QA系统

susht@cis

2017.09.11

问答系统的发展历程

Template-based QA

Expert System

IR-based QA

KB-based QA

Community QA

机器人问答

1960 **1990** **2000** **2010**

BaseBall

LUNAR

MACSYMA

SHRDLE

MASQUE

TREC

START

YAHOO! ANSWERS

ask. answer. discover.

搜狗问问 知乎

WolframAlpha computational knowledge engine

ProBase

问答系统的形式

- 一问一答

- 交互式问答

- 阅读理解

```
Mary journeyed to the den.
Mary went back to the kitchen.
John journeyed to the bedroom.
Mary discarded the milk.
Q: Where was the milk before the den?
A. Hallway

Brian is a lion.
Julius is a lion.
Julius is white.
Bernhard is green.
Q: What color is Brian?
A. White

Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom
```

论文讲解

- 阅读理解
  - 《Teaching Machines to Read and Comprehend》NIPS 2015
  - 《Text Understanding with the Attention Sum Reader Network》ACL 2016
  - 《Gated-Attention Readers for Text Comprehension》arXiv 2017
  - 《Answering Reading Comprehension Using Memory Networks》web.stanford.edu
  - 《End-To-End Memory Networks》NIPS 2015

- 开源问答系统
  - 《Reading Wikipedia to Answer Open-Domain Questions》ACL 2017
  - 《Dataset and Neural Recurrent Sequence Labeling Model for Open-Domain Factoid Question Answering》arXiv 2016

- 强化学习聊天机器人
  - 《Deep Reinforcement Learning for Dialogue Generation》

数据集

- SQuAD
- News Articles—CNN and Daily Mail
- Children’s Book Test
- Wiki QA
- bAbI QA
- MCTest dataset

《Teaching Machines to Read and Comprehend》

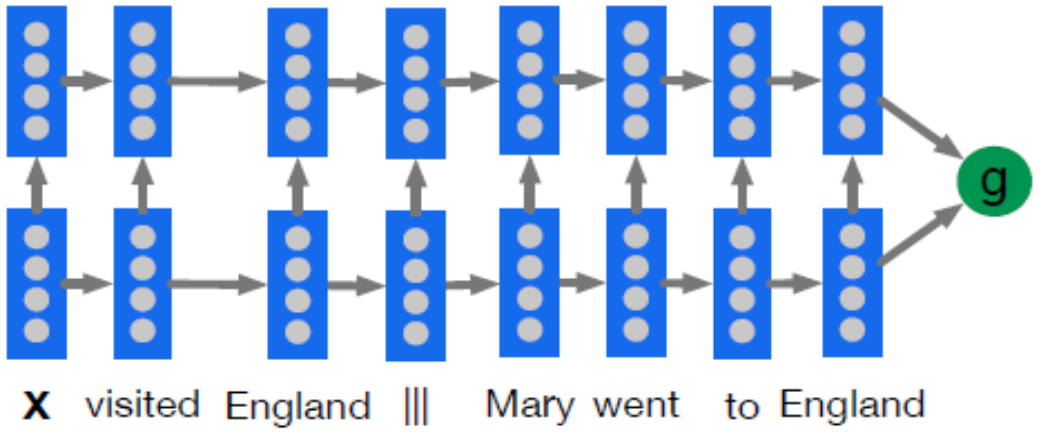
- 多分类任务
  - 三元组问题(a, d, q)，求解p(a|d,q)
  - 建模g(d, q)： Deep LSTM Reader / Attentive Reader/ Impatient Reader

as language modelling or machine translation [17]. We propose three neural models for estimating the probability of word type a from document d answering query q:

$$p(a|d,q) \propto \exp\left(W(a)g(d,q)\right), \quad \text{s.t.} \boxed{a \in V,}$$

where V is the vocabulary[4], and $W(a)$ indexes row a of weight matrix W and through a slight

Deep LSTM Reader



A two layer Deep LSTM Reader with the question encoded before the document.

Attentive Reader

- u

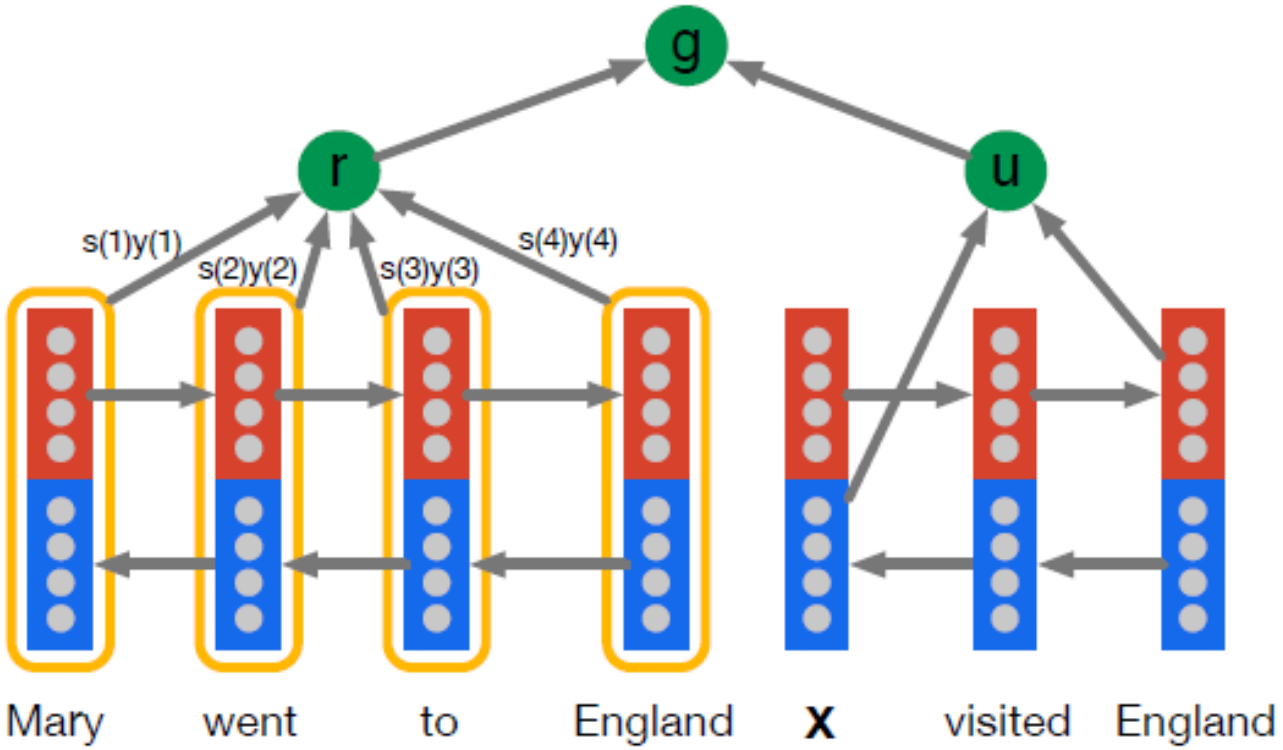
$$u = \overrightarrow{y_q}(|q|) \,||\, \overleftarrow{y_q}(1).$$

- r

$$m(t) = \tanh\left(W_{ym} y_d(t) + W_{um} u\right),$$
$$s(t) \propto \exp\left(\mathrm{w}_{ms}^{\intercal} m(t)\right),$$
$$r = y_d s,$$

- g

$$g^{\mathrm{AR}}(d, q) = \tanh\left(W_{rg} r + W_{ug} u\right)$$



(a) Attentive Reader.

