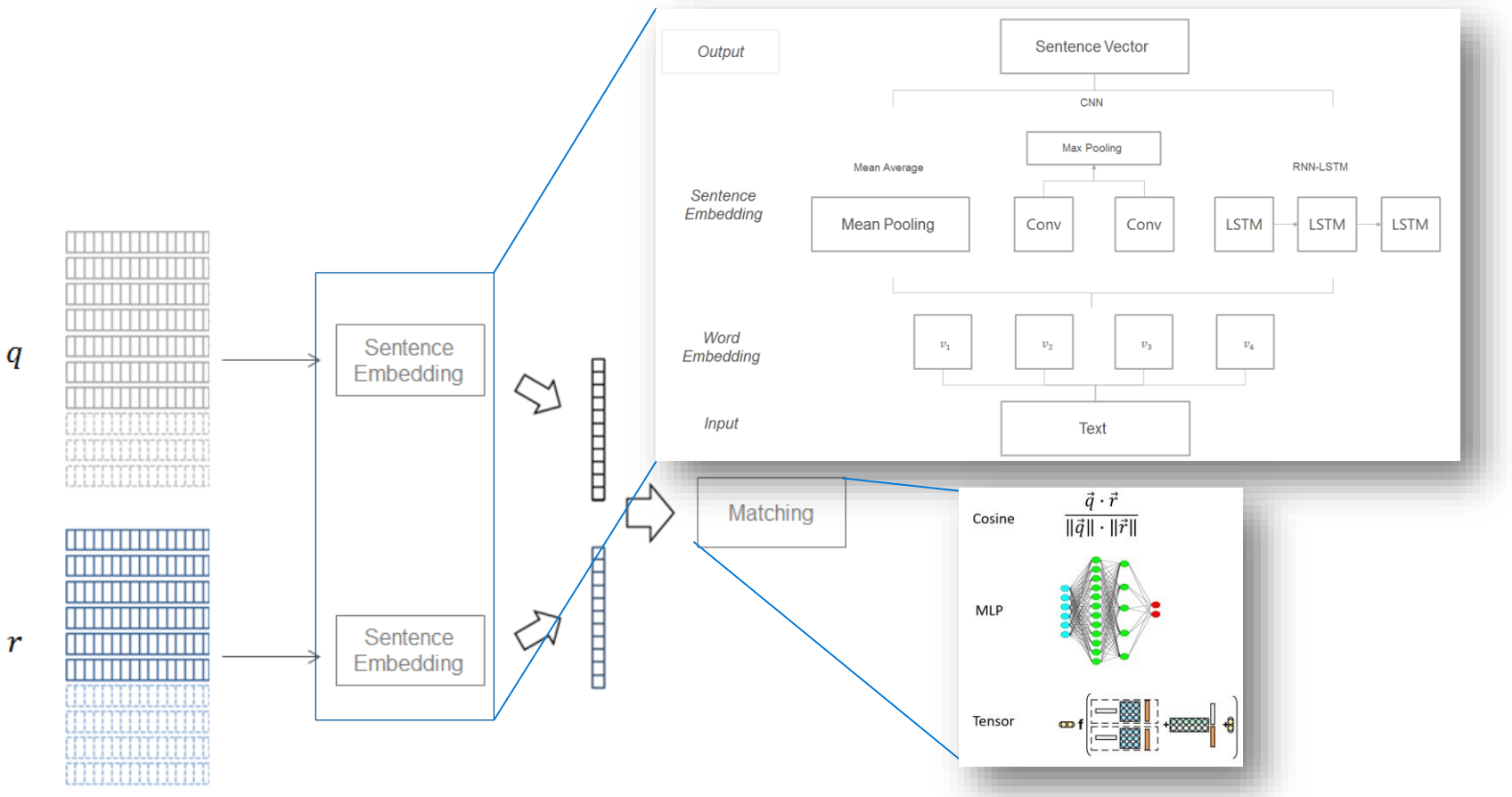
General Chat Engine

Microsoft

Architecture of Retrieval-based Chatbot (Single-Turn)



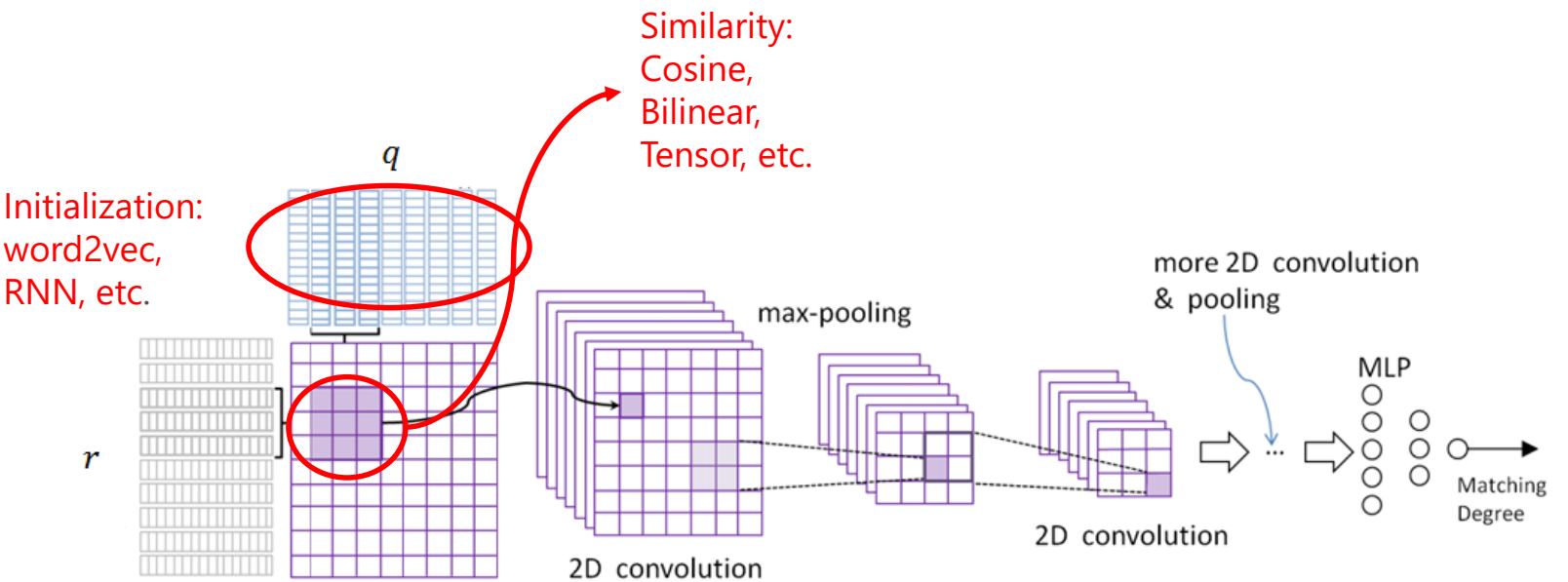
Basic Models for Message-Response Matching : Architecture I



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

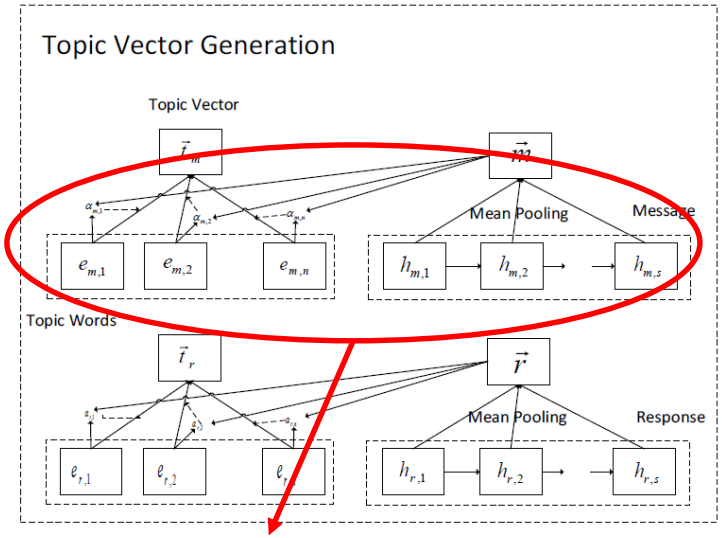
Basic Models for Message-Response Matching : Architecture II

