Question Answering and Some Deep Learning Methods Involved

王石
中科院计算所 智能信息处理重点实验室
wangshi@ict.ac.cn

2

- Introduction
- Methods
- Products
- Looking forward

3

Introduction

Definition

- 传统上，问答系统是一种以自然语言为交互方式的信息检索手段，它接受自然语言形式的问题，输出唯一的精确或合适答案；近年来，问答系统的交互方式由自然语言进一步扩展到语音、图像、手势、表情等多种模式，输出也由自然语言形式的答案扩展到多媒体信息以及服务

- 本研究内容限定在传统的问答系统范围，主要关注自然语言处理部分

https://en.wikipedia.org/wiki/Question_answering

Keywords

personal assistant

human computer conversation

Dialog System

virtual assistant

chat bots

chat companion,

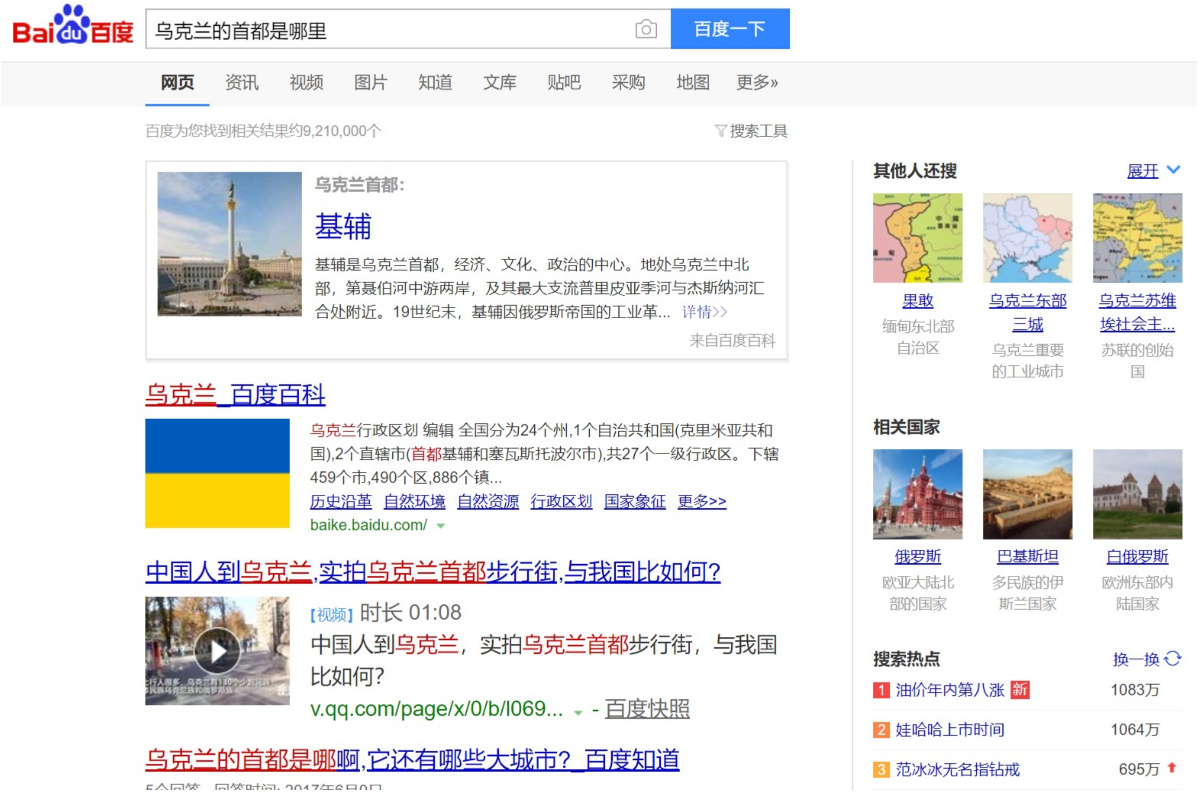
Conversation Systems

NLU

Semantic parsing

Motivation

- 提高效率、场景需求



QA in NLP

李航，今日头条AI Lab主任

曾任MSRA主任研究员、

华为诺亚方舟实验室主任

Fundamental Problems of Statistical Natural Language Processing

- Classification
  - Text classification
  - Sentiment analysis
- Matching
  - Search
  - Question answering
  - Dialogue (single turn)
- Translation
  - Machine translation
  - Speech recognition
  - Hand writing recognition
  - Dialogue (single turn)
- Structured Prediction
  - Named entity extraction
  - Part of speech tagging
  - Sentence parsing
  - Semantic parsing
- Markov Decision Process
  - Dialogue (multi turn, task dependent)