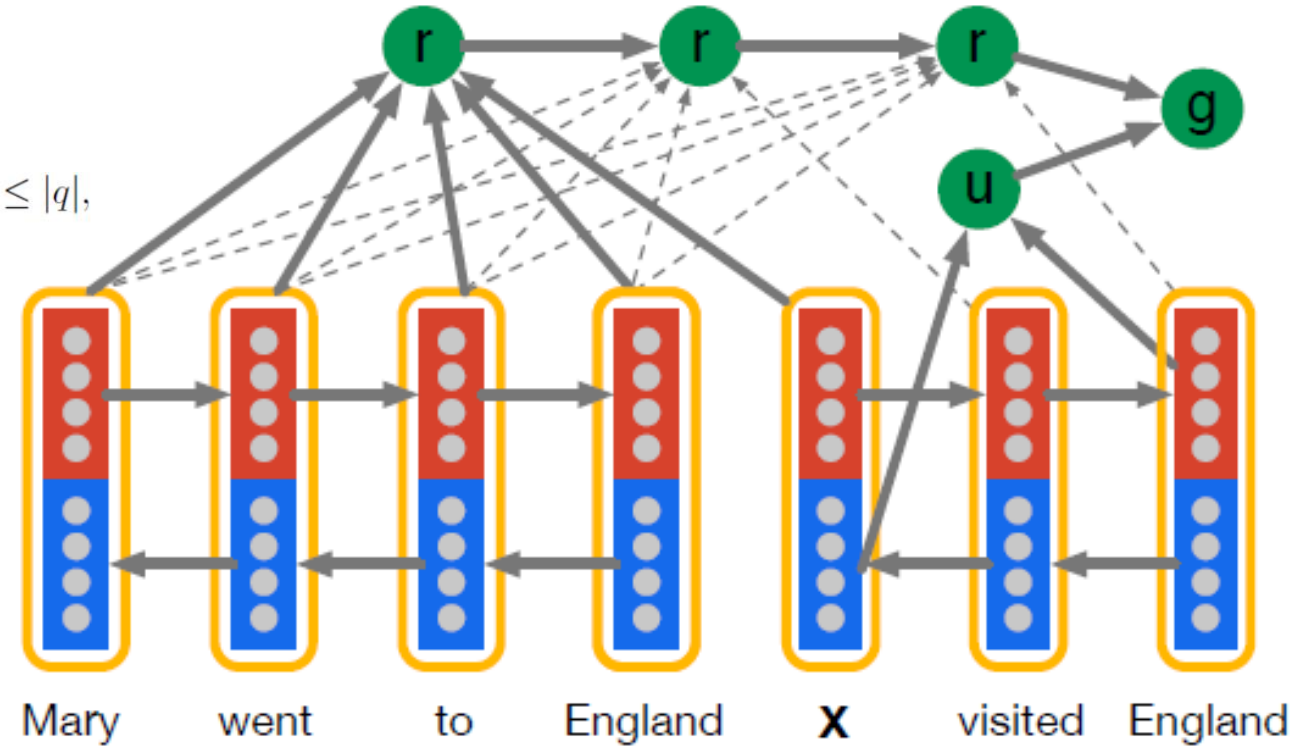
Impatient Reader

- 如何对y分配权重s

$$m(i,t) = \tanh\left(W_{dm}y_d(t) + W_{rm}r(i-1) + W_{qm}y_q(i)\right), \quad 1 \le i \le |q|,$$
$$s(i,t) \propto \exp\left(\mathrm{w}_{ms}^{\intercal} m(i,t)\right),$$

- 如何计算最后的r

$$r(0) = \mathbf{r_0}, \quad r(i) = y_d^{\intercal} s(i) + \tanh\left(W_{rr}r(i-1)\right) \quad 1 \le i \le |q|.$$



(b) Impatient Reader.

举例比较

- Q：国庆节是什么时候？

- Attentive Reader
  - 中华人民共和国国庆节又称十一、国庆节、国庆日、中国国庆节。中央人民政府宣布自1950年起，以每年的10月1日为国庆节。