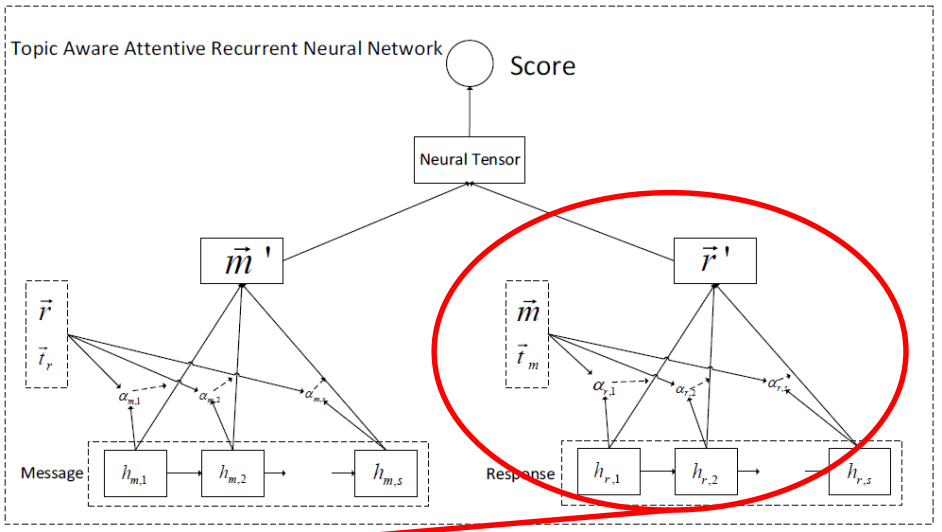


Baotian Hu et al. *Convolutional Neural Network Architectures for Matching Natural Language Sentences,* In NIPS'14
Liang Pang et al. *Text Matching as Image Recognition*, In AAAI'16
Shengxian Wan et al. *A Deep Architecture for Semantic Matching with Multiple Positional Sentence Representations*, In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

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

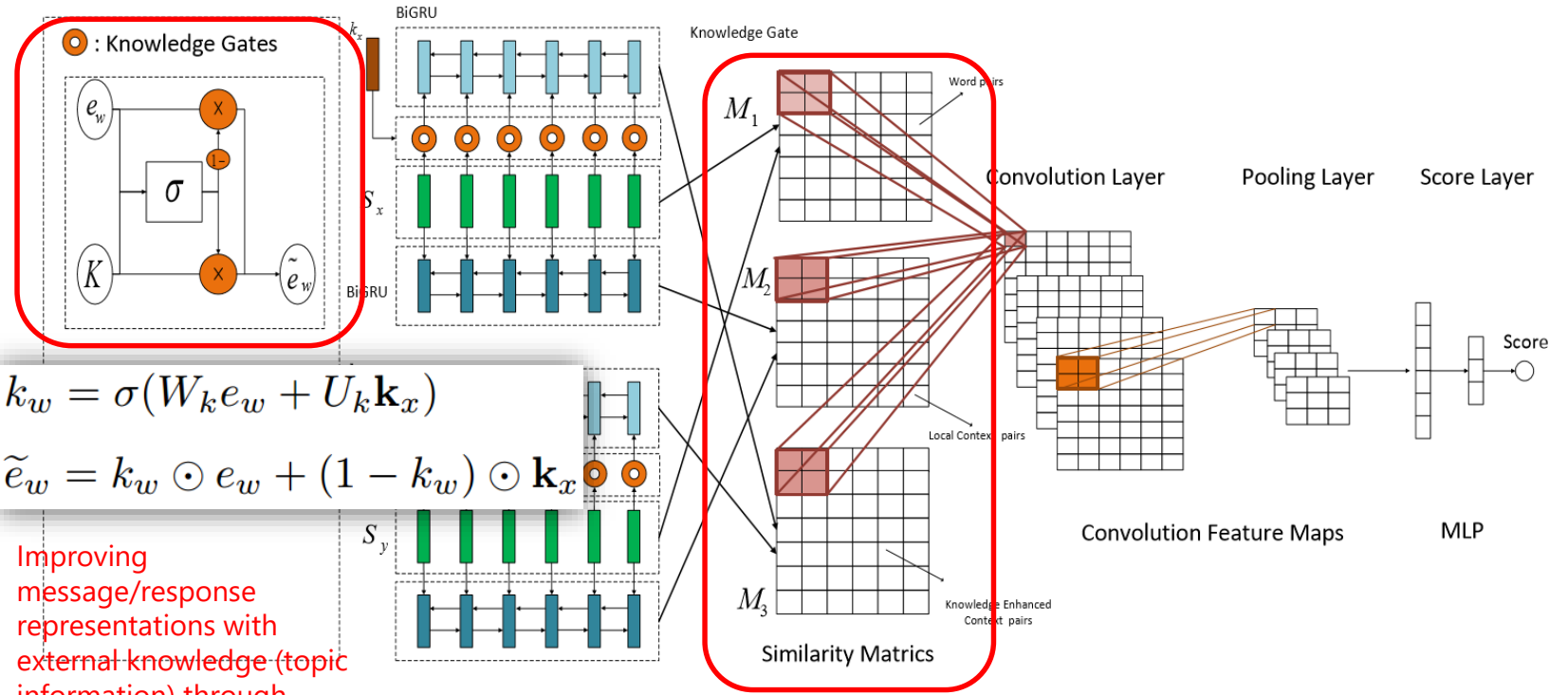
Joint attention with message/response and knowledge (topics)

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

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)



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

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

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

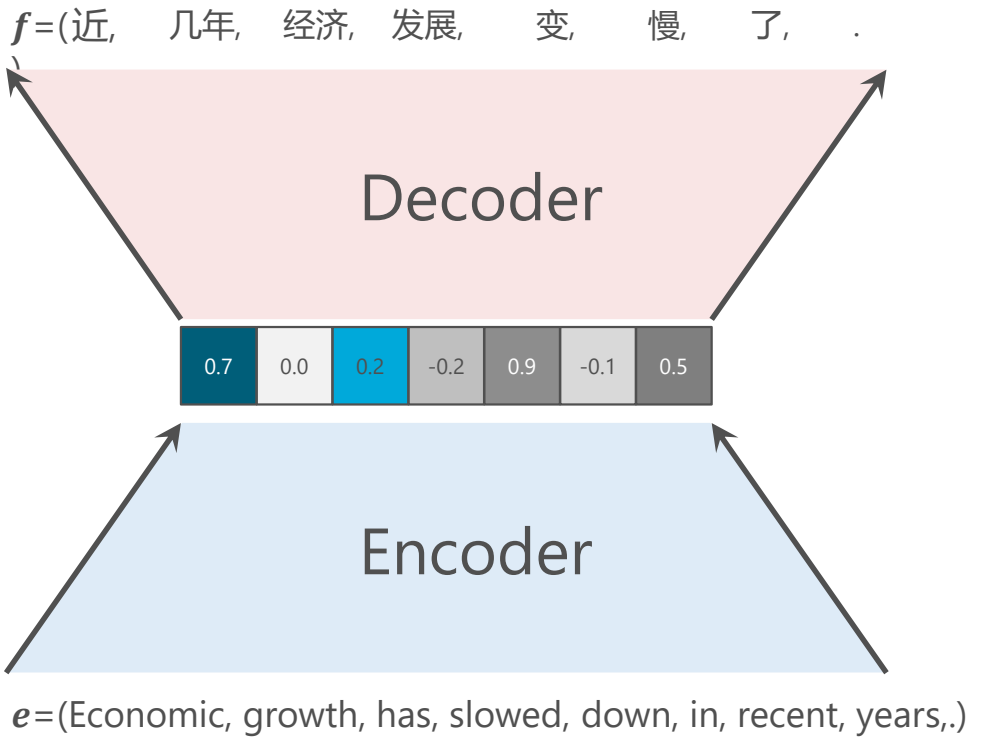
Evaluation on the Largest Public Data - Ubuntu Corpus

