智能问答研究发展趋势

刘康

中国科学院自动化研究所

2016年10月26日

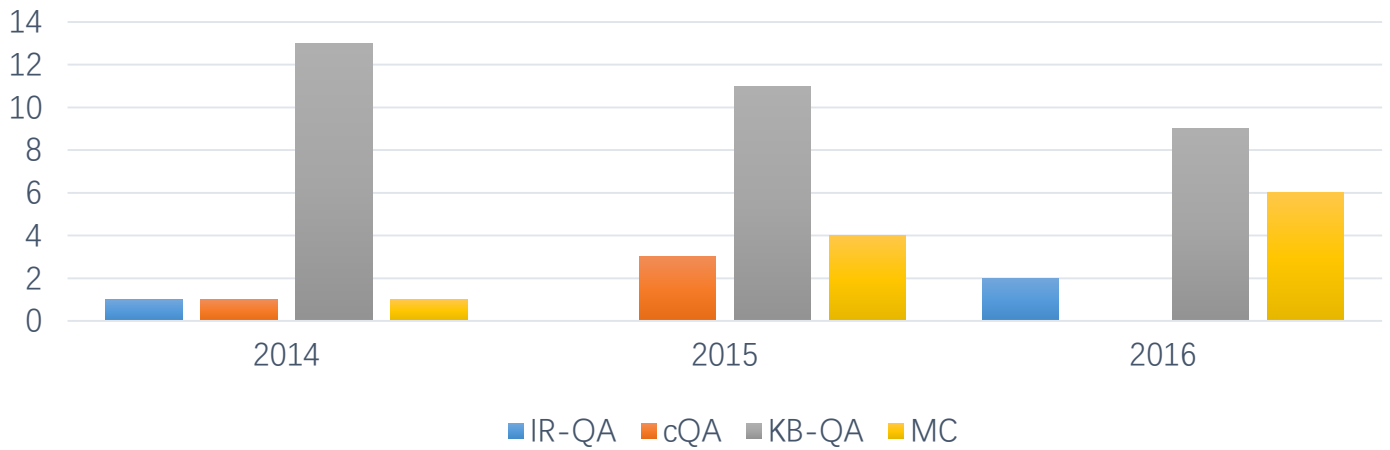
QA

检索式问答

社区问答

知识库问答

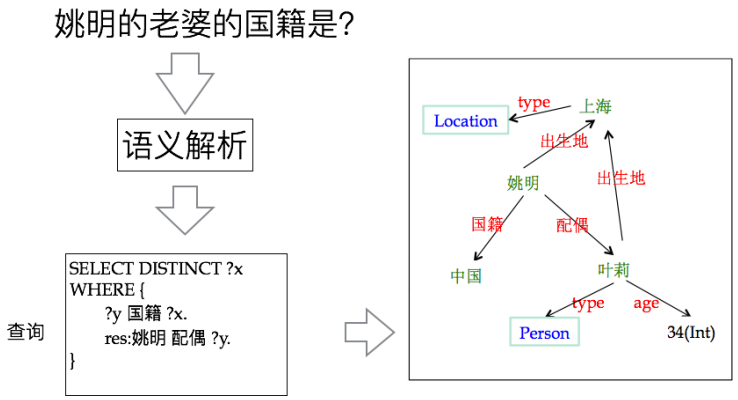
阅读理解



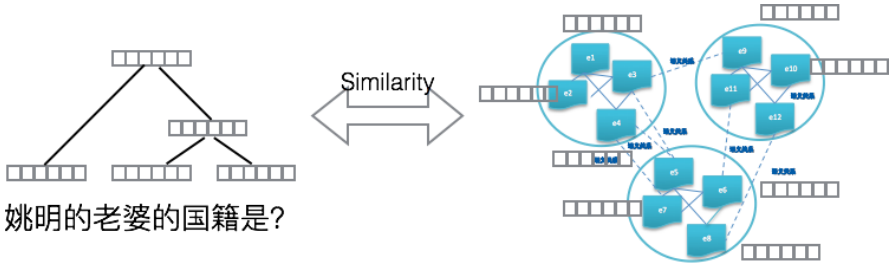
2014-2016
ACL、EMNLP论文统计

知识库问答

- Semantic Parsing （Symbolic Representation）

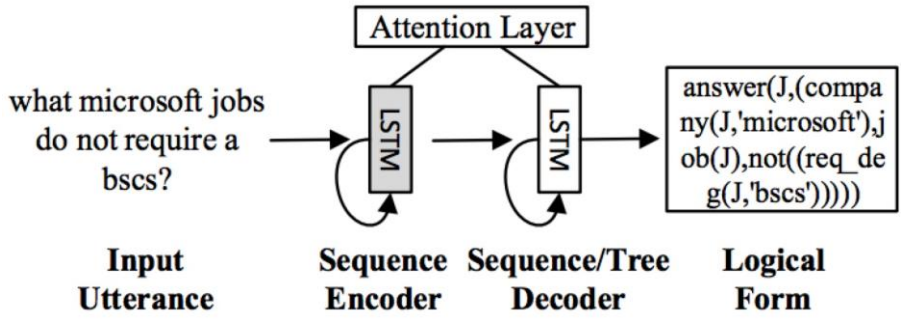