QA in NLP

语言模型	**问答系统**	机器阅读	内容推荐
情感分析	自动写作	机器翻译	自然语言推理
信息检索	信息抽取	推荐系统	知识图谱

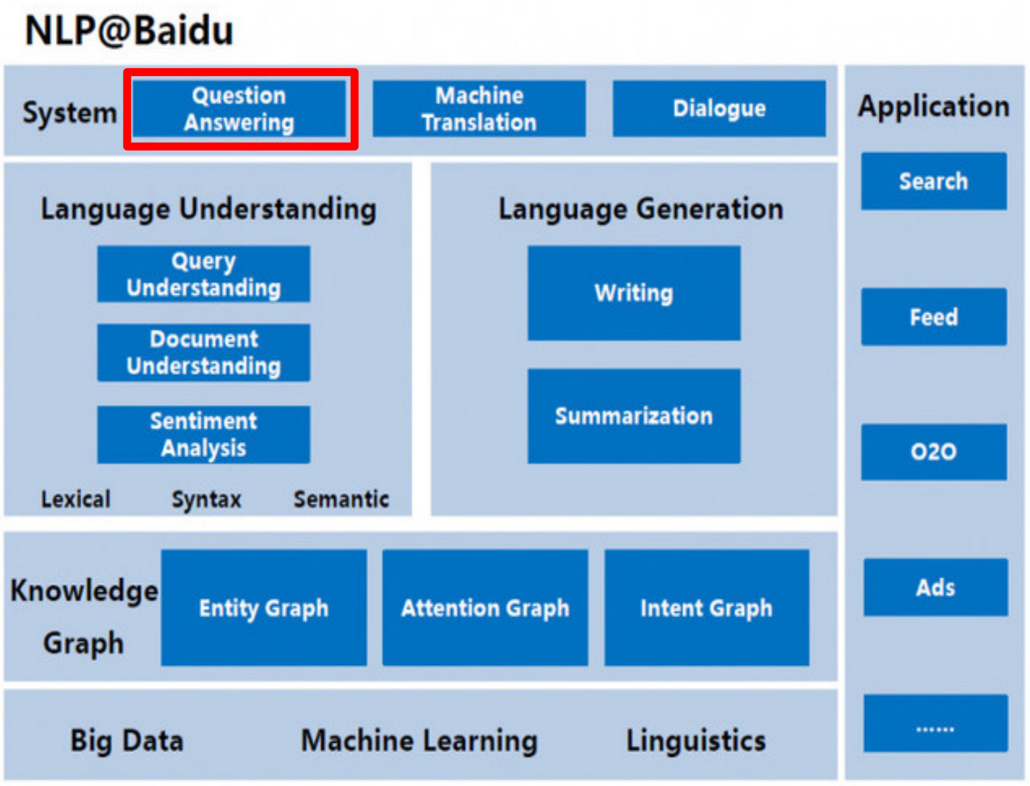
应用系统

语义：WSD、语义角色标注、向量化、指代消解、框架语义分析

句：组块分析、成分句法分析、依存句法分析、语言模型

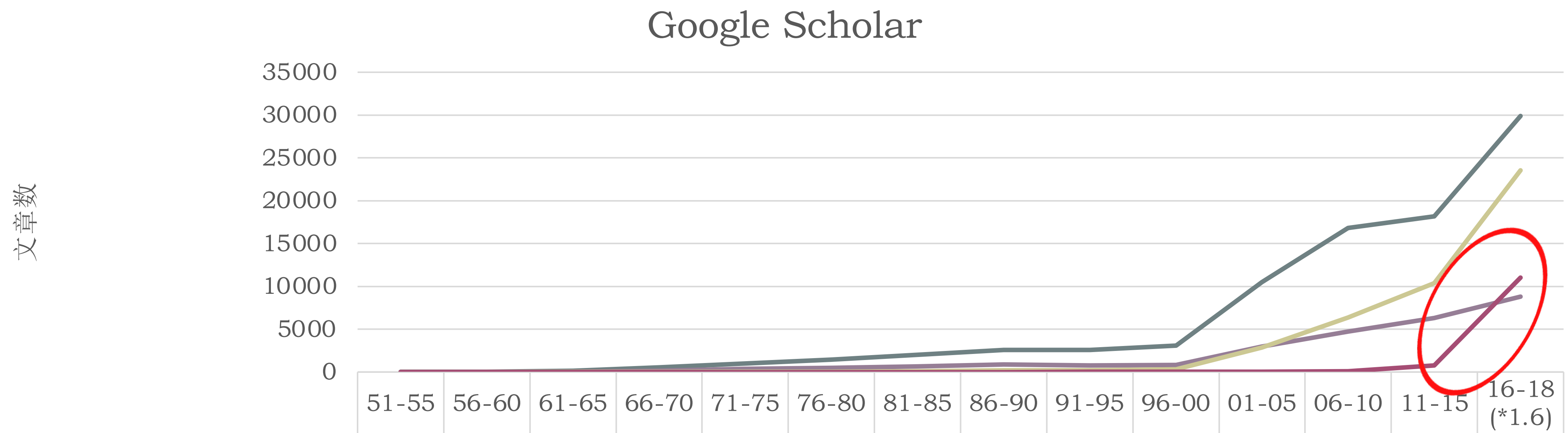
词：分词、新词发现、词性标注、NER、名词短语识别

基础算法



Tech. Trends

- 1950年：图灵发表论文《计算机器与智能》（Computing Machinery and Intelligence），提出著名的图灵测试

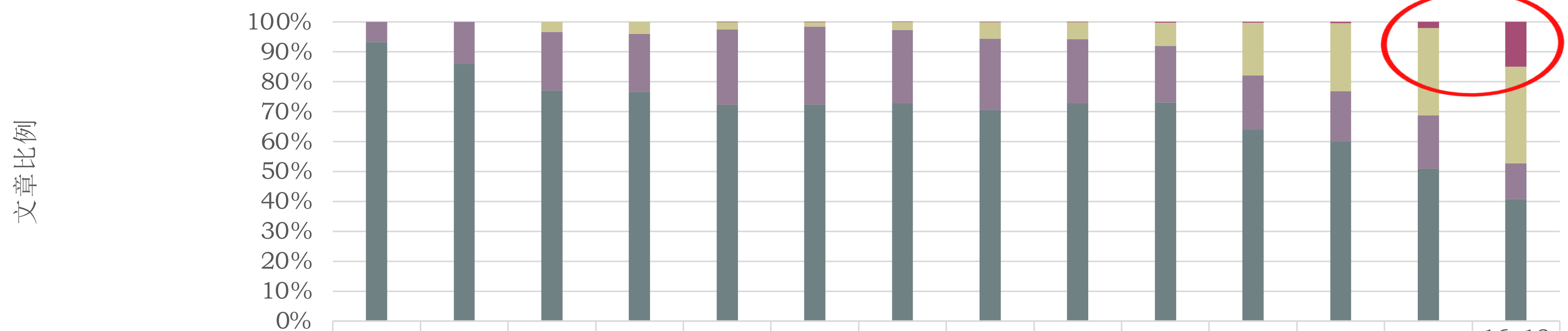


	51-55	56-60	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-00	01-05	06-10	11-15	16-18 (*1.6)
“question answering“	28	55	158	547	1010	1420	2000	2590	2550	3100	10500	16800	18200	29893
AND “grammar“	2	9	40	139	353	510	673	877	751	804	2970	4730	6320	8801
AND “machine learning“	0	0	7	28	34	28	74	202	196	326	2870	6380	10400	23547
AND “deep learning“	0	0	0	0	1	3	1	4	6	13	49	114	752	11022

Tech. Trends

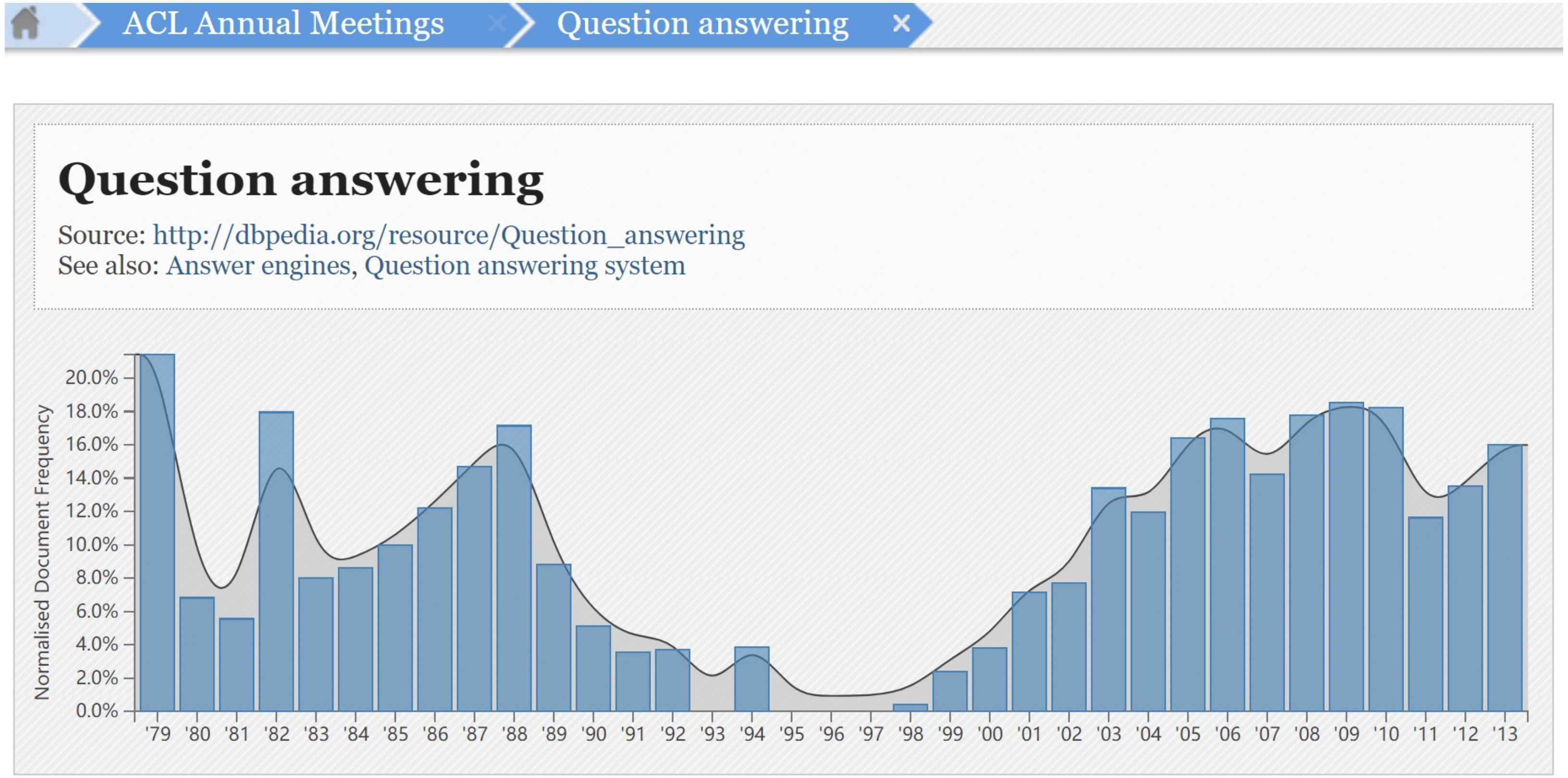
- 1950年：图灵发表论文《计算机器与智能》（Computing Machinery and Intelligence），提出著名的图灵测试

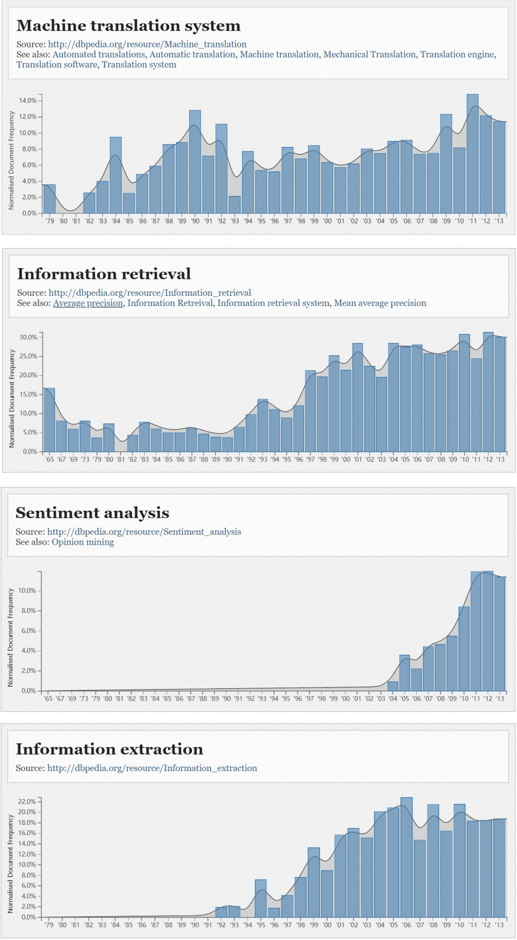
Google Scholar



	51-55	56-60	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-00	01-05	06-10	11-15	16-18 (*1.6)
AND “deep learning“	0	0	0	0	1	3	1	4	6	13	49	114	752	11022
AND “machine learning“	0	0	7	28	34	28	74	202	196	326	2870	6380	10400	23547
AND “grammar“	2	9	40	139	353	510	673	877	751	804	2970	4730	6320	8801
“question answering“	28	55	158	547	1010	1420	2000	2590	2550	3100	10500	16800	18200	29893

Tech. Trends





http://saffron.insight-centre.org/acl_acl/topic/question_answering/

Classification

- 问答系统
  - 分析方法
    - 检索
    - NLU
    - End-to-End
  - 答案形态
    - FAQ
    - 结构化数据
    - 多文档抽取
    - 单文档抽取
    - 自然语言生成
  - 客观性
    - 事实性
    - 非事实
    - 情感性
  - 轮次
    - 单轮
    - 多轮
  - 领域
    - 开放
    - 垂直

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

13

Methods

1. IR-based QA

问答系统

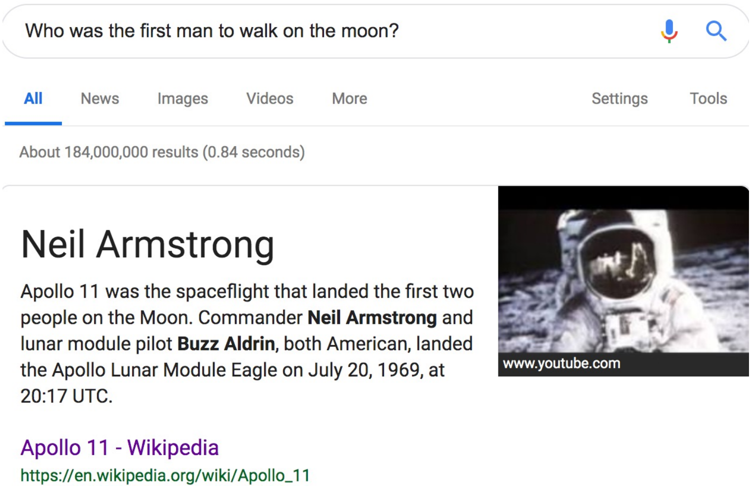
分析方法			答案形态					客观性			轮次		领域	
检索	NLU	End-to-End	FAQ	结构化数据	多文档抽取	单文档抽取	自然语言生成	事实性	非事实	情感性	单轮	多轮	开放	垂直