	$R_2@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
Cosine (I)	68.1	38.3	48.2	68.6
MLP (I)	65.1	25.6	38	70.3
CNN (I)	66.5	22.1	36	68.4
CNN+Tensor (I)	74.3	34.9	51.2	79.7
LSTM (I)	72.5	36.1	49.4	80.1
CNN (II)	73.6	38	53.4	77.7
MatchPyramid (II)	74.3	42	55.4	78.6
MV-LSTM (II)	76.7	41.0	56.5	81.0
TAARNN (I)	**77.0**	**40.4**	**56.0**	**81.7**
KEHNN (II)	**78.6**	**46**	**59.1**	**81.9**

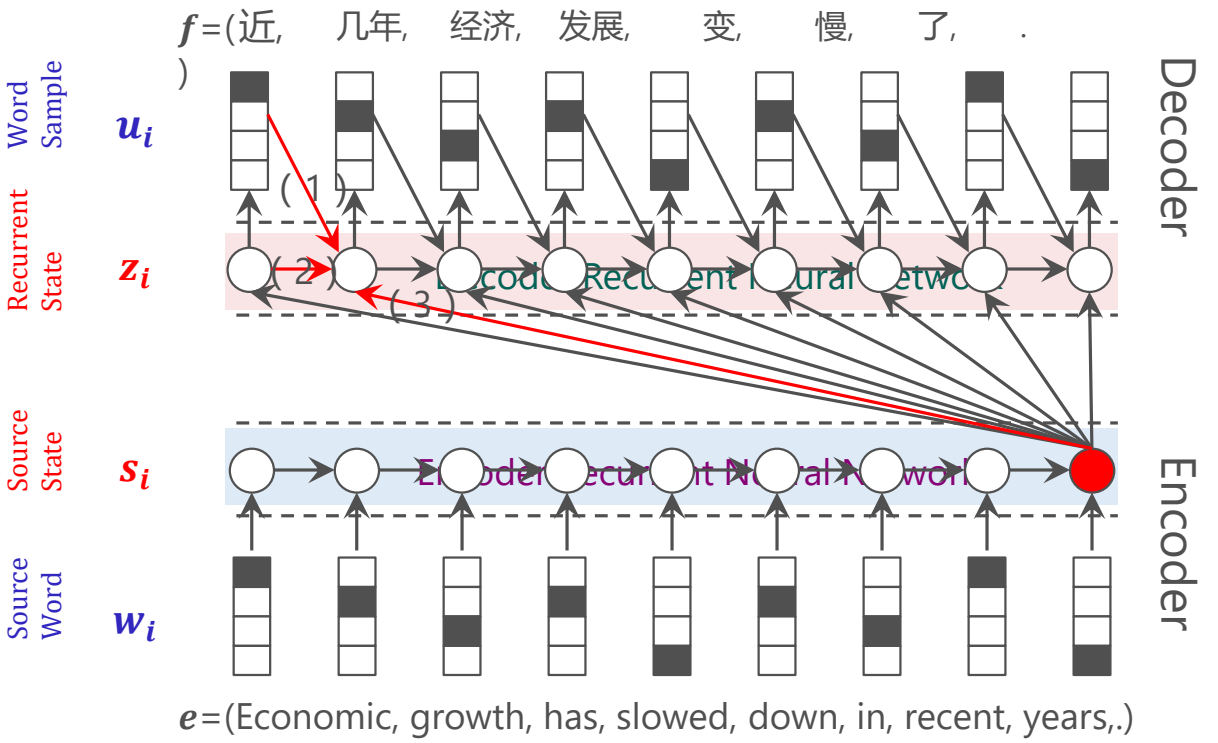
- Train : Validation : Test = 1M : 0.5M : 0.5M
- Negative examples are randomly sampled
- R_n@k means recall at position k in n candidates
- (I) and (II) mean ``in architecture I or II" respectively