- IR based Approach: End2End QA based on Deep Learning (Distributed Representation)



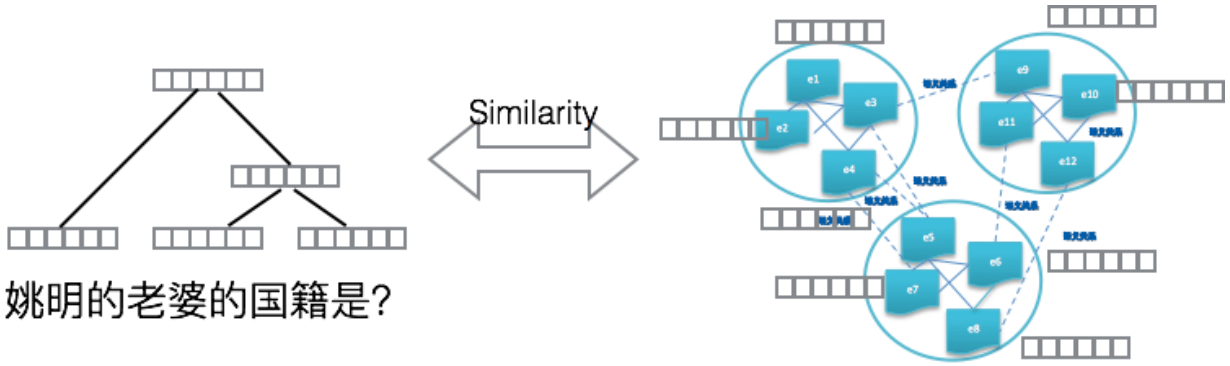
Semantic Parsing

- Sequence to Structure
  - 大规模、开放域，训练数据自动标注（Krishnamurthy, 2012， Fader 2013, Liang 2011, Jia et al. 2016 ）
  - 歧义问题（Chen et al. 2016, Xiao et al. 2016）
  - 面对网页表格（Janhar et al. 2016）
- Neural Semantic Parsing (Dong et al. 2016)