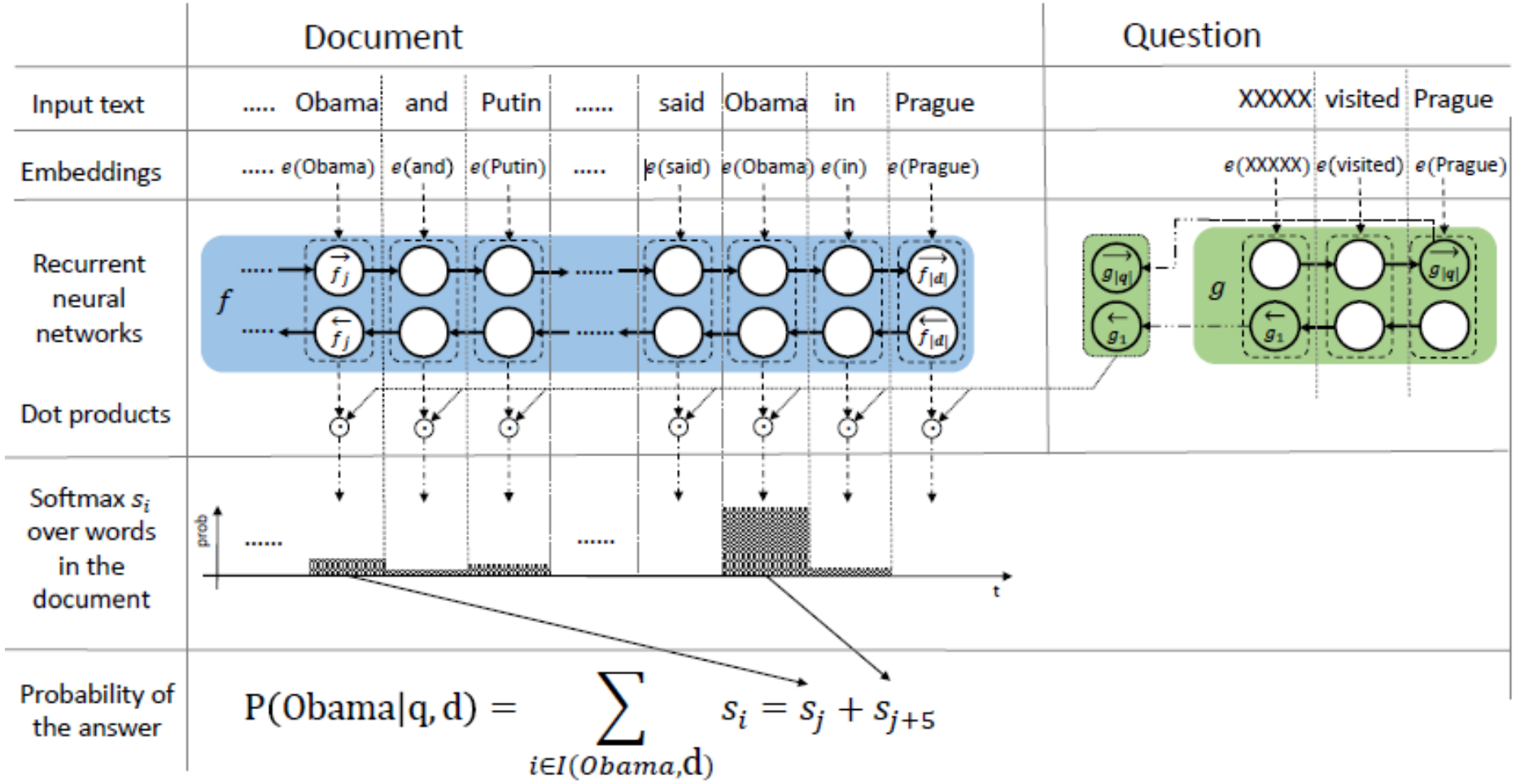
- Impatient Reader
  - 中华人民共和国国庆节又称十一、国庆节、国庆日、中国国庆节。中央人民政府宣布自1950年起，以每年的10月1日为国庆节。
  - 什么时候是国庆节？

AS Reader

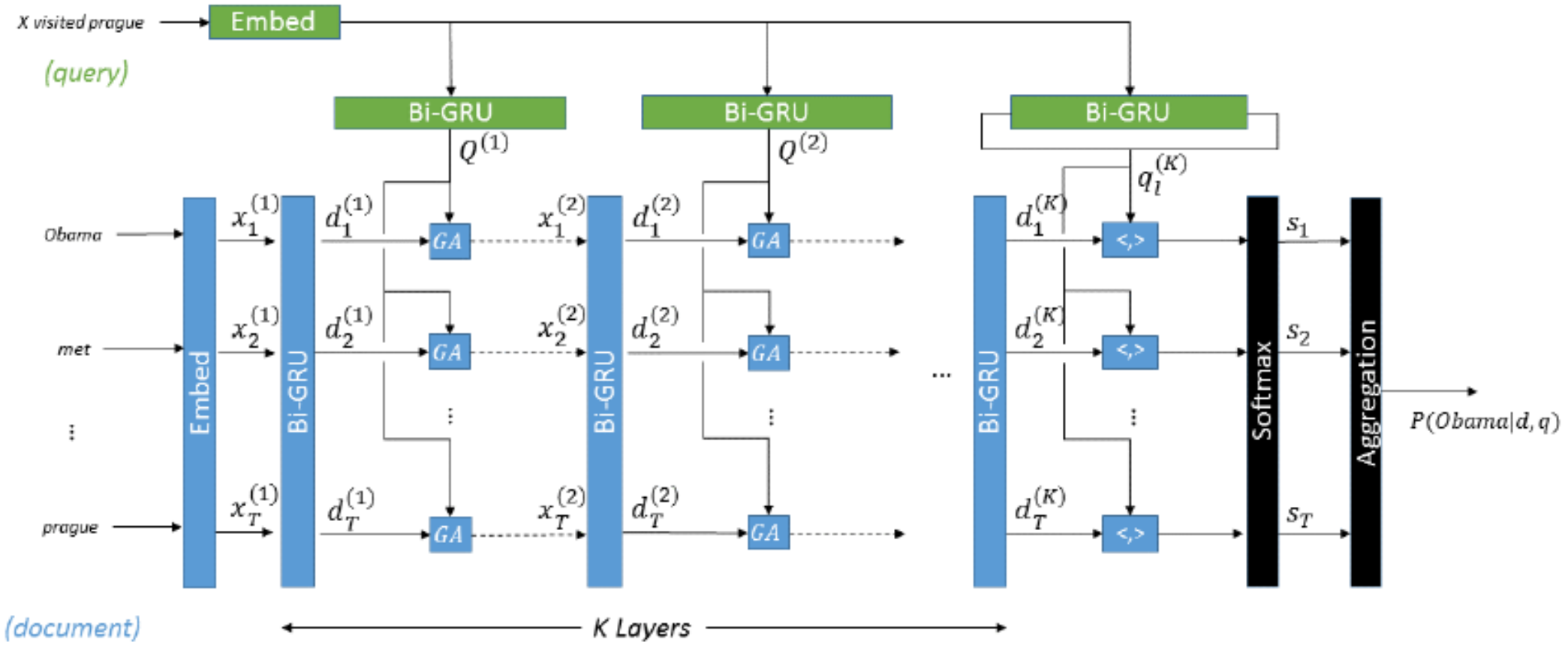
- 《Text Understanding with the Attention Sum Reader Network
  - Attention Sum Reader

- 《Gated-Attention Readers for Text Comprehension》
  - Gated-Attention Reader = Attention Sum Reader * n

Attention Sum Reader



Gated-Attention Reader



LSTM-Readers的比较

- Attentive Reader：使用简单的注意力机制
- Impatient Reader：使用比较复杂的注意力机制，并没有提升
- AS Reader：使用简单的点乘，达到较好的结果
- Gated-Attention Reader：使用多层结构，表现比单层结构更好

- Gated-Attention Reader > AS Reader > Attentive Reader > Impatient Reader
- 选择多层 简单 的结构

如何做到长期记忆？

- LSTM
  - 将state嵌入到低维空间，压缩知识
- Memory Network
  - 加入记忆模块，实现长期记忆
  - 加入推理功能

Sam walks into the kitchen.	Brian is a lion.	Mary journeyed to the den.
Sam picks up an apple.	Julius is a lion.	Mary went back to the kitchen.
Sam walks into the bedroom.	Julius is white.	John journeyed to the bedroom.
Sam drops the apple.	Bernhard is green.	Mary discarded the milk.
Q: Where is the apple?	Q: What color is Brian?	Q: Where was the milk before the den?
A. Bedroom	A. White	A. Hallway