Neural Generation Model

Encoder-Decoder for Sentence Generation

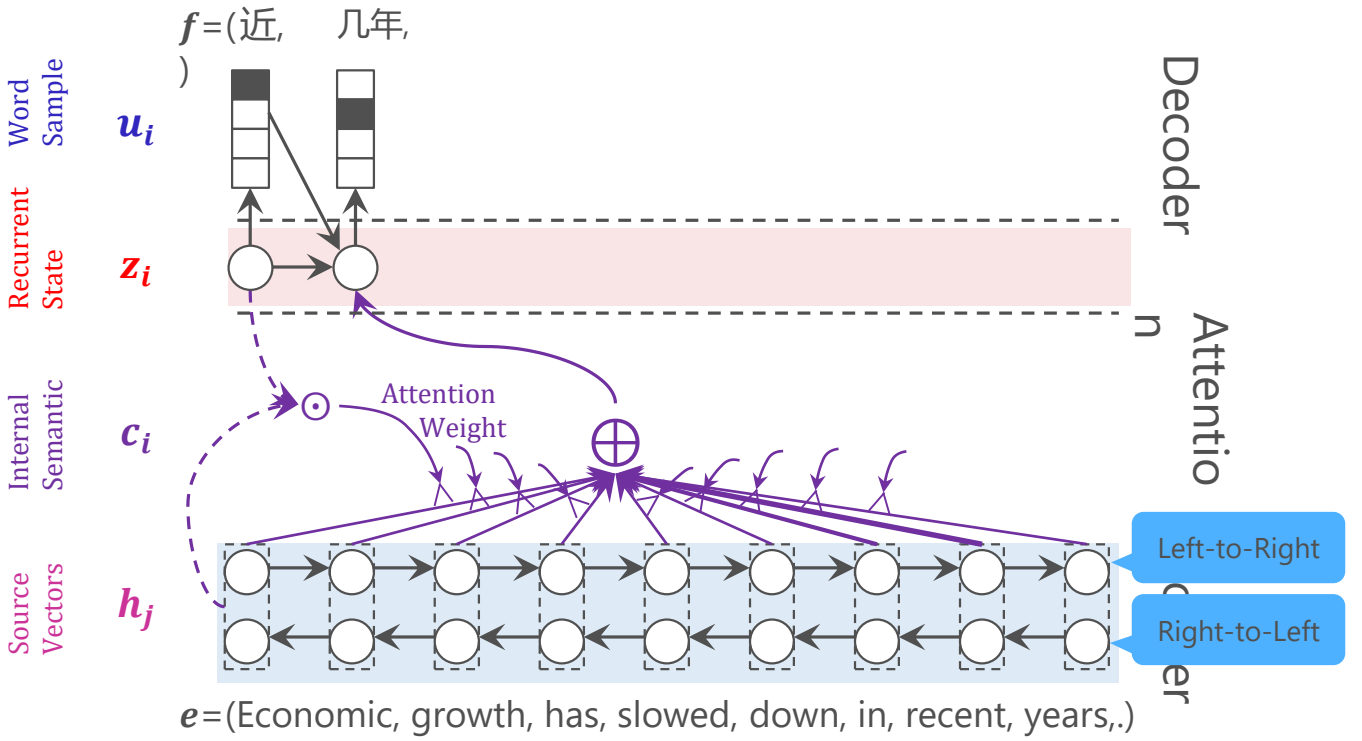


Encoder-Decoder for Sentence Generation



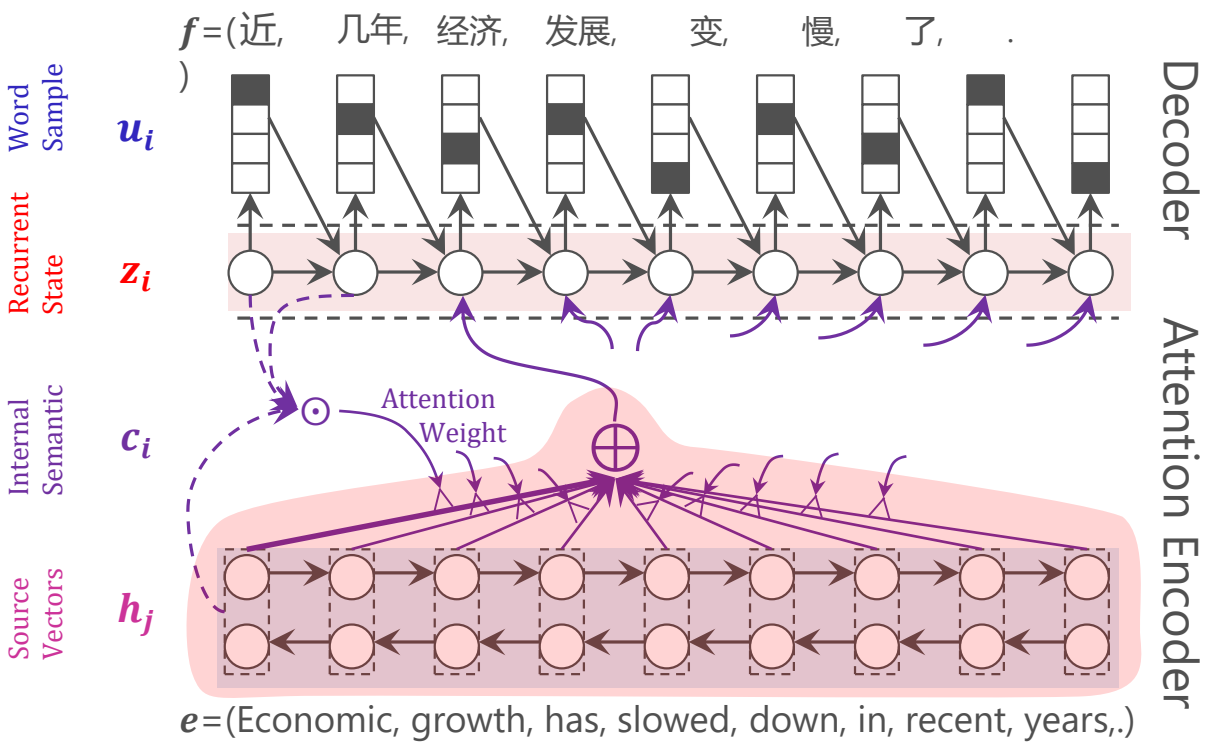
Sutskever et al., NIPS, 2014

Attention based Encoder-Decoder



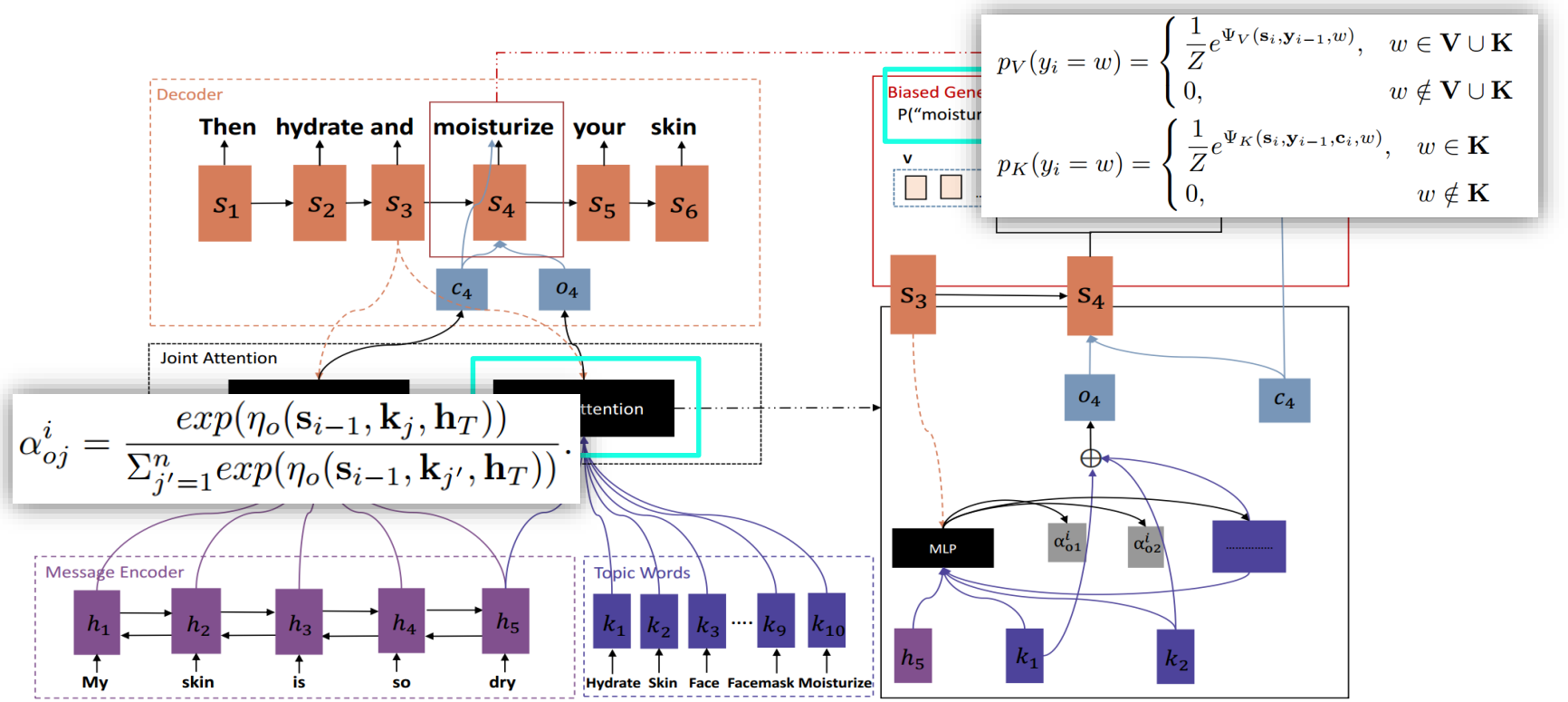
Bahdanau et al., ICLR, 2015

Attention based Encoder-Decoder



Bahdanau et al., ICLR, 2015

Topic-aware Neural Response Generation (TA-Seq2Seq)



$$p_V(y_i = w) = \begin{cases} \frac{1}{Z} e^{\Psi_V(\mathbf{s}_i, \mathbf{y}_{i-1}, w)}, & w \in \mathbf{V} \cup \mathbf{K} \\ 0, & w \notin \mathbf{V} \cup \mathbf{K} \end{cases}$$

$$p_K(y_i = w) = \begin{cases} \frac{1}{Z} e^{\Psi_K(\mathbf{s}_i, \mathbf{y}_{i-1}, \mathbf{c}_i, w)}, & w \in \mathbf{K} \\ 0, & w \notin \mathbf{K} \end{cases}$$

$$\alpha_{oj}^i = \frac{exp(\eta_o(\mathbf{s}_{i-1}, \mathbf{k}_j, \mathbf{h}_T))}{\Sigma_{j'=1}^n exp(\eta_o(\mathbf{s}_{i-1}, \mathbf{k}_{j'}, \mathbf{h}_T))}.$$

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

Evaluation Results

Models	+2	+1	0	Kappa
S2SA	32.3%	36.7%	31.0%	0.8116
S2SA-MMI	33.1%	34.8%	32.1%	0.7848
S2SA-TopicConcat	35.9%	29.3%	34.8%	0.6633
S2SA-TopicAttention	42.3%	27.6%	30.0%	0.8299
TA-Seq2Seq	44.7%	24.9%	30.4%	0.8417

Table 1: Human annotation results

Models	PPL-D	PPL-T	distinct-1	distinct-2
S2SA	147.04	133.11	604/.091	1168/.207
S2SA-MMI	147.04	133.11	603/.151	1073/.378
S2SA-TopicConcat	150.45	132.12	898/.116	2197/.327
S2SA-TopicAttention	133.81	**119.55**	894/.106	2057/.277
TA-Seq2Seq	134.63	**122.82**	1355/.161	2970/.401