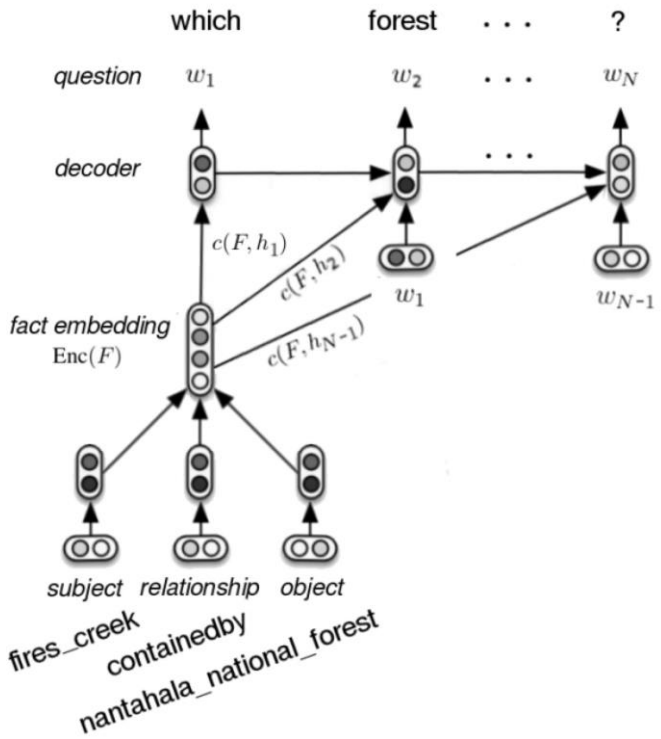
End2End QA based Deep Learning

- Simple Question (单关系问题)
  - 如何构建问句、候选答案的语义表示
    - 词平均（Borders et al. 2014）
    - CNN（Dong et al. 2015），Bi-LSTM（Zhang et al. 2016）
  - 如何学习映射函数
    - Cosine（Borders et al. 2014）
    - 类别、关系、上下文（Borders et al. 2014; Dong et al. 2015; Zhang et al. 2016）