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

1. IR-based QA

Baidu

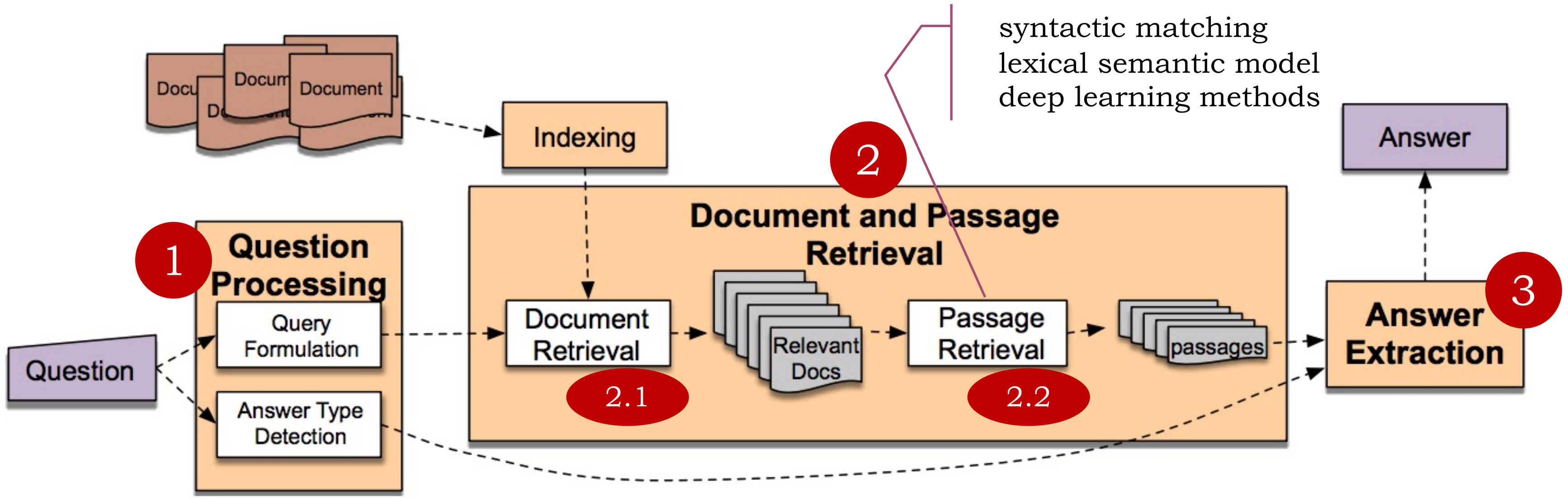
IBM Watson



Google

- 优点：开放域、基于搜索引擎已有技术
- 缺点：精度较低、单轮

1. IR-based QA



- 借助于搜索引擎找到相关文档集合，抽取答案片段
- 核心问题：
  - 问题理解：Query归一化、类型分析、
  - 候选答案抽取、排序：篇章分段、特征提取、段落排序

A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering

- D.Wang, E. Nyberg, A long short-term memory model for answer sentence selection in question answering., in: *ACL, The Association for Computer Linguistics*, 2015, pp. 707–712.
- Citations: 126

Eric Nyberg

Professor and Director, Master of Computational Data Science Program

Carnegie Mellon University
School of Computer Science
Language Technologies Institute

[Contact]

Noted for his contributions to the fields of automatic text translation, information retrieval, and automatic question answering, Nyberg holds a Ph.D. from Carnegie Mellon University (1992) and a B.A. from Boston University (1983). He is a recipient of the Allen Newell Award for Research Excellence (for his contributions to the field of question answering and his work as an original developer on the Watson project) and the BU Computer Science Distinguished Alumna/Alumnus Award. Eric currently directs the Master of Computational Data Science (MCDS) program. He is also co-Founder and Chief Data Scientist at Cognistx, and serves on the Scientific Advisory Board for Fairhair.ai.

http://www.cs.cmu.edu/~ehn/

Tuesday, November 22, 2016 - by Bryan Burtner

For the second consecutive year, Carnegie Mellon came out on top in the LiveQA evaluation – an exercise that requires question-answering (QA) software to respond to real-time questions received by the Yahoo! Answers website – at the Text Retrieval Conference (TREC 2016).

A system designed by Di Wang, a Language Technologies Institute Ph.D. student and member of Prof. Eric Nyberg's Open Advancement of Question Answering (OAQA) research group, out-paced competitors in answering user-generated questions in real time.

The questions, ranging from the mundane to the perplexing (e.g., "How do I convince my mom to pay for my gym?", "Can I keep ferrets and rats in the same room?"), were selected from a collection of Yahoo! Answers submissions that had not yet been answered by human users. The QA systems were allowed one minute to answer each question. Wang's system, which uses a Deep Learning approach to question answering, received the highest average score and success rate among 25 automatic QA systems from 14 teams.

LTI Ph.D. student Di Wang designed a system to rapidly answer questions posed to the Yahoo! Answers website that received the highest score in the LiveQA evaluation track at TREC 2016.

http://www.cs.cmu.edu/~ehn/

A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering

- **Focus**: answer sentence selection
- **Challenge**: match not just words but also **meaning** between Q and A

珍妮弗·卡普里亚蒂_百度图片

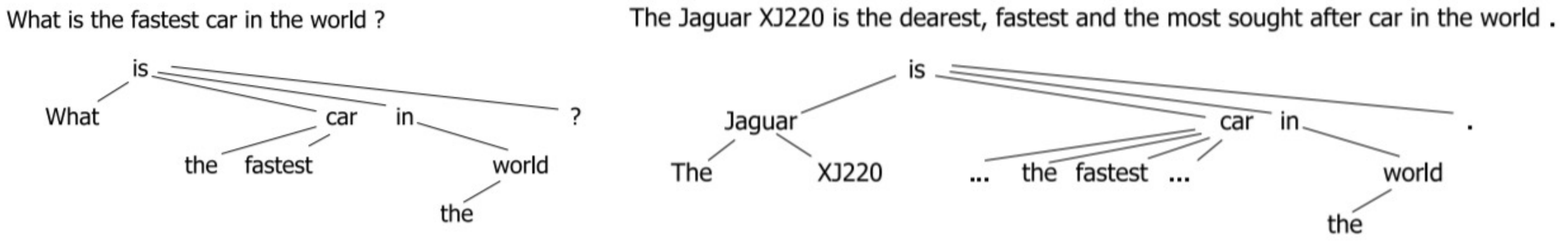
"What sport does Jennifer Capriati play?"

Positive Sentence:
*"Capriati, 19, who has not played competitive **tennis** since November 1994, has been given a wild card to take part in the Paris tournament which starts on February 13."*

Negative Sentence:
"Capriati also was playing in the U.S. Open semifinals in '91, one year before Davenport won the junior title on those same courts."

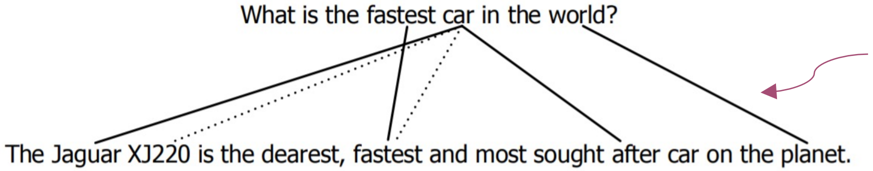
A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering

- **Existing Methods:**
  - Syntactic Matching



Edit Distance: the costs of all operations, including deleting, inserting, and changing a node, needed when transforming an ordered labeled tree to another

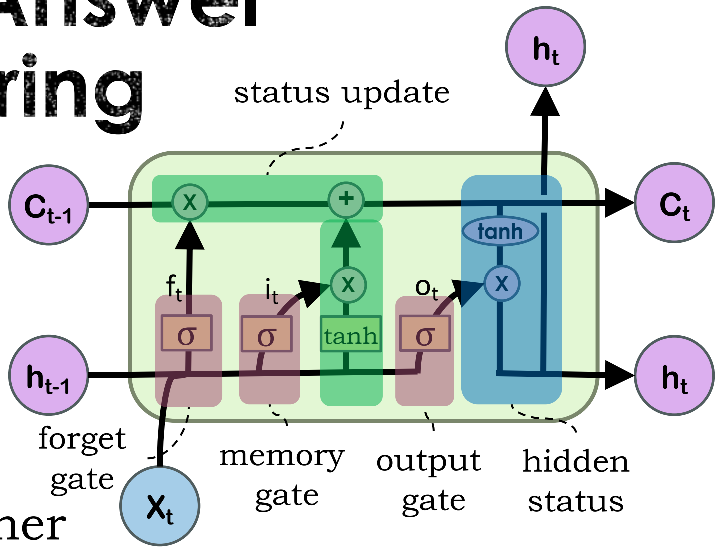
  - Lexical Semantic Matching



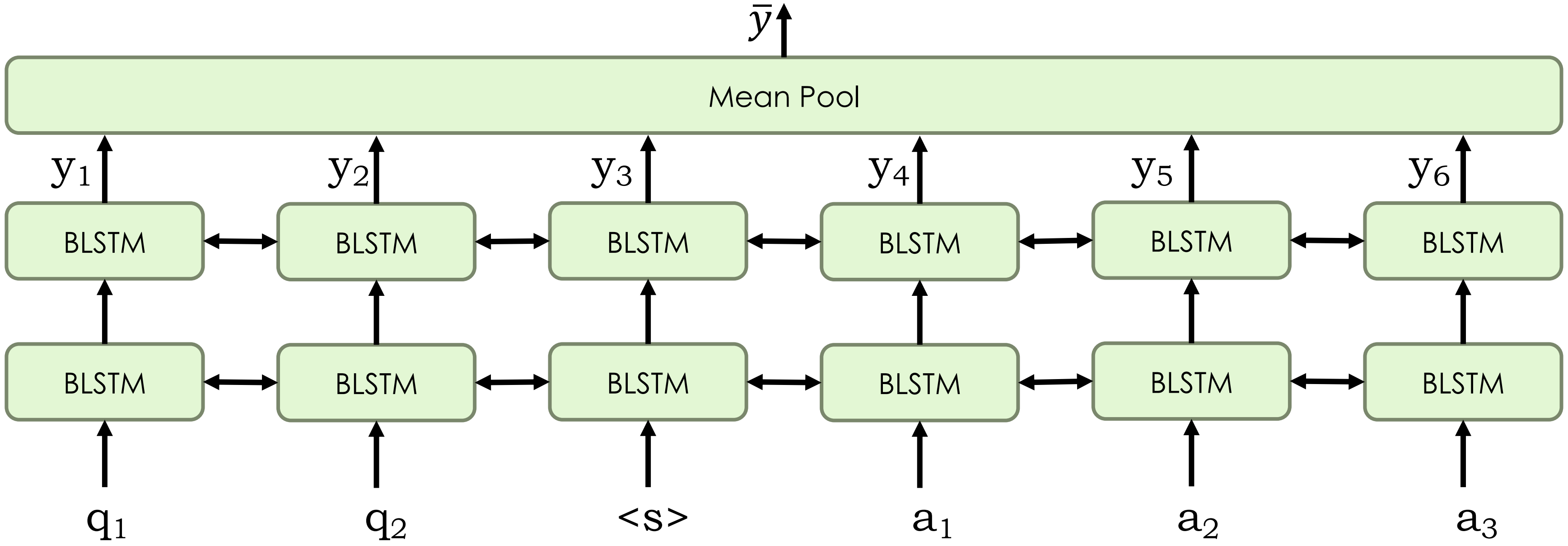
solid lines are clear synonyms or hyponym/hypernym; dashed lines are weaker semantic association