Table 2: Results on automatic metrics

- Data are crawled from Baidu Tieba
- Train : Validation : Test = 5M : 10K : 1K

Examples

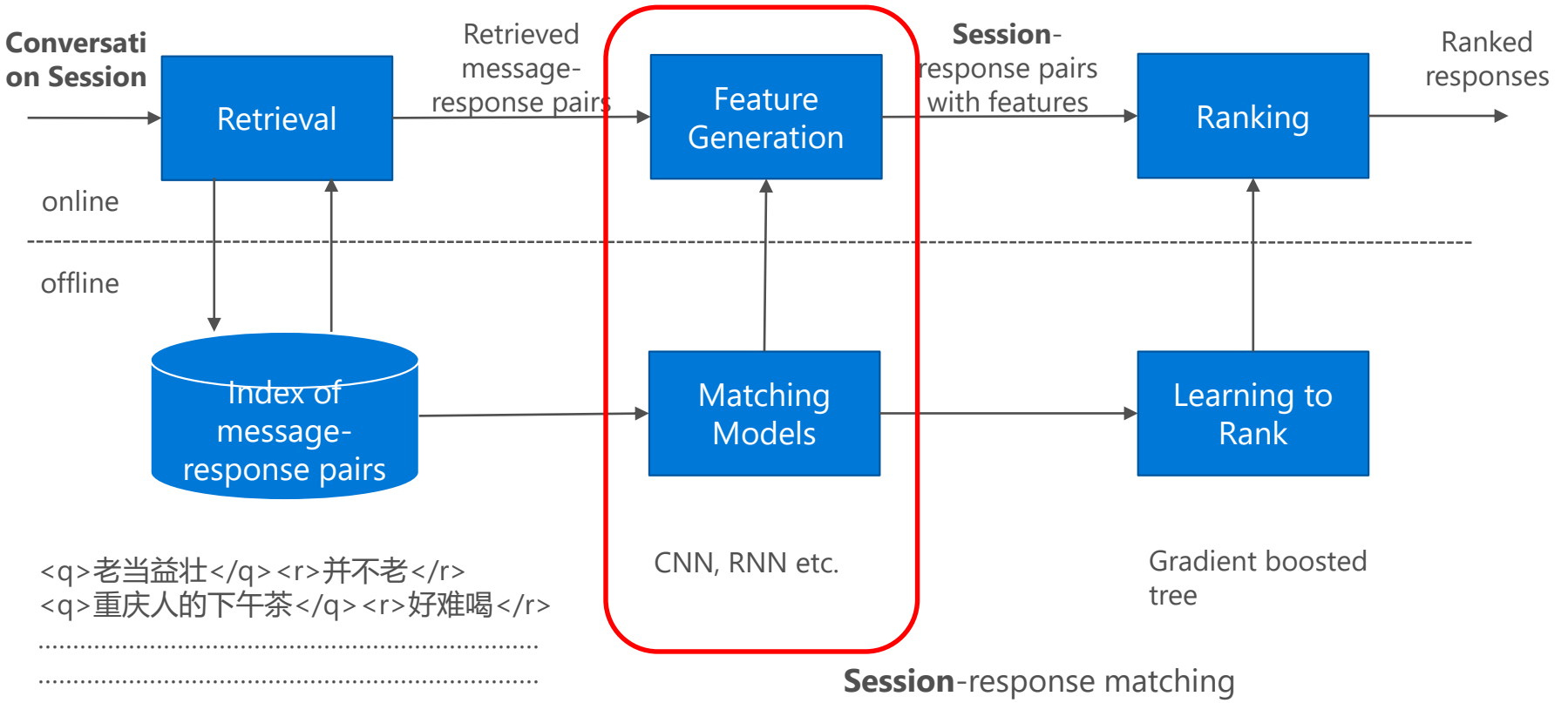
Message	TA-Seq2Seq	Baseline 1 S2SA-MMI	Baseline 2 S2SA
你也玩摄影 You also like photography	我不玩**摄影**，只是想拍个**照片**而已 I don't work on **photography**. I just want to take a few **pictures**.	额额额. . . Well well well....	额. . . Well...
我的左脸颊上有一些红血丝 There is some redness on my left cheek.	我也是**敏感肌** My **skin** is also very **sensitive**.	我也是啊 Me too.	啊啊啊啊 Oh my god oh my god
大学生的话可以去实习吗 Can a college student apply for an internship	可以的，如果你着急的话可以先填个**申请表** Yes, you can fill an **application form** first if you are in a hurry.	可以的可以的 Yes, you can.	可以的 Yes.
我的皮肤好干 My skin is so dry.	那就**补水保湿**吧 Then **hydrate** and **moisturize** your skin.	我也是啊 Me too.	啊啊啊啊 Oh my god oh my god.

* Words in bold are topic words

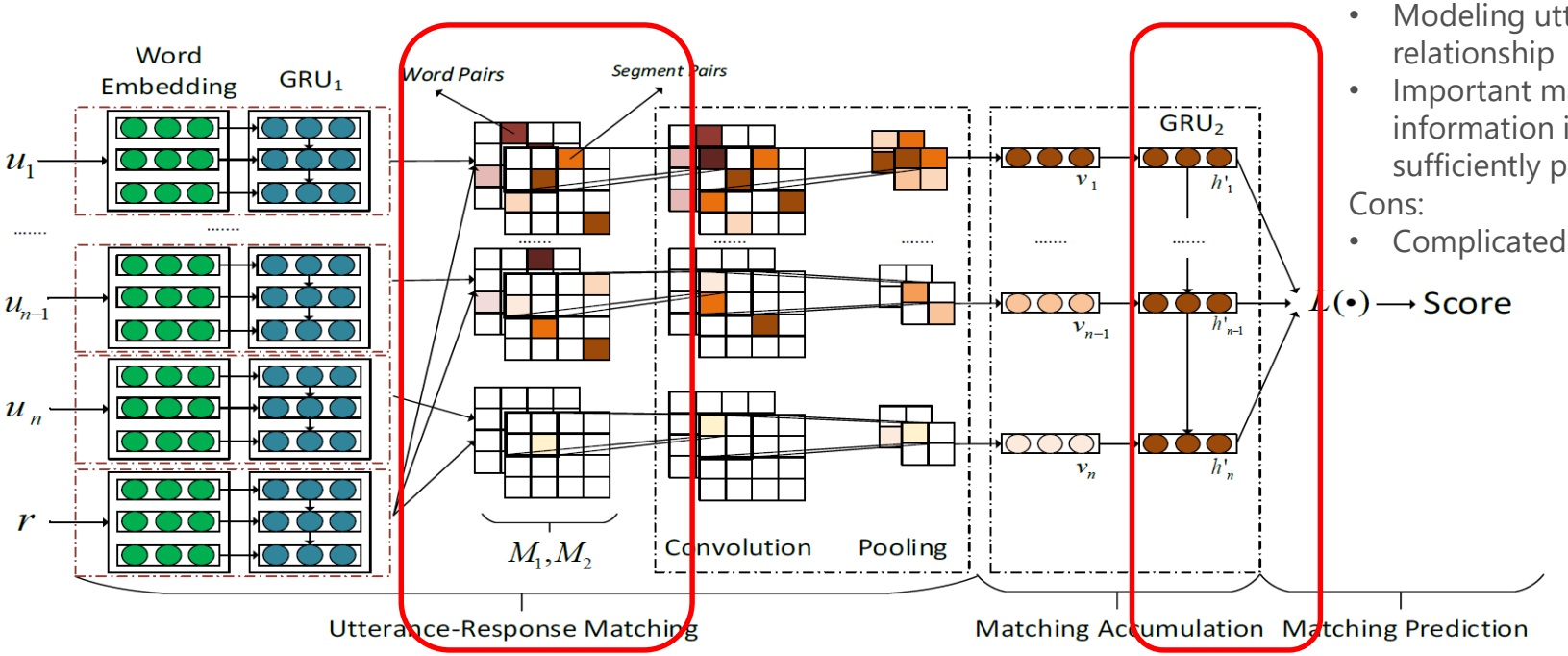
Multi-Turn Conversation

Microsoft

System Overview



Session-Response Matching : Sequential Matching Network



Pros:
- Modeling utterance relationship
- Important matching information is sufficiently preserved