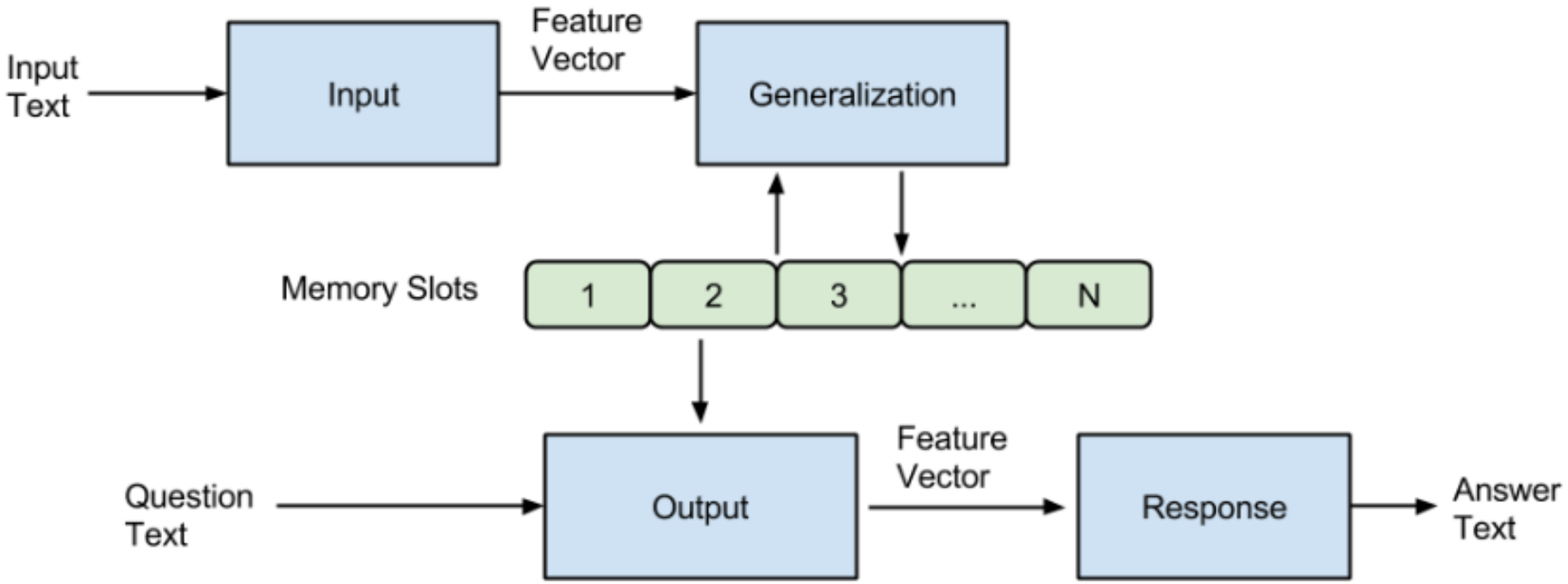
Memory Networks

- 《Answering Reading Comprehension Using Memory Networks》
  - Memory Network

- 《End-To-End Memory Networks》
  - Gated Memory Network

Memory Network Architecture

- Input -> (G + M) -> Output -> R



举例

Memory

Joe went to the kitchen.
Fred went to the kitchen.
Joe picked up the milk.
Joe travelled to the office.
Joe left the milk.
Joe went to the bathroom.

Input + Generalization

x = Where is the milk now? → **Question**

选择记忆Memory

$\mathbf{m}_{o1}$ = Joe left the milk.
$\mathbf{m}_{o2}$ = Joe travelled to the office.

Scoring memories

$$o_1 = O_1(x, \mathbf{m}) = \arg\max_{i=1,\ldots,N} s_O(x, \mathbf{m}_i)$$

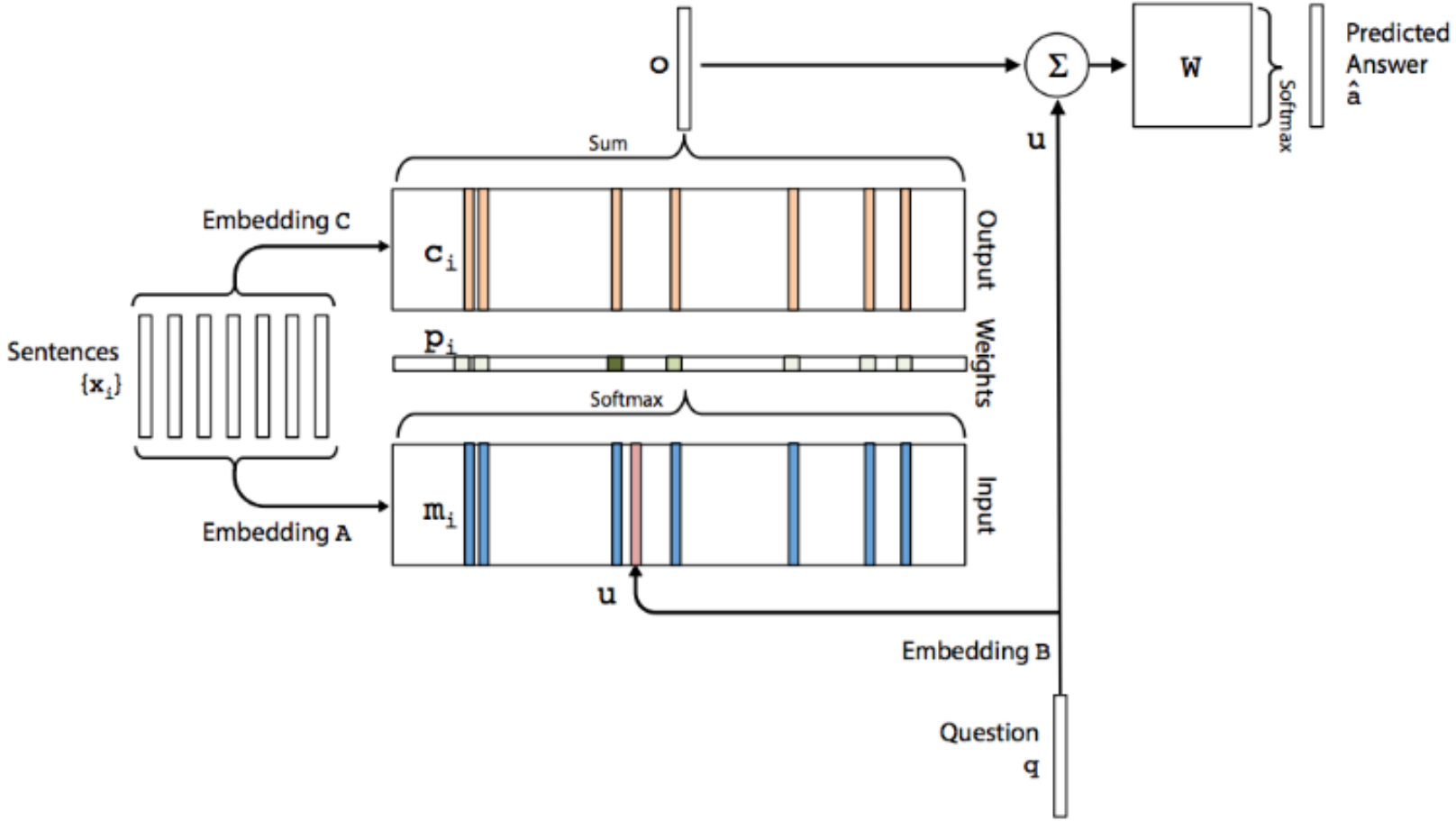
$$o_2 = O_2(x, \mathbf{m}) = \arg\max_{i=1,\ldots,N} s_O([x, \mathbf{m}_{o_1}], \mathbf{m}_i)$$