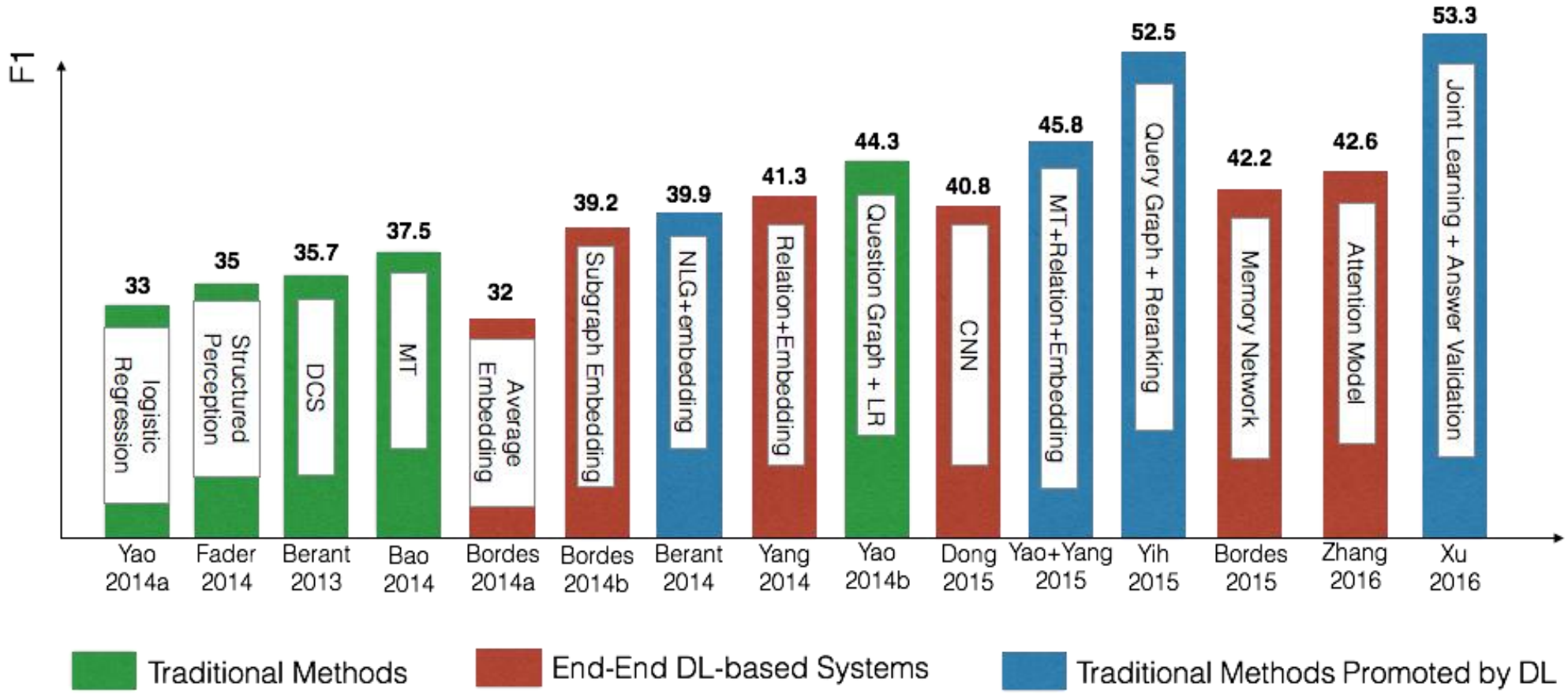


End2End QA based Deep Learning

- Simple Question (单关系问题)
  - NMT（Serban et al. 2016）



Comparisons on WebQuestion



小结

- 基于Semantic Parsing的知识库问答
  - 大规模、开放域
- 基于深度学习的知识库问答
  - 三元组的映射与抽取
  - 复杂问句

机器阅读理解

- 任务描述
  - 阅读理解
  - 完形填空

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.
One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.
His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

2) What did James pull off of the shelves in the grocery store?
A) pudding
B) fries
C) food
D) splinters

3) Where did James go after he went to the grocery store?
A) his deck
B) his freezer
C) a fast food restaurant
D) his room

数据集

- 从小规模数据集→大规模数据集
- 仿真数据集→真实数据集

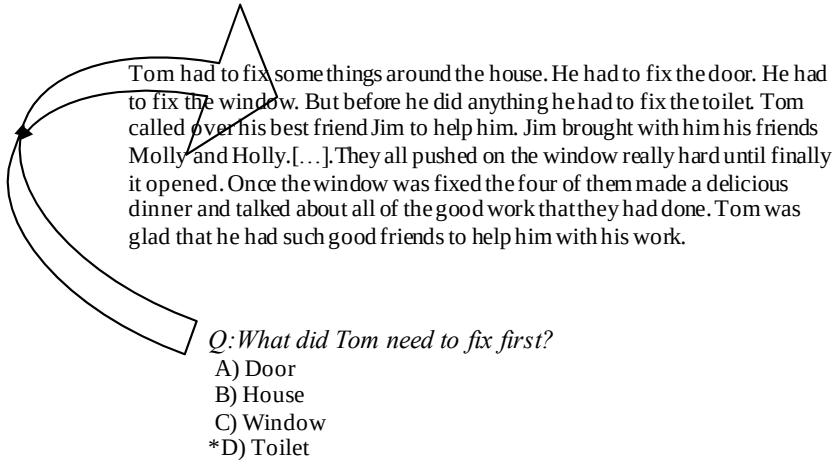
时间	数据集	描述	题型	大小
2015	AI2 Science Exams (UW)	美国科学小学考试（真实数据）	选择题	很小（百级）
2015	bAbi (Facebook)	考察20个方面文本理解和推理能力，简单的文本表达（模拟数据）	问答	很小（1000个训练、1000个测试、每个任务150个词）
2013	MCTest (Microsoft)	儿童读物（真实数据，众包）	选择题	很小（MC160：160篇；MC500：500篇）
2015	CNN/Daily Mail (Deep Mind)	新闻文本（真实数据，自动标注）	完形填空	较大（9w篇）
2016	CBTest (Facebook)	儿童读物（真实数据，自动标注）	完型填空	大（68w）
2016	Chinese RCD (iFlytek and HIT)	新闻文本儿童读物（真实数据，自动标注）	完形填空	大（87w）
2016	SQuAD (Stanford)	Wikipedia文本（真实数据，众包）	问答	较大（10w篇）
2016	Rochester Story	Short Commonsense Story（真实数据，众包）	选择题（句子级）	中等（1w篇）