Cons:
- Complicated

Matching each utterance with response at the beginning

Modeling utterance relationship

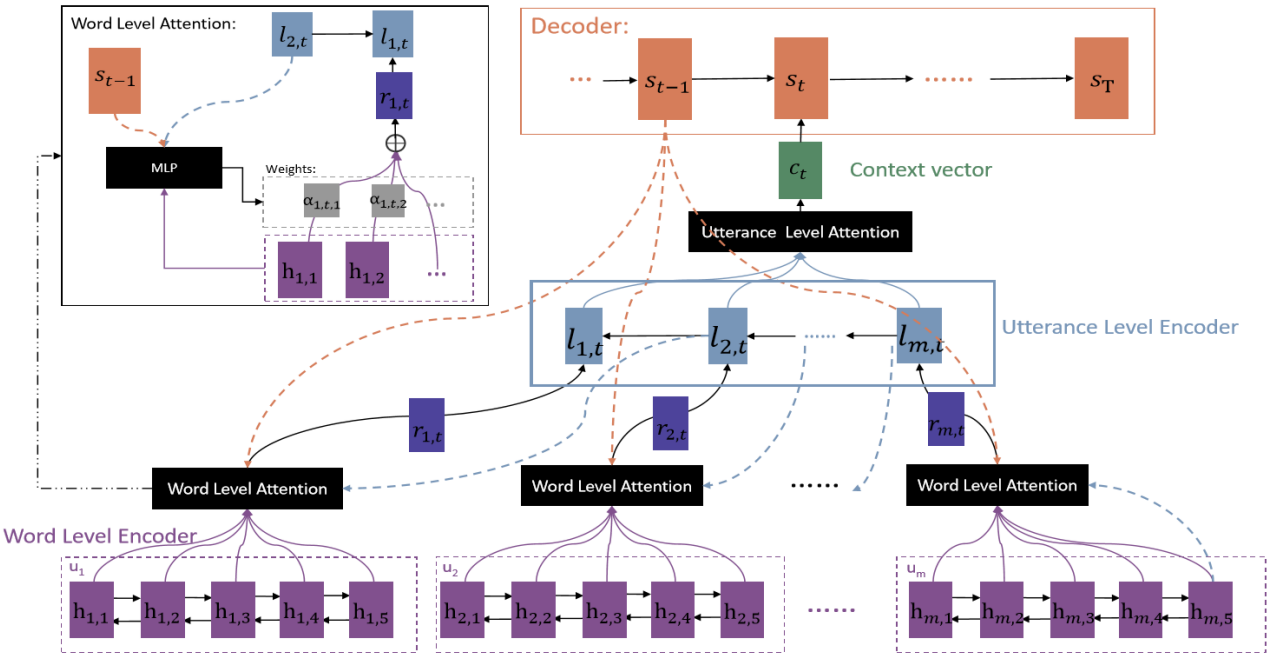
Yu Wu et al., *Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-based Chatbots,* To appear in ACL'17

Evaluation on the Ubuntu Corpus

	$R_2@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
Cosine	65.9	41.0	54.5	70.8
CNN	84.8	54.9	68.4	89.6
LSTM	90.1	63.8	78.4	94.9
BiLSTM	89.5	63.0	78.0	94.4
Multi-View	90.8	66.2	80.1	95.1
DL2R	89.9	62.6	78.3	94.4
SMN	**92.6**	**72.6**	**84.7**	**96.1**

- Train : Validation : Test = 1M : 0.5M : 0.5M
- Negative examples are randomly sampled
- R_n@k means recall at position k in n candidates

Multi-turn Response Generation: Hierarchical Recurrent Attention Network



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

Evaluation Results

Model	Validation Perplexity	Test Perplexity
S2SA	43.679	44.508
HRED	46.279	47.467
VHRED	44.548	45.484
HRAN	40.257	**41.138**

Table 1: Perplexity results

Models	Win	Loss	Tie	Kappa
HRAN v.s. S2SA	**27.3**	20.6	52.1	0.37
HRAN v.s. HRED	**27.2**	21.2	51.6	0.35
HRAN v.s. VHRED	**25.2**	20.4	54.4	0.34

Table 2: Human annotation results (in %)

- Data are crawled from Douban Group
- Train : Validation : Test = 1M : 10K : 1K
- Kappa measures the agreement of human labeling

HRED: Iulian V Serban et al., *Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models* , In AAAI'16
VHRED: Iulian V Serban et al., *A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues,* In AAAI'17

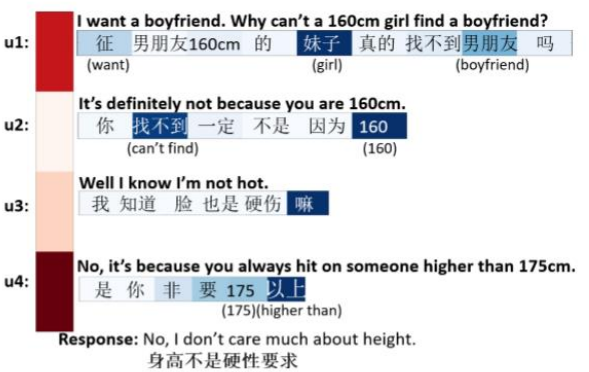
Case Study

ID	Context	Response
1	求唇部死皮怎么去⇒用化妆水一遍遍的擦嘴唇就能擦下来⇒ 好的回去试试⇒我是能擦下来 How to exfoliate my lips?⇒Wipe your lips with lotion again and again. ⇒OK, let me try it. ⇒It works for me.	HRAN: 你用的什么牌子的化妆水 Which brand of lotion are you using? VHRED:嗯我知道了 OK I see.
2	我不能去，你可以找丽丽陪你吃饭⇒她住松江，离市区太远了⇒ 有好吃的多远都要去啊⇒你为什么不能来呢 I can't have dinner with you, but you can ask Lily instead.⇒ She lives in Songjiang, too far away from downtown.⇒I think she'd like to go because of the delicious food.⇒ Why can't you come?	HRAN:吃多了上火了 Because I've eaten too much and gotten inflamed. VHRED:因为我是男的 Because I'm male.
3	征男友，160cm的妹子真的找不到男友吗⇒你找不到一定不是因为160⇒ 我知道脸也是硬伤嘛⇒是你非要175以上 I want a boyfriend. Why can't a 160cm girl find a boyfriend? ⇒It's definitely not because you are 160cm.⇒ Well I know I'm not good-looking. ⇒No, it's because you always hit on someone higher than 175cm.	HRAN:身高不是硬性要求 No, I don't care much about height. VHRED:你是男的还是女的啊 Are you male or female?
4	啦啦啦销售助理在哪里呀在哪里⇒学历不达标⇒ 对做销售有兴趣么⇒没做过不过我不挑工作 I'm looking for a sales assistent.⇒I might not be qualified because of my education.⇒Are you interested in sales? ⇒ I have no experience before but I'm open to any jobs.	HRAN:那你做过设计么 Then have you done any designing jobs before? VHRED:好吧好吧我也觉得 OK, I think so.