Scoring words：answer

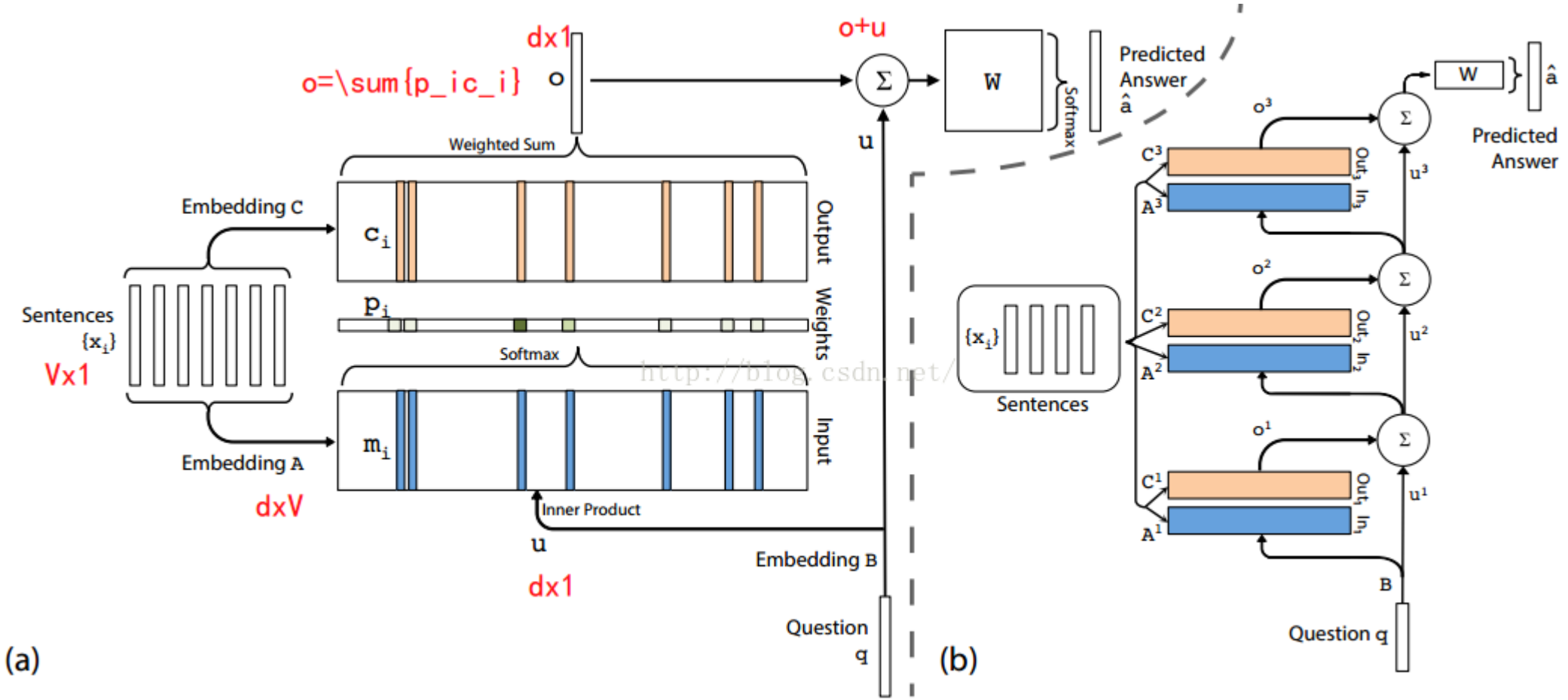
r = office →

$$r = \arg\max_{w \in W} s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], w)$$