- **Goal:** reduce dependency of syntactic features and other (lexical) resources

A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering



- **NN Model**
  - LSTM + Bidirectional + Stacked
  - Input: word embeddings of q + <s> + a
  - Output: whether or not a is selected for q
  - Motivation: combine all contextual information in q and a together



A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering

Methods dependent of syntactic features and external resources

- **Weakness**
  - cardinal numbers and proper nouns matching is crucial, but
  - (1) many of them are OOV
  - (2) embeddings have noise:
    - "*China*" vs. "*Japan*" : close in embedding space but very different in *q*-*a* matching
- **Solution**
  - Combine BLSTM with keywords overlapping by GBDT
- **Experiment**
  - (TREC) QA track (8-13) data
    - training set: 1229 questions, each on average associated with 33 candidate labeled answers
    - test set: 100 questions

Reference	MAP	MRR
Yih et al. (2013) – Random	0.3965	0.4929
Wang et al. (2007)	0.6029	0.6852
Heilman and Smith (2010)	0.6091	0.6917
Wang and Manning (2010)	0.5951	0.6951
Yao et al. (2013)	0.6307	0.7477
Severyn and Moschitti (2013)	0.6781	0.7358
Yih et al. (2013) – BDT	0.6940	0.7894
Yih et al. (2013) – LCLR	0.7092	0.7700

Features	MAP	MRR
BM25	0.6370	0.7076
Single-Layer LSTM	0.5302	0.5956
Single-Layer BLSTM	0.5636	0.6304
Three-Layer BLSTM	0.5928	0.6721
Three-Layer BLSTM + BM25	**0.7134**	**0.7913**

Recommended References

- Tellex, Stefanie, et al. "Quantitative evaluation of passage retrieval algorithms for question answering." *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2003.
- Cui, Hang, et al. "Question answering passage retrieval using dependency relations." *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2005.
- Kolomiyets, Oleksandr, and Marie-Francine Moens. "A survey on question answering technology from an information retrieval perspective." *Information Sciences* 181.24 (2011): 5412-5434.
- Monz, Christof. "From document retrieval to question answering." (2003).
- Zhou, Guangyou, et al. "Learning continuous word embedding with metadata for question retrieval in community question answering." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Vol. 1. 2015.
- Dumais, Susan, et al. "Web question answering: Is more always better?." *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2002.

2. Community QA

问答系统

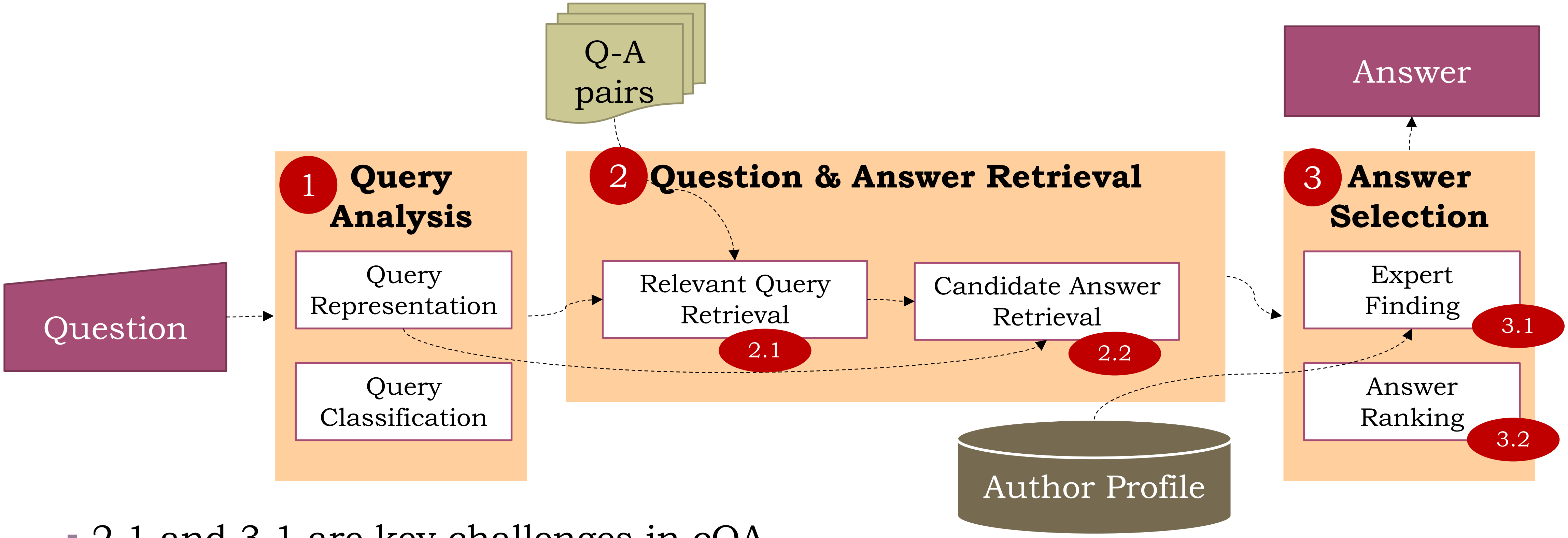
分析方法	答案形态	客观性	轮次	领域
检索	FAQ	事实性	单轮	开放
NLU	结构化数据	非事实	多轮	垂直
End-to-End	多文档抽取	情感性		
	单文档抽取			
	自然语言生成			

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

2. Community QA

- 优点：数据丰富、实现简单
- 缺点：数据质量不佳（作弊、噪声、错误）

2. Community QA



- 2.1 and 3.1 are key challenges in cQA
- SemEval-2017 task 3: Community question answering

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

- Guangyou Zhou, Tingting He, Jun Zhao, and Po Hu. *Proceedings of the 53rd ACL and the 7th International Joint Conference on Natural Language Processing.* 2015.
- Citations: 126

Guangyou Zhou
Central China Normal University;
Chinese Academy of Sciences, Institute of Automation

[+] [–] Guangyou Zhou

> Home > Persons

[–] Person information

- *affiliation:* Central China Normal University, Wuhan, China
- *affiliation:* Chinese Academy of Sciences, Institute of Automation, Beijing, China

[–] 2010 – today

2019

[j10] Yehong Peng, Yizhen Fang, Zhiwen Xie, Guangyou Zhou: **Topic-enhanced emotional conversation generation with attention mechanism.** Knowl.-Based Syst. 163: 429-437 (2019)

2017

[j9] Zhiwen Xie, Zhao Zeng, Guangyou Zhou, Weijun Wang: **Topic enhanced deep structured semantic models for knowledge base question answering.** SCIENCE CHINA Information Sciences 60(11): 110103:1-110103:15 (2017)

[j8] Guangyou Zhou, Jimmy Xiangji Huang: **Modeling and Learning Distributed Word Representation with Metadata for Question Retrieval.** IEEE Trans. Knowl. Data Eng. 29(6): 1226-1239 (2017)

[j7] Guangyou Zhou, Jimmy Xiangji Huang: **Modeling and Mining Domain Shared Knowledge for Sentiment Analysis.** ACM Trans. Inf.

[j7] Guangyou Zhou, Jimmy Xiangji Huang: **Modeling and Mining Domain Shared Knowledge for Sentiment Analysis.** ACM Trans. Inf.

2011

[c5] Guangyou Zhou, Li Cai, Jun Zhao, Kang Liu: **Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives.** ACL 2011: 653-662

[c4] Guangyou Zhou, Jun Zhao, Kang Liu, Li Cai: **Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing.** ACL 2011: 1556-1565

[c3] Li Cai, Guangyou Zhou, Kang Liu, Jun Zhao: **Large-scale question classification in cQA by leveraging Wikipedia semantic knowledge.** CIKM 2011: 1321-1330

[c2] Guangyou Zhou, Li Cai, Kang Liu, Jun Zhao: **Improving Dependency Parsing with Fined-Grained Features.** IJCNLP 2011: 228-236

[c1] Li Cai, Guangyou Zhou, Kang Liu, Jun Zhao: **Learning the Latent Topics for Question Retrieval in Community QA.** IJCNLP 2011: 273-281

Jun Zhao 赵军

Professor

Research Group of Speech and Language Technology,
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
Office: Room 703, Intelligence Building, Institute of Automation, Chinese Academy of Sciences, Zhongguancun East Road 95#, Beijing, China, 100190
Email: jzhao(at)nlpr(dot)ia(dot)ac(dot)cn

Bio

Dr Zhao Jun is a professor at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. He received his PhD degree from Tsinghua University in 1998. Before joining NLPR in 2002, he worked in the Hong Kong University of Science and Technology as a postdoctoral research fellow. His current research focuses on natural language processing, information extraction and question answering. Prof Zhao has published over 50 peer-reviewed papers in the prestigious conferences and journals, including ACL, SIGIR, TKDE, JLMR, IJCAI, EMNLP, etc. His paper "Relation Classification via Convolutional Deep Neural Network" obtained best paper award of COLING-2014. His paper "Collective entity linking in web text: a graph-based method" ranks 2 in the highest referenced papers of SIGIR in recent five years with the Google academic search. He also served as workshop cochair for ACL-2016.