Visualization

(a) Visualization of case 1

(b) Visualization of case 2



(c) Visualization of case 3

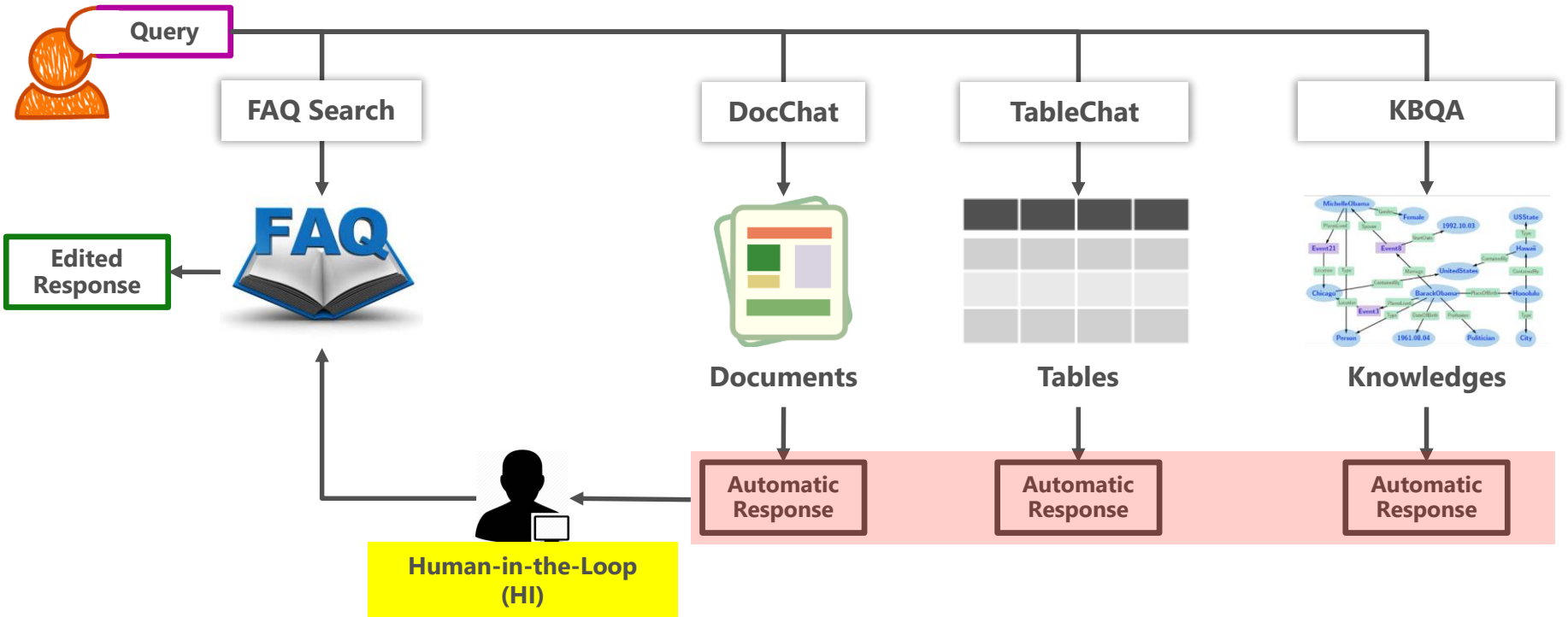
(d) Visualization of case 4

Information and Answer(InfoBot)

Microsoft

InfoBot Overview

Build conversational bots for information queries based on various genre of content and knowledge

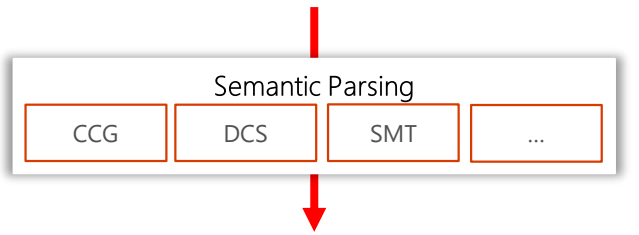


Knowledge-based QA (KB-QA)

CCG: Combinatory Categorial Grammar
DCS: Dependency-based Compositional Semantics
SMT: Statistical Machine Translation

Semantic Parsing-based KB-QA(SP-QA)

Query: where was Barack Obama born ?

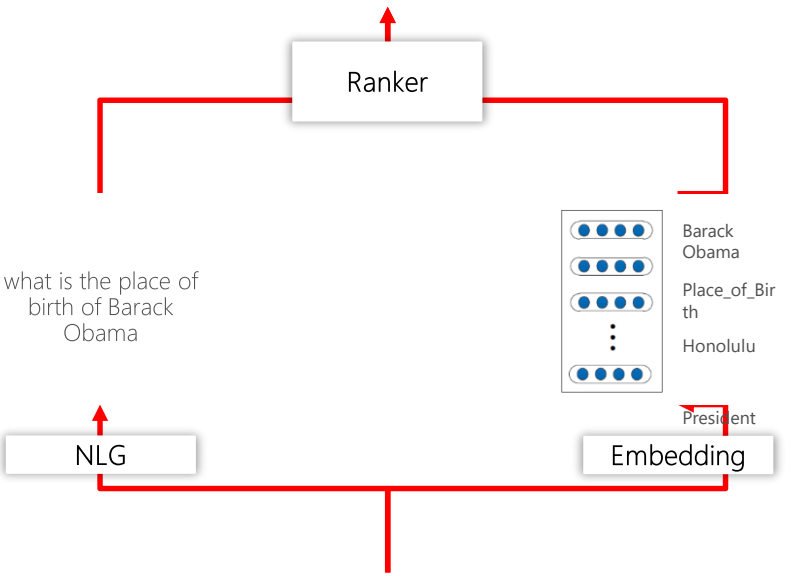


Logical Form: $\lambda x.Place\_of\_Birth(Barack\ Obama, x)$

<Barack Obama, Place_of_Birth, Honolulu>

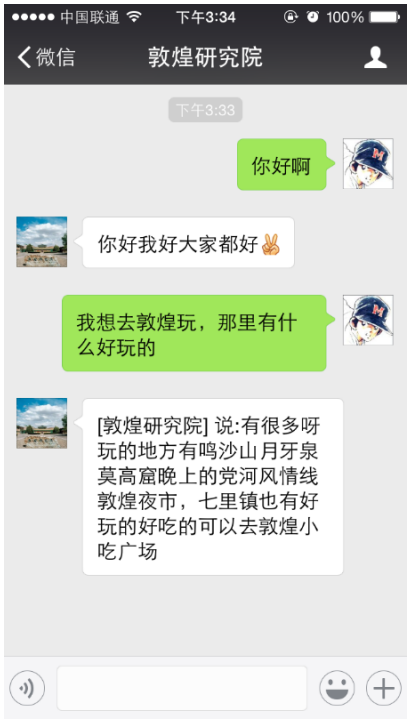
Information Retrieval-based KB-QA(IR-QA)

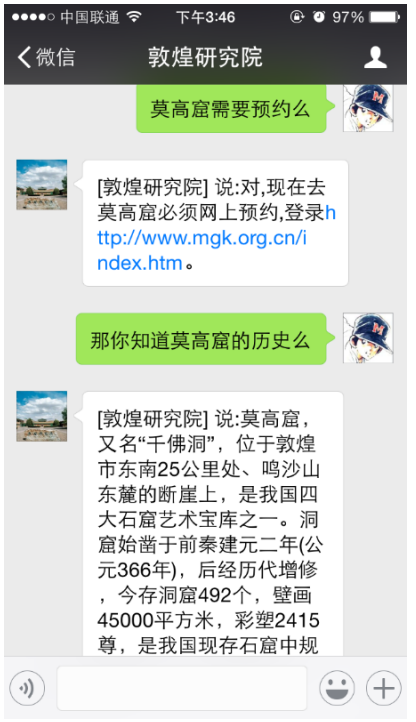
Query: where was Barack Obama born ?