Weakly Supervised Memory Network



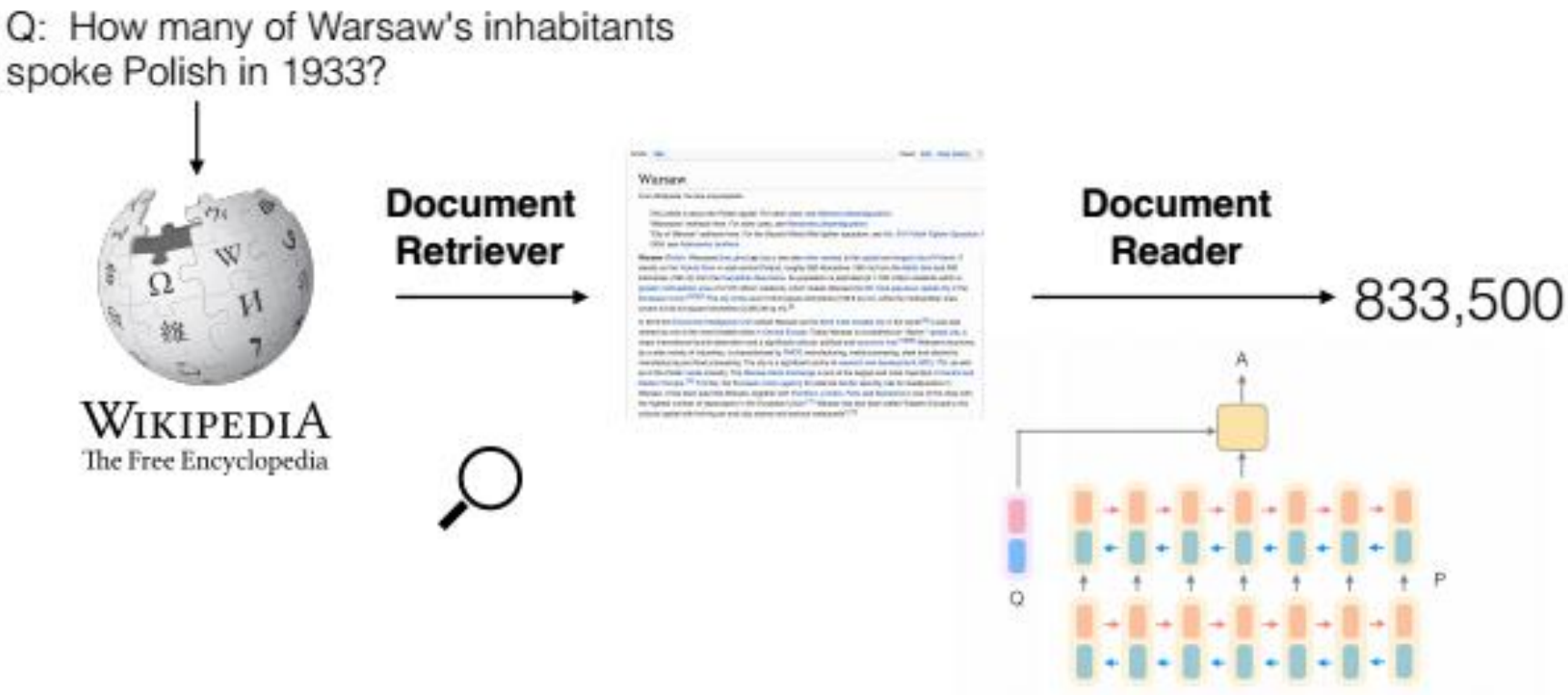
Gated Memory Networks



Model	CNN		Daily Mail		CBT-NE		CBT-CN	
	Valid	Test	Valid	Test	Valid	Test	Valid	Test
Human(query)	-	-	-	-	-	52.0	-	64.4
Human(context+query)	-	-	-	-	-	81.6	-	81.6
LSTM(context+query)	-	-	-	-	51.2	41.8	62.6	56.0
Deep LSTM Reader	55.0	57.0	63.3	62.2	-	-	-	-
Attentive Reader	61.6	63.0	70.5	69.0	-	-	-	-
Impatient Reader	61.8	63.8	69.0	68.0	-	-	-	-
Memory Networks(single)	63.4	66.8	-	-	70.4	66.6	64.2	63.0
Memory Networks(ensemble)	66.2	69.4	-	-	-	-	-	-
Attention Sum Reader(single)	68.6	69.5	75.0	73.9	73.8	68.6	68.8	63.4
Attention Sum Reader(ensemble)	73.9	75.4	78.7	77.7	76.2	71.0	71.1	68.9
Dynamic Entity Representation	71.3	72.9	-	-	-	-	-	-
Gate Attention Reader(single)	73.0	73.8	76.7	75.7	74.9	69.0	69.0	63.9
Gate Attention Reader(ensemble)	76.4	**77.4**	79.1	**78.1**	75.5	**71.9**	72.1	69.4

基于阅读理解做QA任务

- 文档检索 + 阅读理解

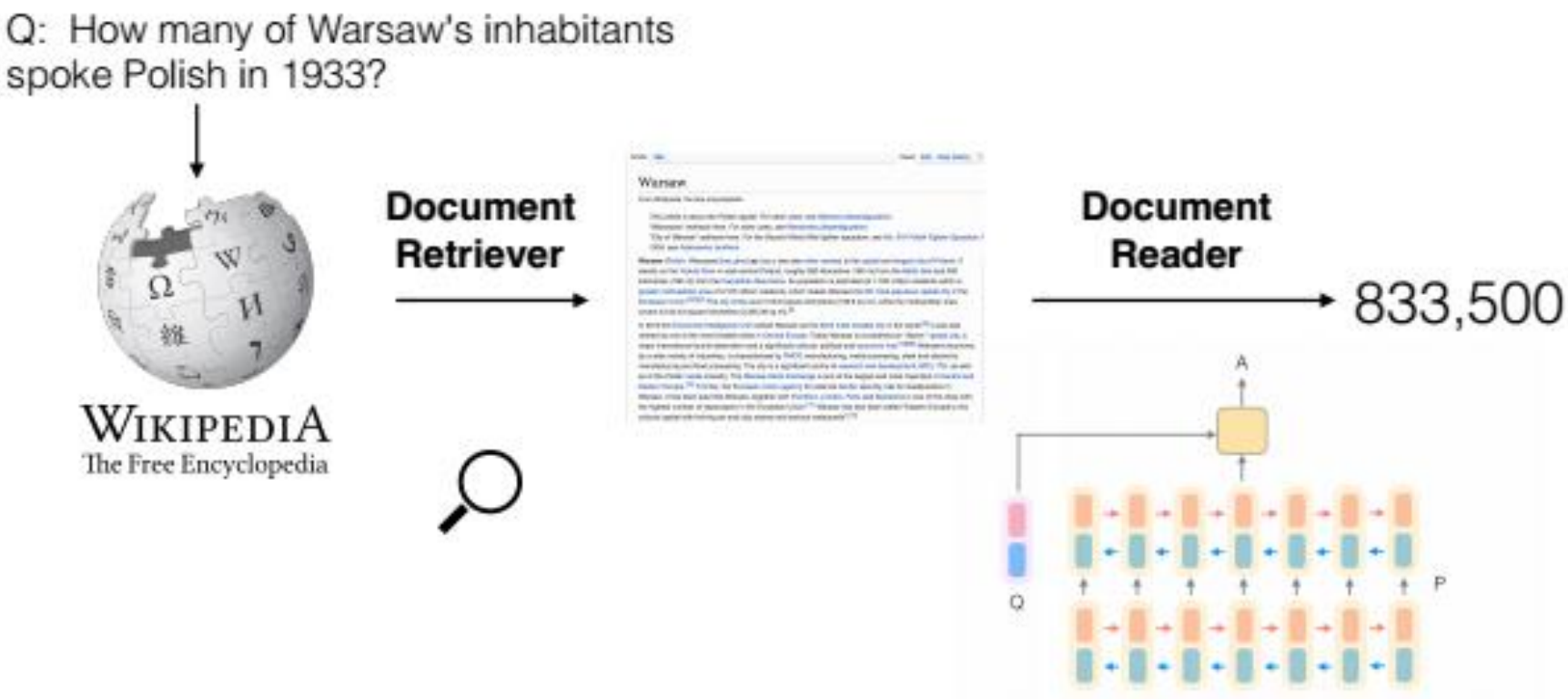


QA 系统

- 《Reading Wikipedia to Answer Open-Domain Questions》
  - 基于维基百科的DrQA
  - 分类任务

- 《Dataset and Neural Recurrent Sequence Labeling Model for Open-Domain Factoid Question Answering》
  - 基于百度知道的webQA
  - 序列标注任务