方法

- Feature Engineering

$$p(a|q,D) = \underbrace{p(S|q,D)}_{\textbf{\textit{AS}}}\underbrace{p(a|q,S)}_{\textbf{\textit{RTE}}}$$

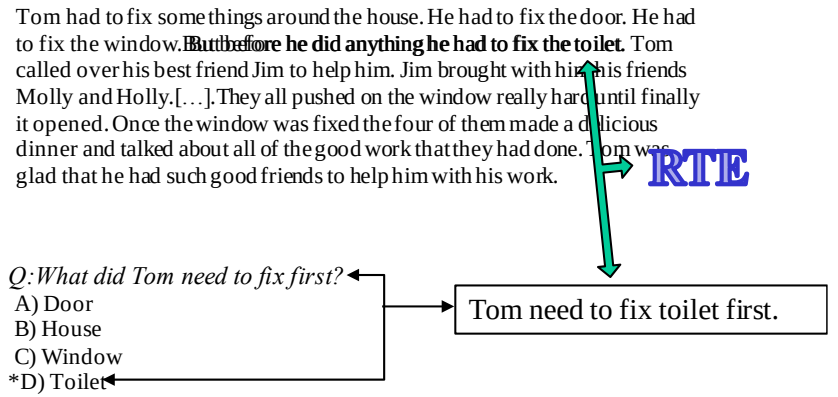
Step1：Answer Selection



Features:

- Alignments among words (Sachan, et al. ACL 2015)
- Discourse Relation: Causal, Temporal, Explanation (Narasimhan et al. ACL 2015)
- Linguistic Features: POS, Dependency, Word Embeddings, Frames (Wang et al. ACL 2015)