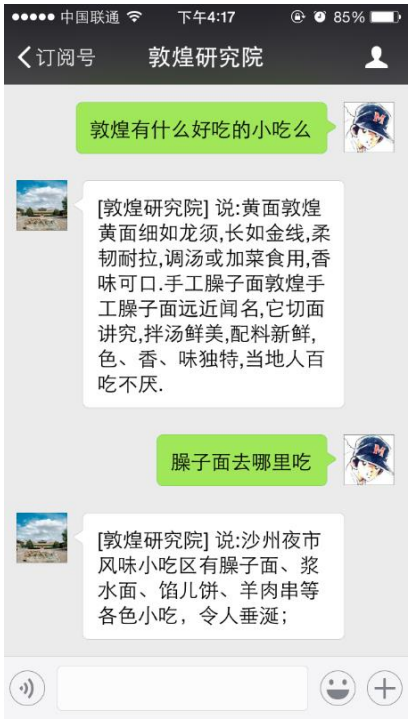


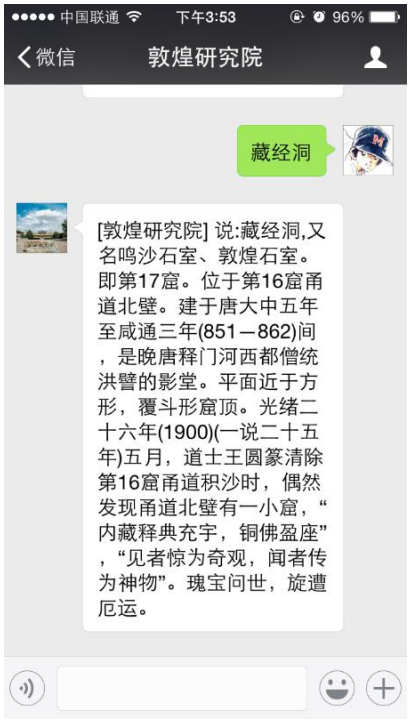
<Barack Obama, Place_of_Birth, Honolulu>

敦煌小冰


中国联通 下午3:34 100%
微信
敦煌研究院
下午3:33
你好啊
你好我好大家都好✌
我想去敦煌玩，那里有什么好玩的
[敦煌研究院] 说:有很多呀玩的地方有鸣沙山月牙泉莫高窟晚上的党河风情线敦煌夜市，七里镇也有好玩的好吃的可以去敦煌小吃广场


中国联通 下午3:46 97%
微信
敦煌研究院
莫高窟需要预约么
[敦煌研究院] 说:对,现在去莫高窟必须网上预约,登录http://www.mgk.org.cn/index.htm。
那你知道莫高窟的历史么
[敦煌研究院] 说:莫高窟，又名“千佛洞”，位于敦煌市东南25公里处、鸣沙山东麓的断崖上，是我国四大石窟艺术宝库之一。洞窟始凿于前秦建元二年(公元366年)，后经历代增修，今存洞窟492个，壁画45000平方米，彩塑2415尊，是我国现存石窟中规

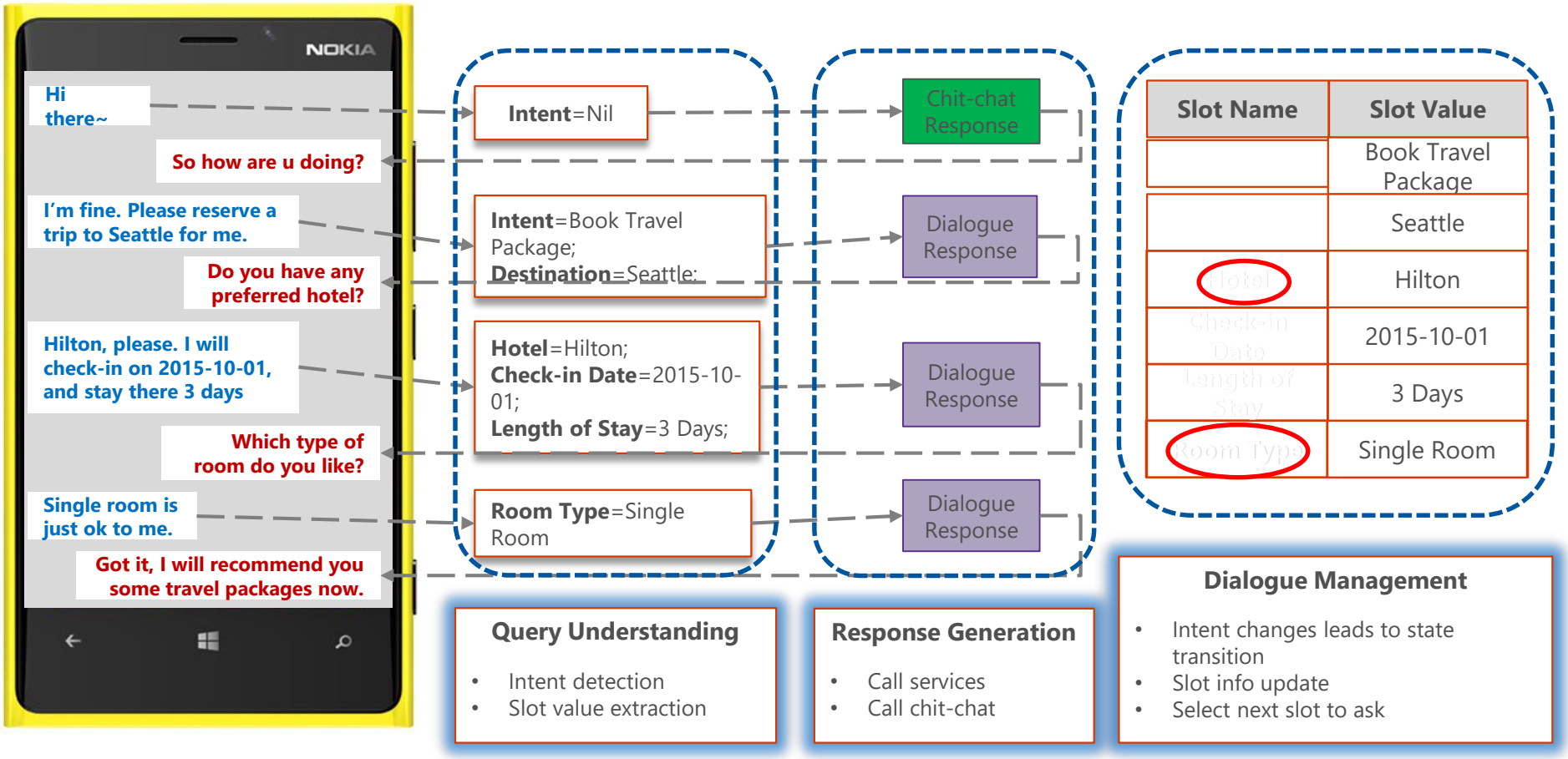

中国联通 下午4:17 85%
订阅号
敦煌研究院
敦煌有什么好吃的小吃么
[敦煌研究院] 说:黄面敦煌黄面细如龙须,长如金线,柔韧耐拉,调汤或加菜食用,香味可口.手工臊子面敦煌手工臊子面远近闻名,它切面讲究,拌汤鲜美,配料新鲜,色、香、味独特,当地人百吃不厌.
臊子面去哪里吃
[敦煌研究院] 说:沙州夜市风味小吃区有臊子面、浆水面、馅儿饼、羊肉串等各色小吃，令人垂涎；


中国联通 下午3:53 96%
微信
敦煌研究院
藏经洞
[敦煌研究院] 说:藏经洞,又名鸣沙石室、敦煌石室。即第17窟。位于第16窟甬道北壁。建于唐大中五年至咸通三年(851—862)间，是晚唐释门河西都僧统洪譬的影堂。平面近于方形，覆斗形窟顶。光绪二十六年(1900)(一说二十五年)五月，道士王圆箓清除第16窟甬道积沙时，偶然发现甬道北壁有一小窟，“内藏释典充宇，铜佛盈座”，“见者惊为奇观，闻者传为神物”。瑰宝问世，旋遭厄运。

Task-Oriented Dialogue System

Microsoft

Dialogue Process



Conclusion and Future Work

CAAP

- Chit-chat: Retrieval-based and generation-based, single turn and multi-turn, fusing with knowledge
- Infobot based on FAQ, Doc-Chat, KB-QA
- Task-Oriented dialogue

Future work

- Better modelling the context and multi-turn
- Fusing user (type) profile into response generation
- Response style transformation