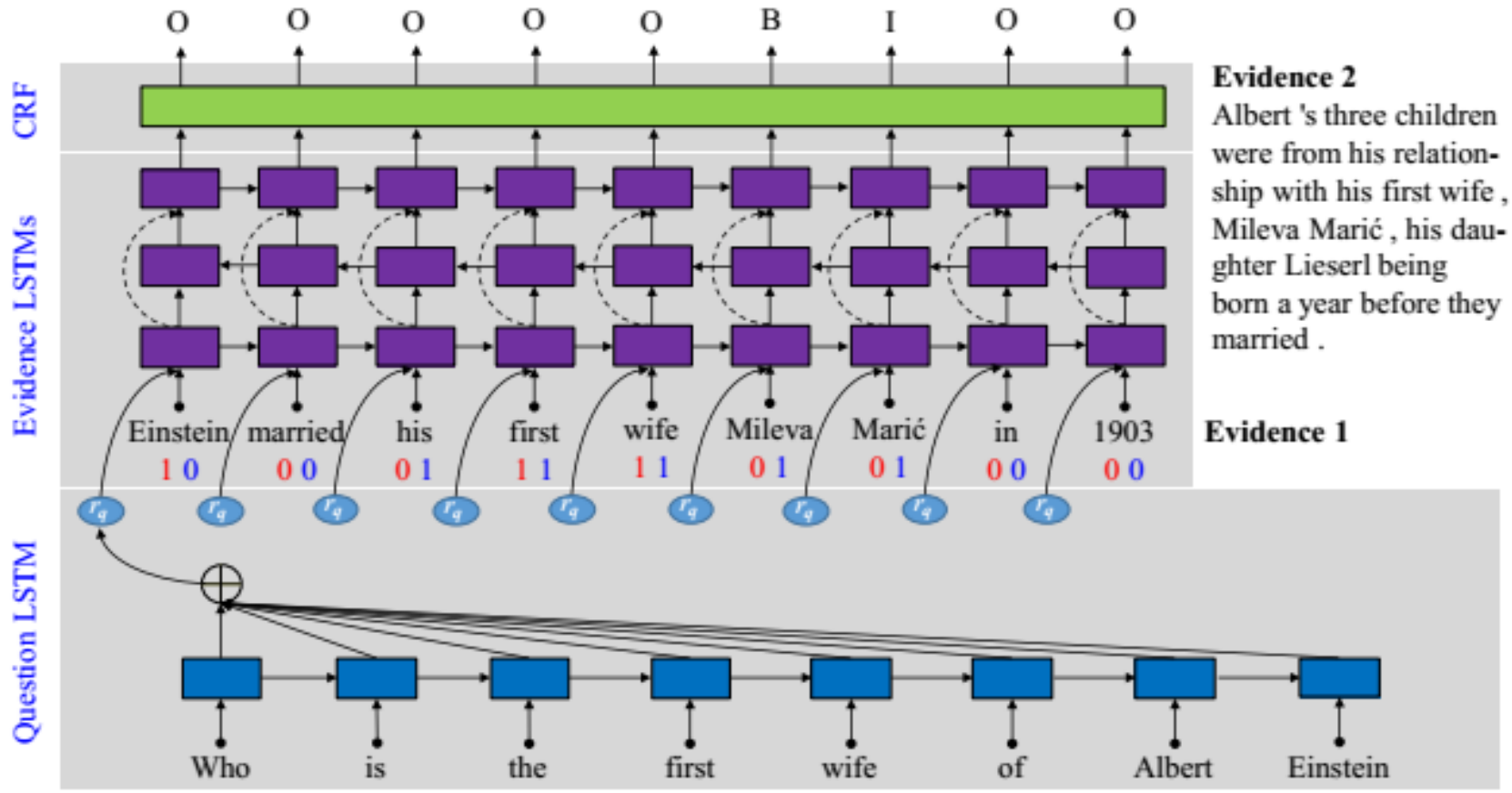
DrQA


Q: How many of Warsaw's inhabitants spoke Polish in 1933?
WIKIPEDIA
The Free Encyclopedia
Document Retriever
Warsaw
Document Reader
833,500
A
Q
P

Document Reader

- document建模
  - p = f(word embeddings) + f(exact match) + f(token features) + f(aligned question embedding)
- question建模
  - self attention rnn
- 预测
  - $P_{start(i)}$ ， $P_{end(i)}$
  - max($P_{start(i)}$ * $P_{end(j)}$)

WebQA


O O O O O B I O O
CRF
Evidence LSTMs
Einstein married his first wife Mileva Marić in 1903
1 0 0 0 0 1 1 1 1 1 0 1 0 1 0 0 0 0
Question LSTM
Who is the first wife of Albert Einstein
Evidence 2
Albert 's three children were from his relation-ship with his first wife , Mileva Marić , his dau-ghter Lieserl being born a year before they married .
Evidence 1

Experience

- Tricks
  - 动态loss
  - 变长LSTM + Attentive Reader
  - 基于字的模型，预测后进行分词补全

- 效果
  - 测试集：f1 = 0.68
  - 线上测试：f1 = 0.62

webQA

* 本系统的事实依据来源于 https://zhidao.baidu.com

春秋战国时期，法家的代表人是

搜索

随机

韩非子 (4)

韩非 (4)