Step2：Inference

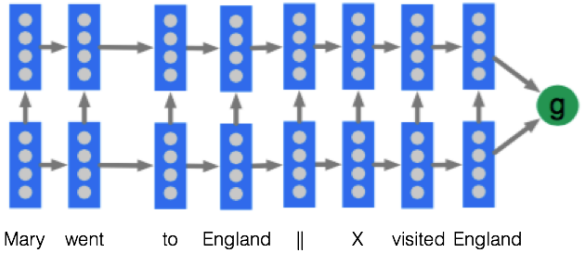


方法

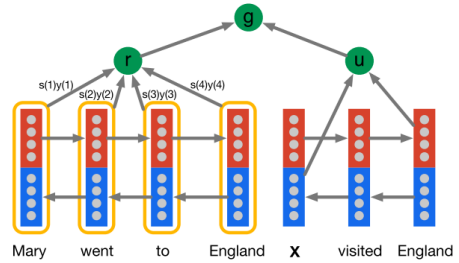
- Deep Learning

$$p(a|d,q) \propto \exp\left(W(a)g(d,q)\right)$$

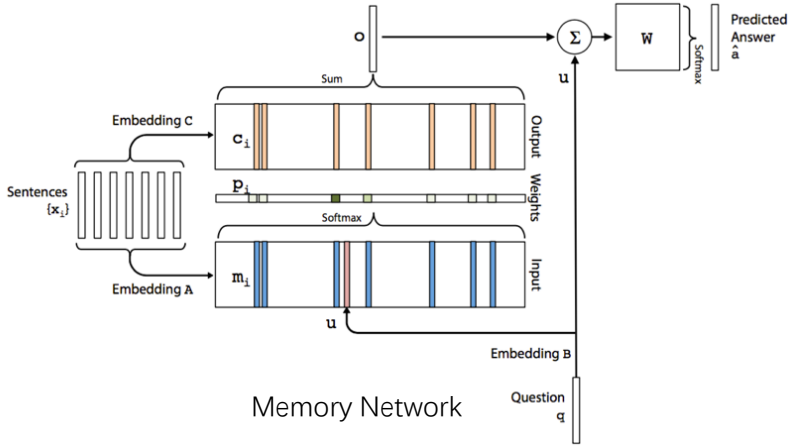
Representation Learning



LSTM



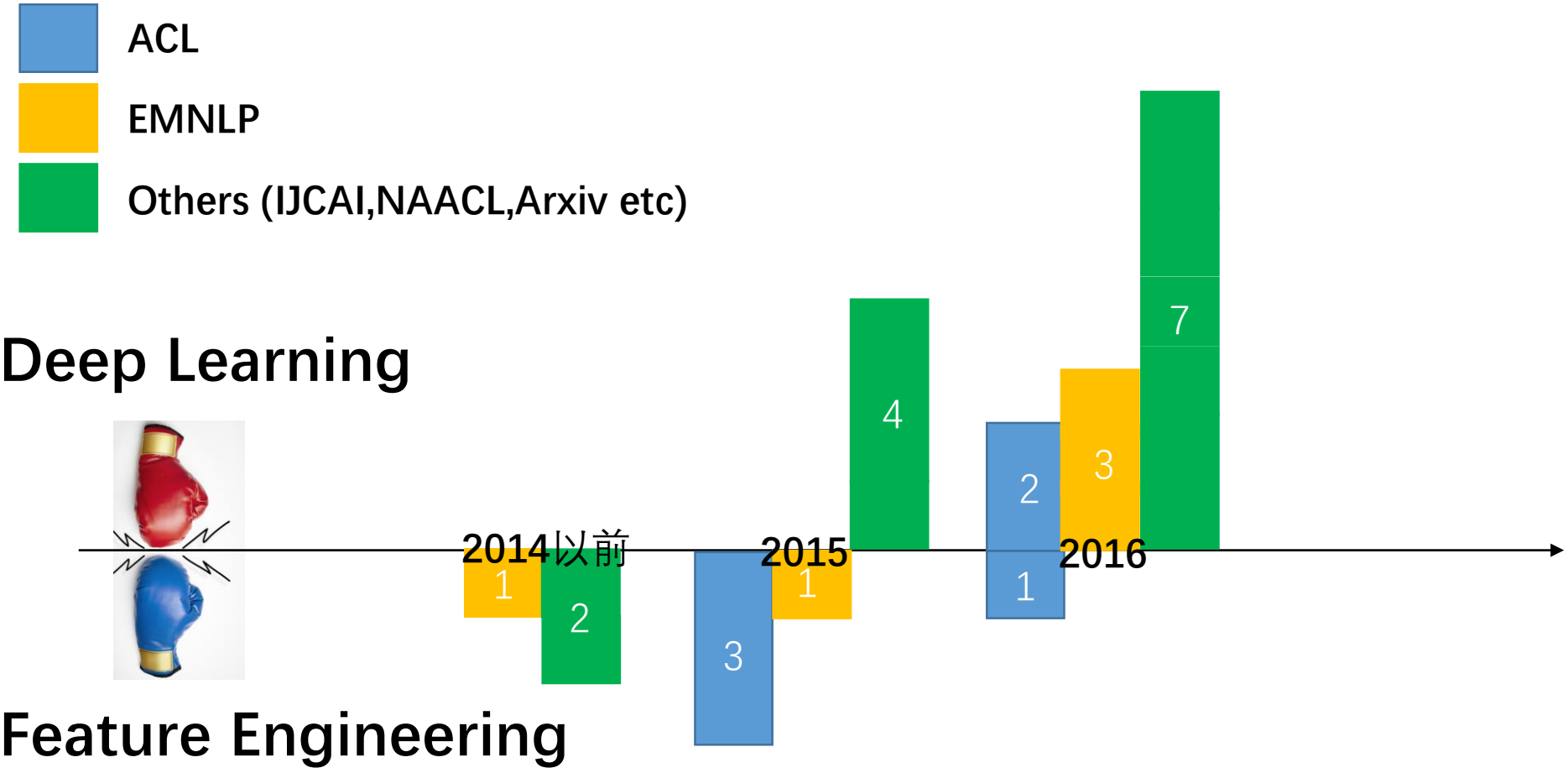
Attention



Memory Network

LSTM (Hermann, et al. 2015)
Attention based Model (Chen, et al. 2016, Hermann, et al. 2015)
Memory Network (Weston, et al. 2015)
Hierarchical CNN (Yin, et al. 2015)
Hierarchical Attention (Fang, et al. 2016)
Neural Turing Machine (Graves, et al. 2014)