http://nlpr-web.ia.ac.cn/cip/english/~junzhao/index.html#

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

- **Focus**: semantically equivalent or relevant questions retrieval
- **Challenge**: lexical gap problem
  - "*how do I get **knots** out of my cats fur?*"
  - "*how can I remove a **tangle** in my cat's fur?*"
- **Existing Methods:**
  - **Translation model / Topic-based model**: semantically related *q-a* retrieval
    - The basic assumption is questions and answers are "**parallel texts**" or "**same topic**", which is not true.
    - (1) One question have multiple answers which are diverse and contain much more information than the question
    - (2) Many answers are low quality and make the learnt translation probabilities / topic models unreliable
  - **Deep linguistic analysis**

Table 1: An example from Yahoo! Answers

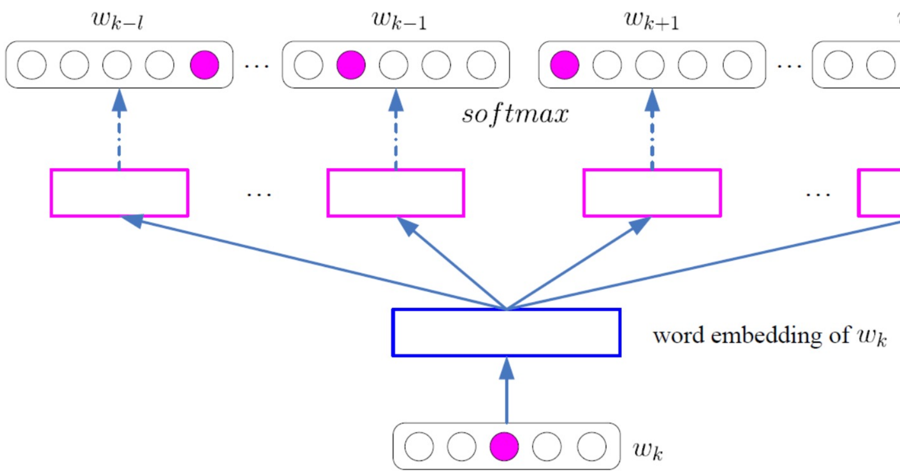
Question (Question title)	Connected to the Internet but cannot get online!??
Description (Question body)	I have put a clean install of XP on my laptop and installed all the network and internet drivers from Dell......
Answer 1	Does the Wi-Fi Connection have a passkey, because if it does the internet wont work if you have XP SP1. However you can upgrade your XP to SP2. Then it should work.Or you could take the passkey off your Wi-Fi
Answer 2	how have you posted this if you cant go on the internet

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

- **Method**:
  - (1) transform question into BoEW (Bag-of-**Embedded**-Words)
  - (2) aggregate variable-cardinality BoEWs into **fixed-length** ones by **Fisher kernel**

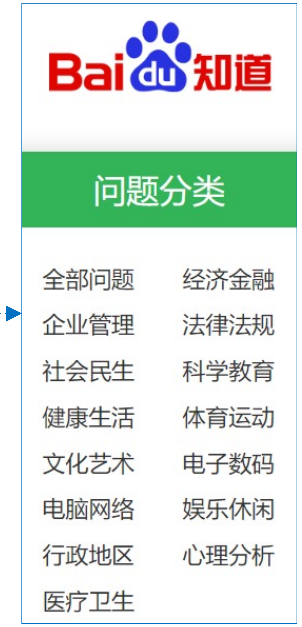
Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

- **Method**:
  - (1) transform question into BoEW (Bag-of-**Embedded**-Words)
    - existing word embeddings mainly based on word co-occurrence, which perform poorly when i) similar words with **very little context** or ii) context could be noisy or biased
    - present a novel method to learn word **embeddings with metadata** on cQA data set
    - use the **Skip-gram model** for learning word embeddings, since it is much more efficient as well as memory-saving than other approaches

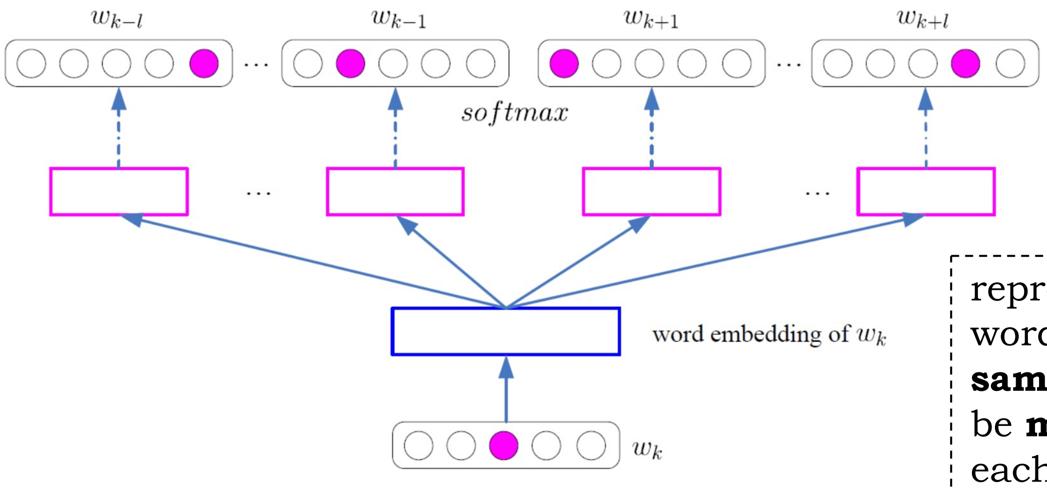


Traditional Skip-gram Model

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

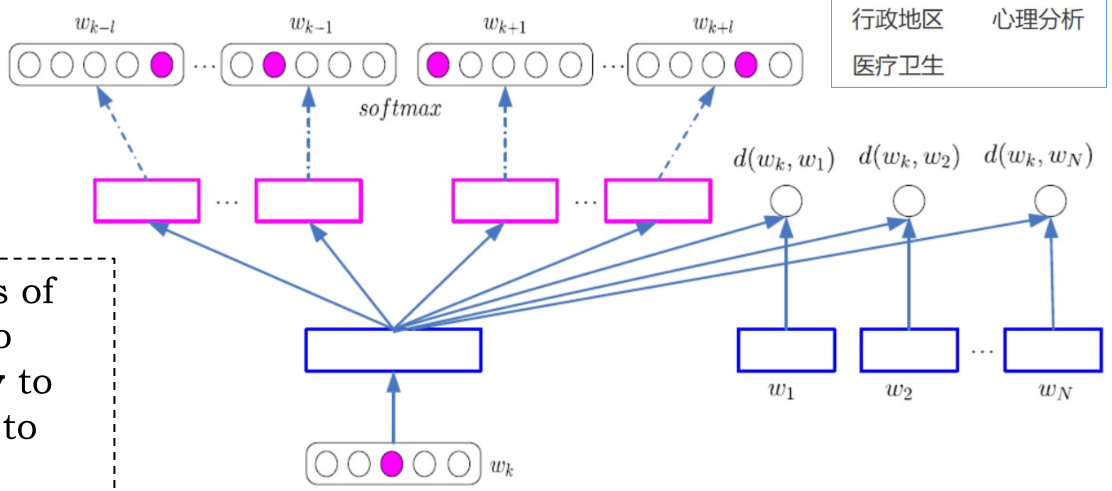


- **Method**:
  - (1) transform question into BoEW (Bag-of-**Embedded**-Words)
    - cQA metadata, such as "**category**", "voting", can **benefits embedding learning**
    - Assumption: **category encodes (implies)** attributes or properties of question words
    - "*What are the security issues with java?*" ∈ "Computers & Internet >> Security"



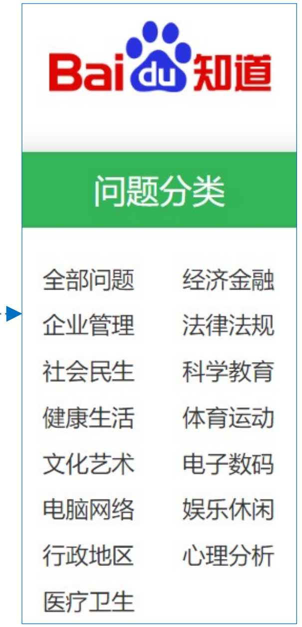
Traditional Skip-gram Model

representations of words belong to **same category** to be **more close** to each other



M-NET: continuous skip-gram model with metadata of category information

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering



- **Method**:
  - (1) transform question into BoEW (Bag-of-**Embedded**-Words)
    - cQA metadata, such as "**category**", "voting", can **benefits embedding learning**
    - Assumption: **category encodes (implies)** attributes or properties of question words
    - “*What are the security issues with java?*” ∈ “Computers & Internet >> Security”

category information

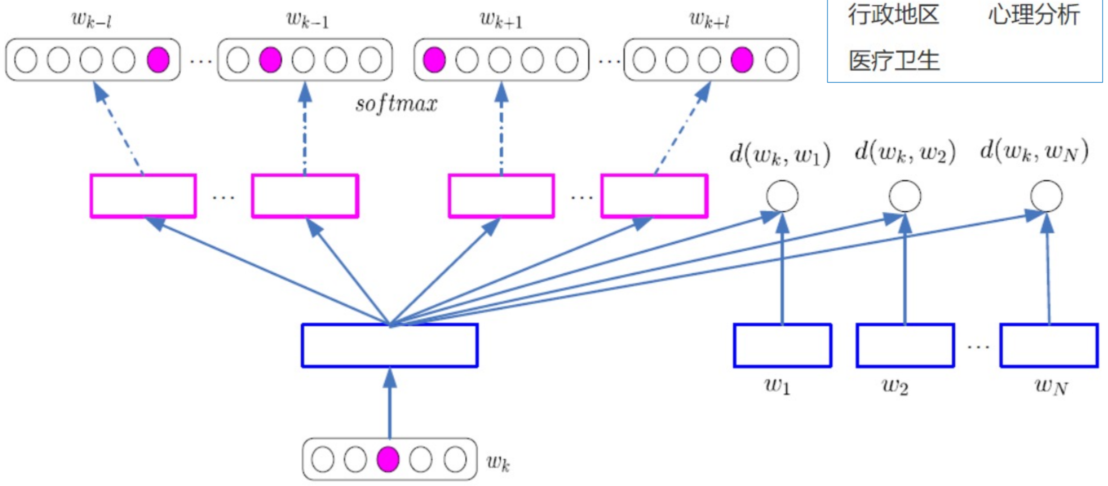
Objective: $J_c = J(\theta) + \beta \boxed{E_c}$