关于阅读理解任务的思考

- 人类如何做阅读理解？
  - 从问题中找到关键词
  - 联系关键词所在的句子，联系上下文
  - 仔细理解相关句子最终得到答案

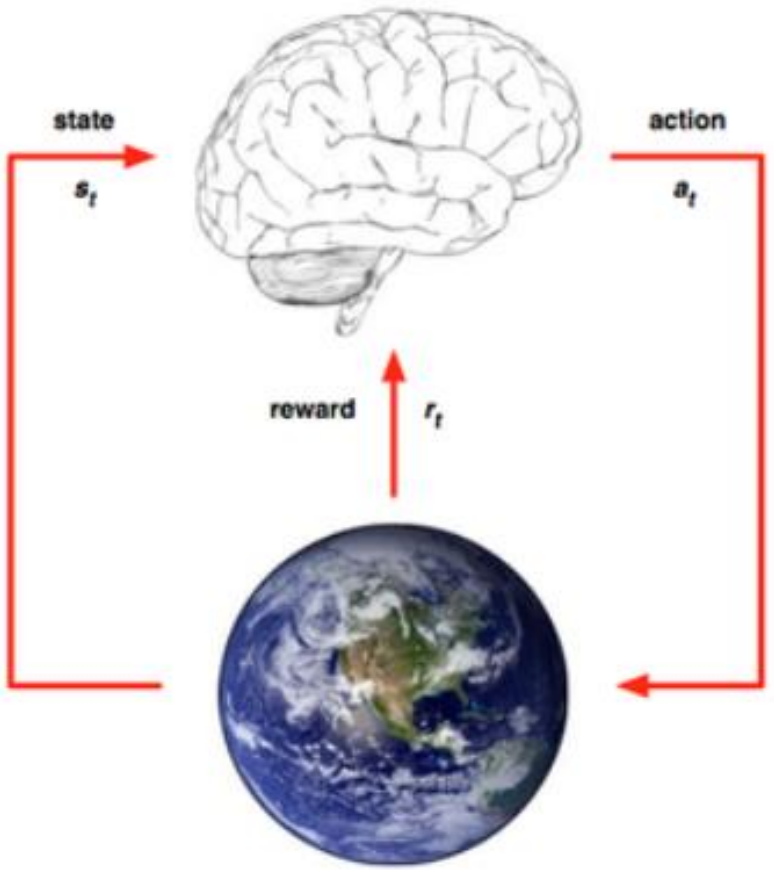
- Attention机制

基于强化学习的对话生成

- Seq2seq
  - 无意义答复 （呵呵 / 我不知道）
  - 对话死循环

- Seq2seq + RL
  - 《Deep Reinforcement Learning for Dialogue Generation》
  - 《A Deep Reinforcement Learning Chatbot》

强化学习四要素

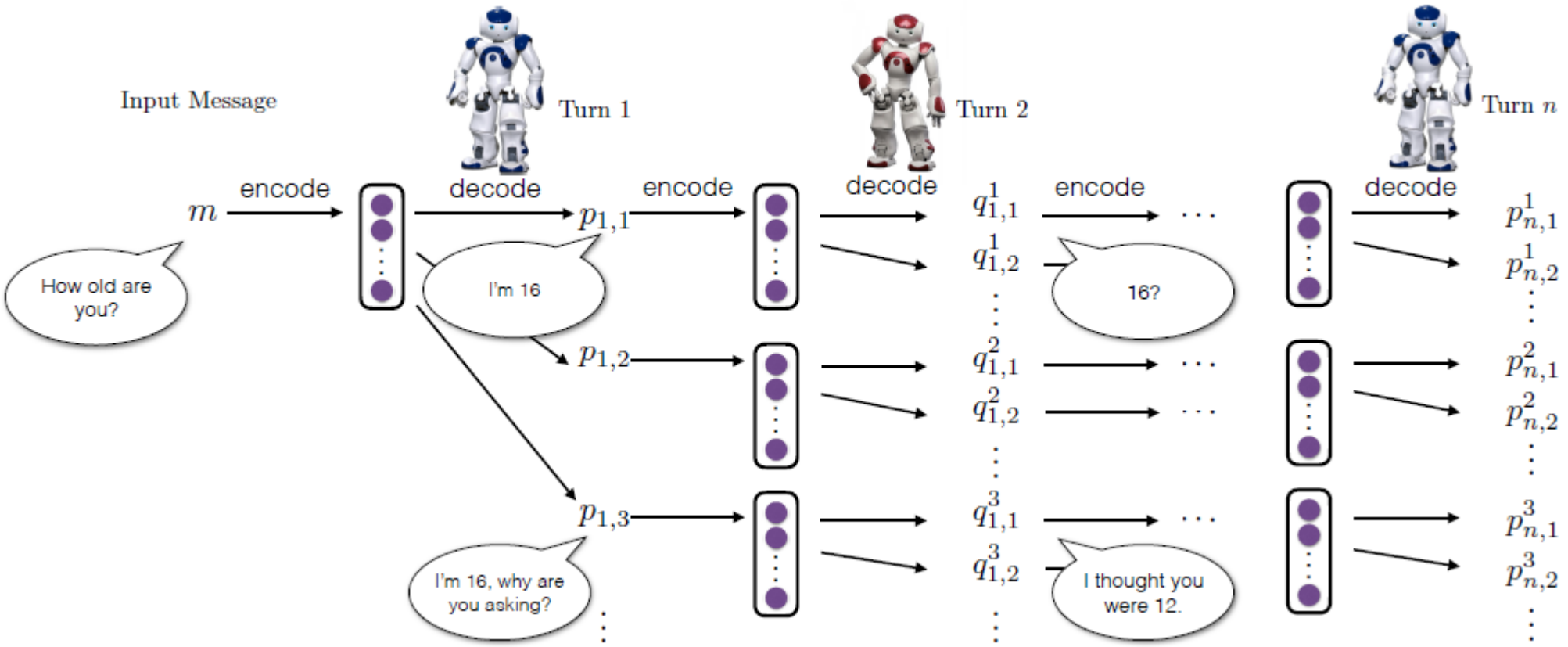


- At each step t the agent:
  - Receives state s_t
  - Receives scalar reward r_t
  - Executes action a_t
- The environment:
  - Receives action a_t
  - Emits state s_t
  - Emits scalar reward r_t

基于强化学习的对话生成

- Action
  - 这一轮生成的replay
- State
  - LSTM encoder，上一轮两个agent的对话表示[pi,qi]
- Policy
  - LSTM encoder-decoder，计算$p_{RL}(p_{i+1} \mid p_i; q_i)$
- Reward
  - Ease of Answering + Information Flow + Semantic Coherence
  - 判断replay是否有意义 / replay是否重复 / replay是否连贯

Dialogue simulation



实验结果

- 改进
  - 对话轮数增加
  - 对话多样性提高
  - 话题延续，提出相关问题

Input Message	Mutual Information Model	RL
How old are you ?	I'm 16.	I'm 16, why are you asking ?
What's your full name ?	I have no idea .	What's yours ?
I don't want to go home tonight.	Really ?	Why ?
Do you have any feelings for me ?	I don't know what you are talking about.	I don 't want to hurt your feelings .
How much time do you have here?	Not long enough. Sorry, sir.	Ten seconds .
Shall we get started ?	Of course !	Yes . We 've got a lot of work to do here .
Do you play football ?	No, i don't	Yes. I love football !
We'd have to talk to him.	I mean, he's a good guy	About what ?
How come you never say it?	Because I don't want to hurt you .	I don 't think it 's a good idea to say it .

Thanks