两类方法比较


ACL
EMNLP
Others (IJCAI,NAACL,Arxiv etc)
Deep Learning
2014以前
2015
2016
1
2
3
1
4
1
2
3
7
Feature Engineering

不同类型问题细致化的分析

- 问题的类别

No.	Category	(%)
1	Exact match	13
2	Paraphrasing	41
3	Partial clue	19
4	Multiple sentences	2
5	Coreference errors	8
6	Ambiguous / hard	17

问题类别

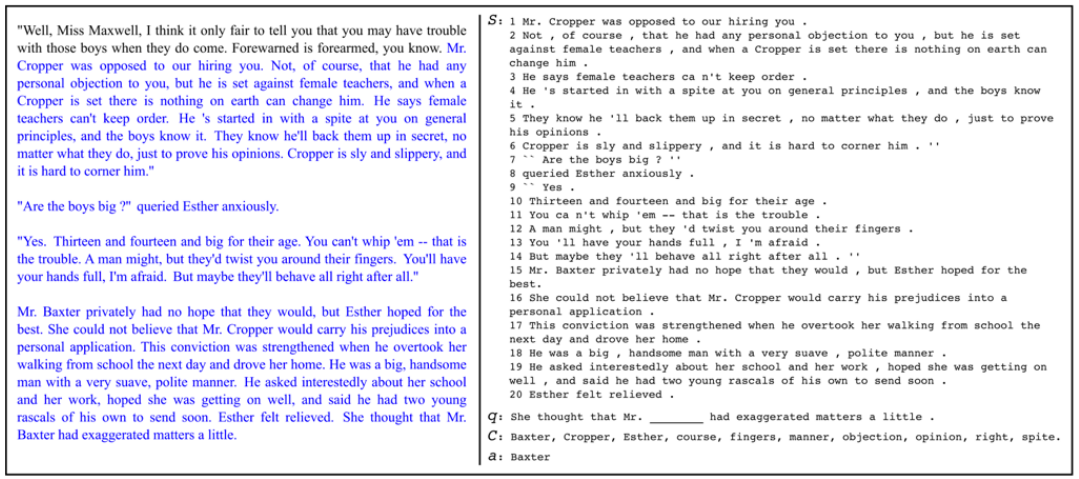
Category	Classifier	Neural net
Exact match	13 (100.0%)	13 (100.0%)
Paraphrasing	32 (78.1%)	39 (95.1%)
Partial clue	14 (73.7%)	17 (89.5%)
Multiple sentences	1 (50.0%)	1 (50.0%)
Coreference errors	4 (50.0%)	3 (37.5%)
Ambiguous / hard	2 (11.8%)	1 (5.9%)
All	66 (66.0%)	74 (74.0%)

结果

简单的Attention Model似乎已经可以达到性能的”天花板“
但是对于需要深度推理的问题，效果仍然不好

数据规模的影响（Bajgar et al. 2016）IBM

- 按照CBTest数据集的方式构建大数据集



"Well, Miss Maxwell, I think it only fair to tell you that you may have trouble with those boys when they do come. Forewarned is forearmed, you know. Mr. Cropper was opposed to our hiring you. Not, of course, that he had any personal objection to you, but he is set against female teachers, and when a Cropper is set there is nothing on earth can change him. He says female teachers can't keep order. He 's started in with a spite at you on general principles, and the boys know it. They know he'll back them up in secret, no matter what they do, just to prove his opinions. Cropper is sly and slippery, and it is hard to corner him."

"Are the boys big ?" queried Esther anxiously.