Euclidean distance

$$E_c = \sum_{k=1}^{N}\sum_{i=1}^{N} s(w_k, w_i)\boxed{d(w_k, w_i)}$$

category of w_k

$$s(w_k, w_i) = \begin{cases} 1 & \text{if } \boxed{c(w_k)} = c(w_i) \\ 0 & \text{otherwise} \end{cases}$$



M-NET: continuous skip-gram model with metadata of category information

Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering

- **Method**:
  - (2) aggregate variable-cardinality BoEWs into **fixed-length** ones by **Fisher kernel**
    - BoEWs are variable-size sets and most of the index methods in IR are not suitable
    - Fisher kernel: a probability density function. See More
    - 核方法：将不可分难题拉到高维进行可分；核函数：将极难的计算拉回到低维进行计算

Model	dim	Yahoo data				Baidu data			
		MAP	MRR	R-Prec	P@5	MAP	MRR	R-Prec	P@5
LM (baseline)	-	0.435	0.472	0.381	0.305	0.392	0.413	0.325	0.247
(Jeon et al., 2005)	-	0.463	0.495	0.396	0.332	0.414	0.428	0.341	0.256
(Xue et al., 2008)	-	0.518	0.560	0.423	0.346	0.431	0.435	0.352	0.264
(Zhou et al., 2011)	-	0.536	0.587	0.439	0.361	0.448	0.450	0.367	0.273
(Ji et al., 2012)	-	0.508	0.544	0.405	0.324	0.425	0.431	0.349	0.258
(Zhang et al., 2014a)	-	0.527	0.572	0.433	0.350	0.443	0.446	0.358	0.265
Skip-gram + FV	50	0.532	0.583	0.437	0.358	0.447	0.450	0.366	0.272
	100	0.544	0.605†	0.440	0.363	0.454	0.457	0.373	0.274
	300	0.550†	0.619†	0.444	0.365	0.460†	0.464†	0.374	0.277
M-NET + FV	50	0.548†	0.612†	0.441	0.363	0.459†	0.462†	0.374	0.276
	100	0.562‡	0.628‡	0.452†	0.367‡	0.468‡	0.471	0.378†	0.280†
	300	**0.571**‡	**0.643**‡	**0.455**‡	**0.374**‡	**0.475**‡	**0.477**‡	**0.385**‡	**0.283**‡

- **Experiment:**

	#queries	#candidate	#relevant
Yahoo data	1,000	13,000	2,671
Baidu data	1,000	8,000	2,104

Recommended References

- Srba, Ivan, and Maria Bielikova. "A comprehensive survey and classification of approaches for community question answering." *ACM Transactions on the Web (TWEB)* 10.3 (2016): 18.
- Liu, Yandong, Jiang Bian, and Eugene Agichtein. "Predicting information seeker satisfaction in community question answering." *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008.
- Li, Baichuan, and Irwin King. "Routing questions to appropriate answerers in community question answering services." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.
- Riahi, Fatemeh, et al. "Finding expert users in community question answering." *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012.
- Liu, Xiaoyong, W. Bruce Croft, and Matthew Koll. "Finding experts in community-based question-answering services." *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 2005.
- Ji, Zongcheng, et al. "Question-answer topic model for question retrieval in community question answering." *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012.
- Zhou, Guangyou, et al. "Learning continuous word embedding with metadata for question retrieval in community question answering." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Vol. 1. 2015.
- Qiu, Xipeng, and Xuanjing Huang. "Convolutional neural tensor network architecture for community-based question answering." *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

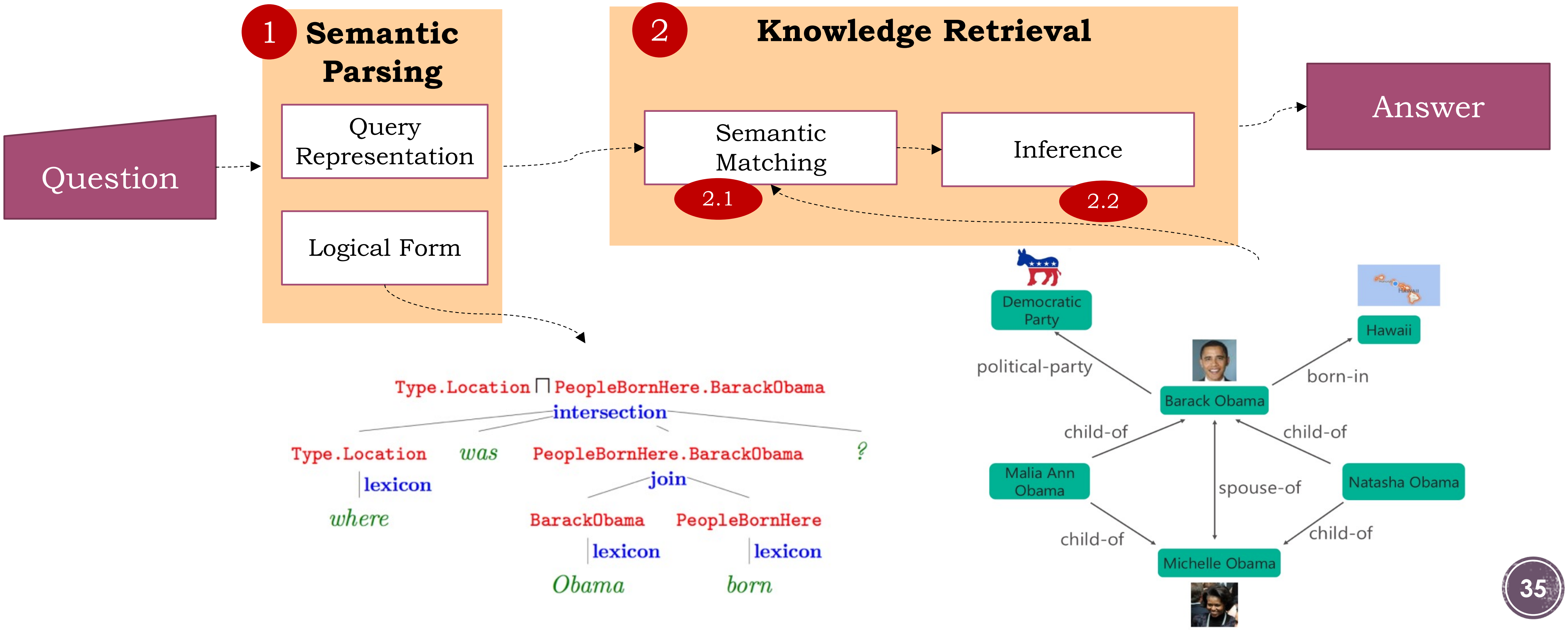
3. Knowledge base QA

问答系统

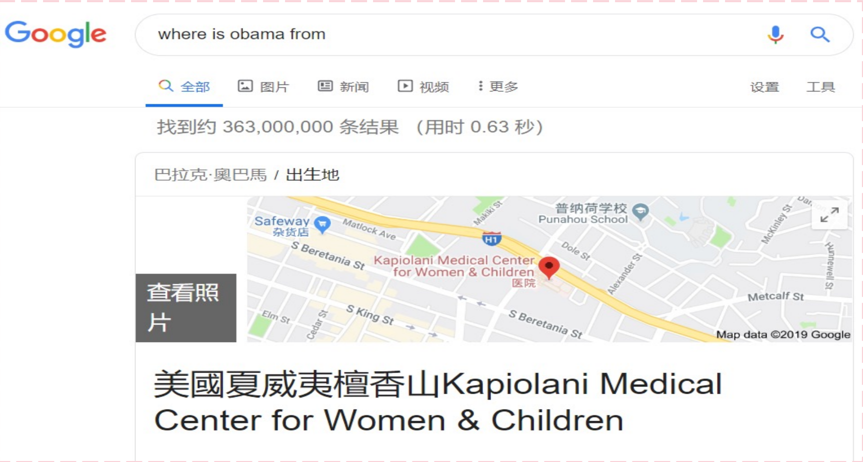
分析方法			答案形态					客观性			轮次		领域	
检索	NLU	End-to-End	FAQ	结构化数据	多文档抽取	单文档抽取	自然语言生成	事实性	非事实	情感性	单轮	多轮	开放	垂直

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

3. Knowledge base QA



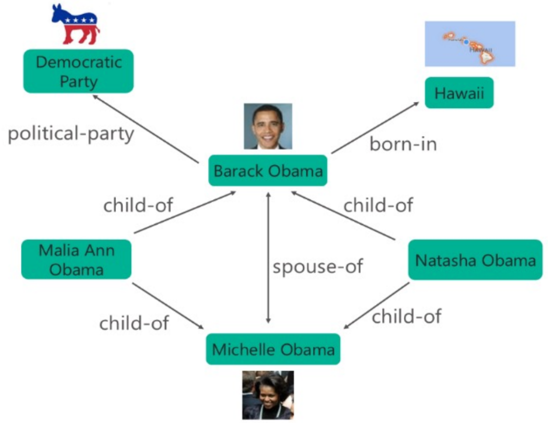
3. Knowledge base QA



- 优点：提供高精度答案，支持推理
- 缺点：依赖知识图谱，语义解析困难

Knowledge Base

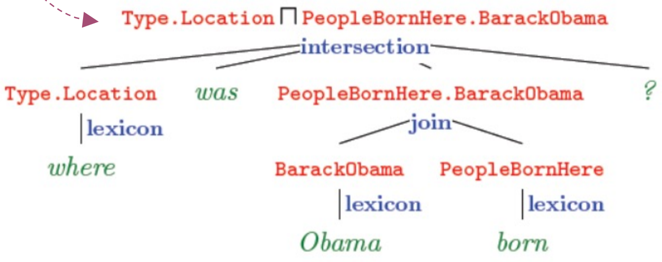
- KB: **facts in a structured form** (SPO triple, subject-predicate-object)
  - **DBPedia**: 2014 版拥有超过458万的物件，包括144.5万人、73.5地点、12.3张唱片、8.7万部电影、1.9万种电脑游戏、24.1万个组织、25.1种物种、6000个疾病 http://blog.dbpedia.org/
  - **Freebase**: 一个由元数据组成、允许全球所有人**和机器**快捷访问的资源库，内容主要来自社区成员贡献。由Metaweb开发，2007年3月公开，2010年7月被谷歌收购，并推出“Knowledge Graph”概念。2014年12月，Google宣布6个月后关闭Freebase，并将其数据迁移至维基数据