"Yes. Thirteen and fourteen and big for their age. You can't whip 'em -- that is the trouble. A man might, but they'd twist you around their fingers. You'll have your hands full, I'm afraid. But maybe they'll behave all right after all."

Mr. Baxter privately had no hope that they would, but Esther hoped for the best. She could not believe that Mr. Cropper would carry his prejudices into a personal application. This conviction was strengthened when he overtook her walking from school the next day and drove her home. He was a big, handsome man with a very suave, polite manner. He asked interestedly about her school and her work, hoped she was getting on well, and said he had two young rascals of his own to send soon. Esther felt relieved. She thought that Mr. Baxter had exaggerated matters a little.

S: 1 Mr. Cropper was opposed to our hiring you .
2 Not , of course , that he had any personal objection to you , but he is set against female teachers , and when a Cropper is set there is nothing on earth can change him .
3 He says female teachers ca n't keep order .
4 He 's started in with a spite at you on general principles , and the boys know it .
5 They know he 'll back them up in secret , no matter what they do , just to prove his opinions .
6 Cropper is sly and slippery , and it is hard to corner him . ''
7 `` Are the boys big ? ''
8 queried Esther anxiously .
9 `` Yes .
10 Thirteen and fourteen and big for their age .
11 You ca n't whip 'em -- that is the trouble .
12 A man might , but they 'd twist you around their fingers .
13 You 'll have your hands full , I 'm afraid .
14 But maybe they 'll behave all right after all . ''
15 Mr. Baxter privately had no hope that they would , but Esther hoped for the best.
16 She could not believe that Mr. Cropper would carry his prejudices into a personal application .
17 This conviction was strengthened when he overtook her walking from school the next day and drove her home .
18 He was a big , handsome man with a very suave , polite manner .
19 He asked interestedly about her school and her work , hoped she was getting on well , and said he had two young rascals of his own to send soon .
20 Esther felt relieved .

Q: She thought that Mr. _____ had exaggerated matters a little .

C: Baxter, Cropper, Esther, course, fingers, manner, objection, opinion, right, spite.

a: Baxter

	CNN	Daily Mail	CBT CN	CBT NE	BookTest
# queries	380,298	879,450	120,769	108,719	14,140,825
Max # options	527	371	10	10	10
Avg # options	26.4	26.5	10	10	10
Avg # tokens	762	813	470	433	522
Vocab. size	118,497	208,045	53,185	53,063	1,860,394

语料规模较CBT增加了60倍

	Named entity		Common noun		
	valid	test	valid	test	
Humans (query) (Hill et al., 2015)	NA	52.0	NA	64.4	
Humans (context+query) (Hill et al., 2015)	NA	**81.6**	NA	**81.6**	
LSTMs (context+query) (Hill et al., 2015)	51.2	41.8	62.6	56.0	CBT training data
Memory Networks (Hill et al., 2015)	70.4	66.6	64.2	63.0	
AS Reader (sing				63.4	
AS Reader (avg				**68.9**	
AS Reader (gre				67.5	
GA Reader (ensemble) (Dhingra et al., 2016)	75.3	71.9	72.1	69.4	
EpiReader (ensemble) (Trischler et al., 2016b)	76.6	71.8	73.6	70.6	
IA Reader (ensemble) (Sordoni et al., 2016)	76.9	**72.0**	74.1	**71.0**	
AoA Reader (single model) (Cui et al., 2016a)	77.8	**72.0**	72.2	69.4	
AS Reader (single model)	80.5	76.2	83.2	80.8	BookTest training data
AS Reader (greedy ensemble)	82.3	**78.4**	85.7	**83.7**	

数据规模对于DL-based性能的影响极大的超过了模型本身

小结

- 数据：
  - 人工数据→真实数据
  - 类型多样：选择题、完形填空、生成题
  - 实体型问题→句子型问题
  - 大规模数据的自动生成

- 方法：基于Deep Learning的阅读理解方法
  - Representation Learning: 问题和篇章
  - Attention Model成为主流
  - 基于深度学习的推理模型

谢谢

请批评指正