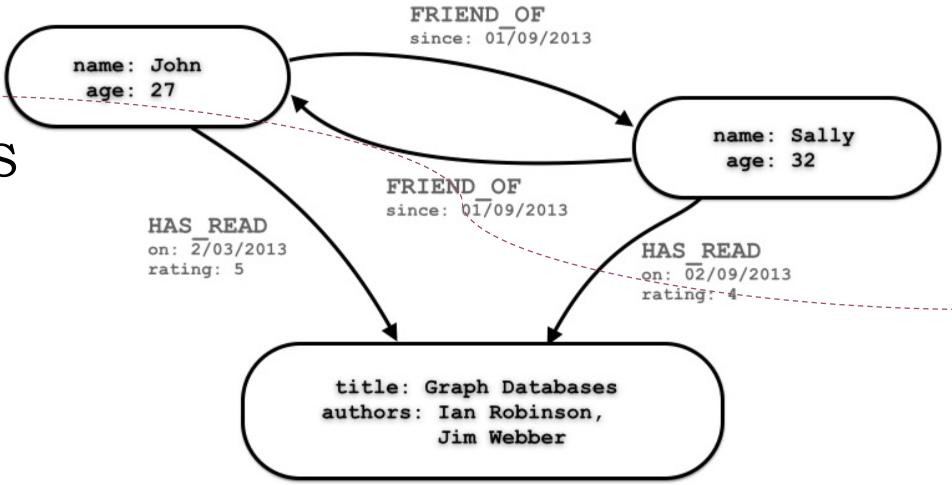
- KB query language
  - **Lambda calculus(λ-calculus)**
    - 1930s初，普林斯顿大学的逻辑学家阿伦佐·丘奇 (Alonzo Church, 1903-1995) 开发出了一种新的形式系统拉姆达运算/演算 (λ-calculus)，其核心是λ表达式，以此形成函数定义、函数应用和递归的形式系统
    - λ运算是函数式编程语言共同的祖先，典型代表是Lisp(Scheme)、ML、Haskell和Erlang等。任何一个可计算函数都能用λ运算来表达和求值，它等价于图灵机
    - “*number of dramas starring Tom Cruise*”:
      count(λx.Genre(x;Drama)$\wedge\exists y$.Performance(x;y)$\wedge$Actor(y;TomCruise))

Querying Knowledge Base

- KB query language
  - **λ-DCS** (Lambda Dependency-Based Compositional Semantics): a **simpler form** of λ-calculus, which is derived from tree-structured description logic and was designed for building a natural language **interface into Free-base**
    - https://arxiv.org/abs/1309.4408; https://cs.stanford.edu/~pliang/papers/freebase-emnlp2013.pdf
    - `count`$(\lambda x.\mathrm{Genre}(x;\mathrm{Drama})\wedge\exists y.\mathrm{Performance}(x;y)\wedge\mathrm{Actor}(y;\mathrm{TomCruise}))$ → `count`$(\mathrm{Genre.Drama}\sqcap\mathrm{Performance.Actor.TomCruise})$



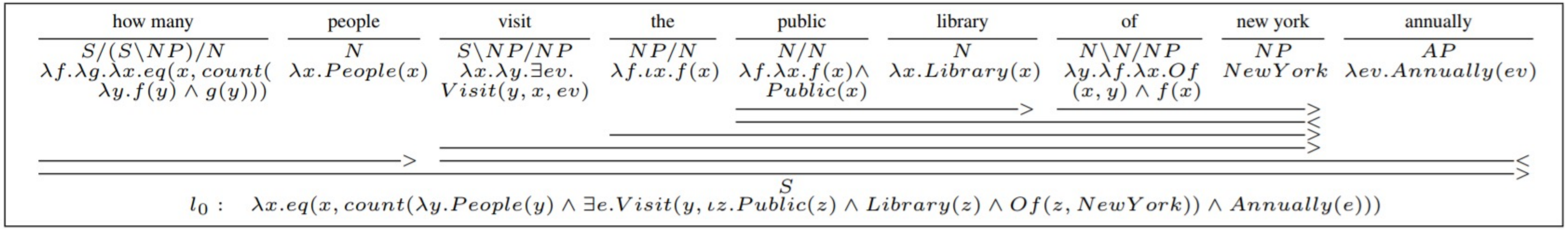
  - **SPARQL**: a graph query languages, and can be seen a **implement** of λ-DCS
    - For graph database
    - 查询即子图匹配
    - W3C标准




```
MATCH (sally:Person { name: 'Sally' })
MATCH (john:Person { name: 'John' })
MATCH (sally)-[r:FRIEND_OF]-(john)
RETURN r.since as friends_since
```

Semantic Parsing

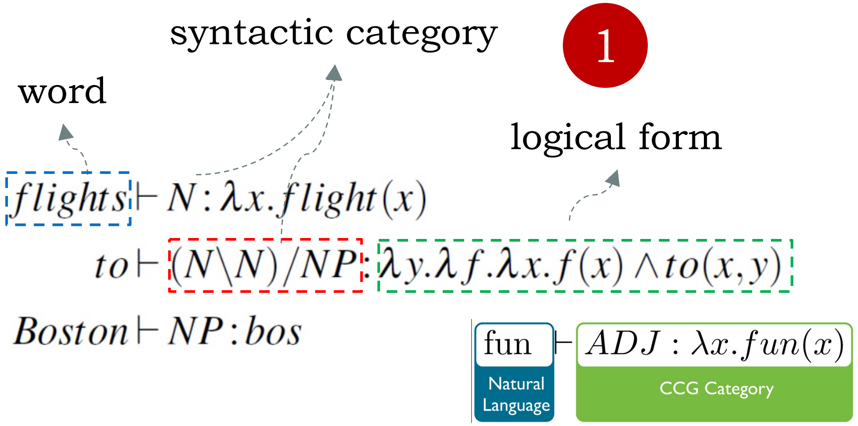
- KB-QA Challenges
  - Semantic parsing
    - Lexical gap (ontology matching), Parsing

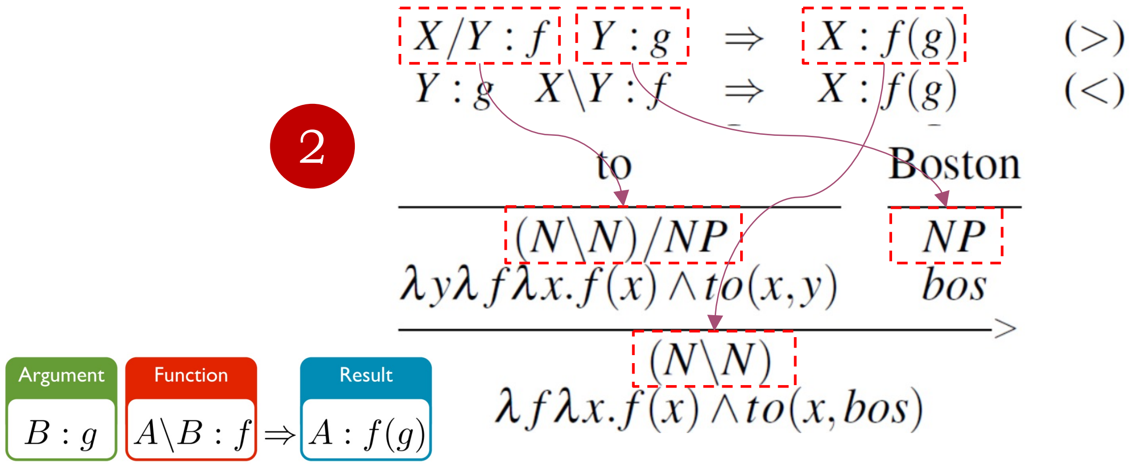


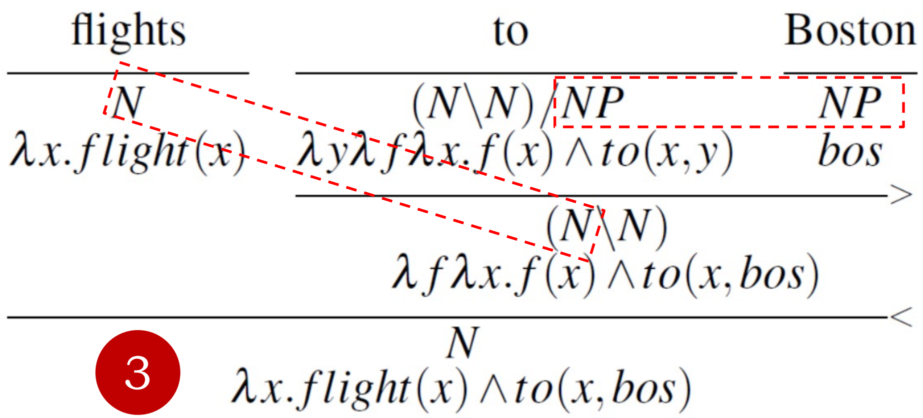
  - 除问答外，语义分析是包括包括信息抽取、NLG、情感分析等诸多应用的瓶颈，是目前NLP领域最大的挑战之一

Semantic Parsing with CCG

- Semantic parsing methods
  - CCG(Combinatory Categorial Grammar) parser
    - 一种通过词汇范畴显式地关联句法和语义的短语结构文法
    - 基于组合逻辑，同时显示给出句法组合和语义组合方法
    - 问题：CCG Induction
      - Lexicon and Combinators learning



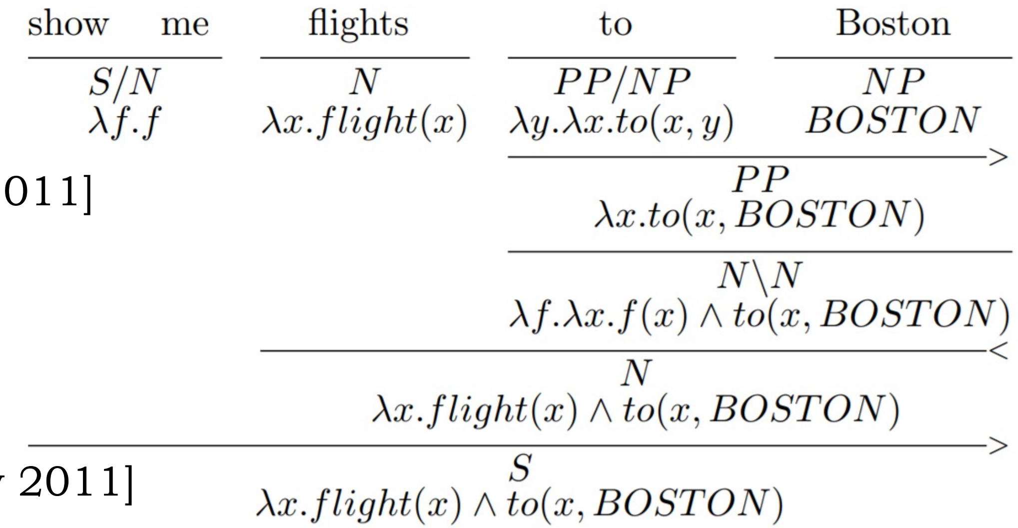




https://yoavartzi.com//pub/afz-tutorial.acl.2013.pdf

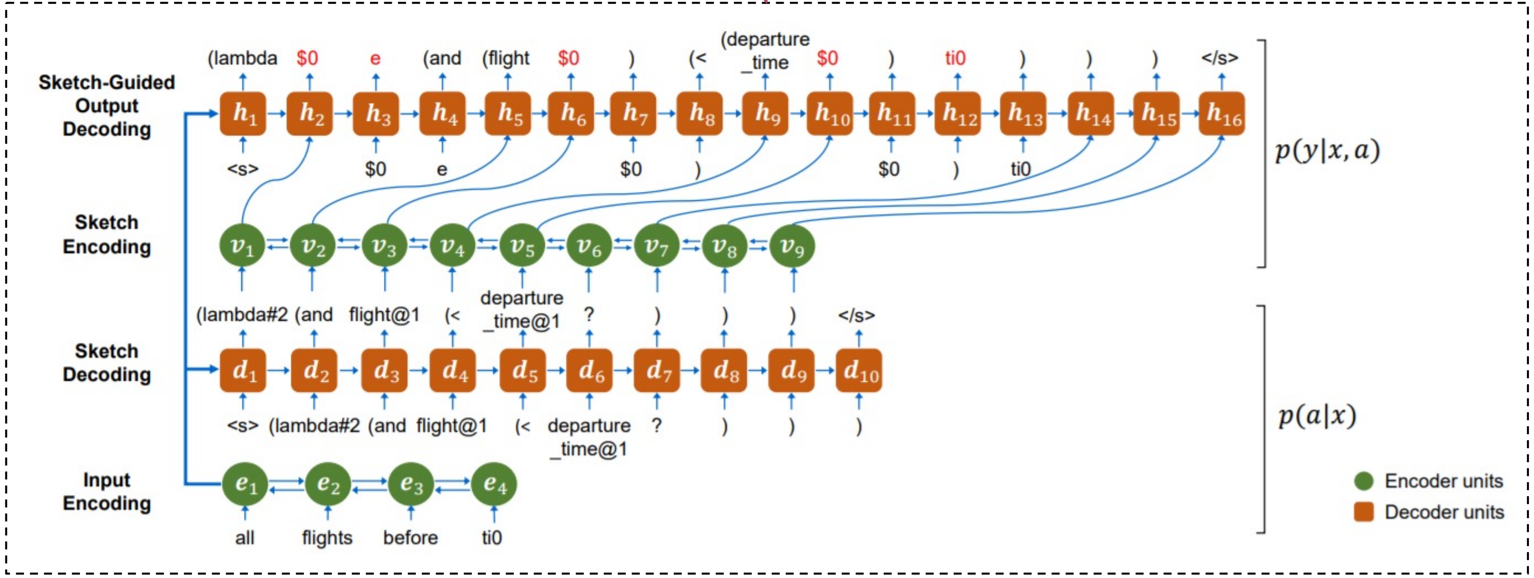
Semantic Parsing with CCG

- Semantic parsing methods
  - **Other Parser**:
    - Inductive Logic Programming [Zelle and Mooney 1996]
    - SCFG [Wong and Mooney 2006]
    - CCG + CKY [Zettlemoyer and Collins 2005]
    - Constrained Optimization + ILP [Clarke et al. 2010]
    - DCS + Projective dependency parsing [Liang et al. 2011]
  - **Gramma Induction**:
    - Annotated parse trees [Miller et al. 1994]
    - Sentence-LF pairs [Zettlemoyer and Collins 2005]
    - Question-answer pairs [Clarke et al. 2010]
    - Instruction-demonstration pairs [Chen and Mooney 2011]
    - Conversation logs [Artzi and Zettlemoyer 2011]
    - Visual sensors [Matuszek et al. 2012a]

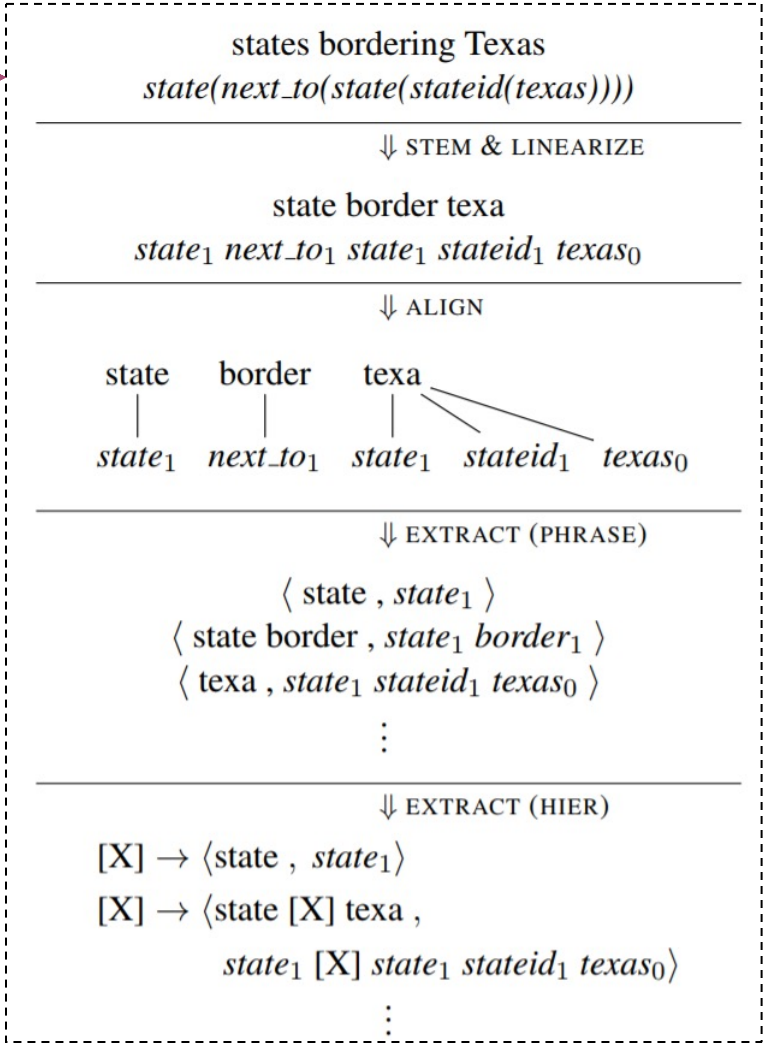


Other Semantic Parsing Methods

- Semantic parsing methods
  - as **Machine Translation**
  - Unsupervised Semantic Parsing with **Markov Logic**
  - **Neural** Semantic Parsing



https://arxiv.org/pdf/1805.04793.pdf



https://www.aclweb.org/anthology/P13-2009

Other Semantic Parsing Methods

- NLU as Intent Recognition and Slot Filling
  - 信息槽 (Information slot)：对应值约束信息
  - 请求槽 (request slot)：对应值省略用以请求信息
    - *I'd like to see Our King of Traitor tonight in Seattle*
    - Request(ticket, movie-name=*Our King of Traitor*, start-time=*tonight*, city=*Seattle*)

Sentence	show	restaurant	at	New	York	tomorrow
Slots	O	O	O	B-desti	I-desti	B-date
Intent	Find Restaurant					
Domain	Order					

Table 1: An Illustrative Example of Natural Language Representation.

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Wen-tau Yih, Ming-Wei Chang, Xiaodong He, Jianfeng Gao. Semantic parsing via staged query graph generation: Question answering with knowledge base. In *Proceedings of the Joint Conference of the 53rd ACL and the 7th International Joint Conference on AFNLP*. 2015. *Outstanding Paper*
- Citations: 224

Scott Yih

Research Scientist

Facebook AI Research (FAIR)

Research Positions

Research Scientist, Facebook AI Research, 2019—Present

Principal Research Scientist, Allen Institute for Artificial Intelligence, 2017—2019

Senior Researcher, Microsoft Research, 2015—2017

Researcher, Microsoft Research, 2006—2015

Post-Doc Researcher, Microsoft Research, 2005—2006

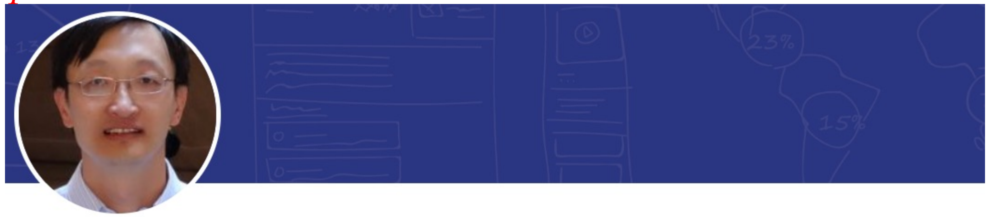
Education

Ph.D. in Computer Science
University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 2005
Dissertation Title: *Learning and Inference for Information Extraction*

M.S. in Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, 1997
Thesis Title: *Template-based Information Extraction from Tree-structured HTML Documents*

B.S. in Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, 1995

http://scottyih.org/



Jianfeng Gao

Partner Research Manager

About Projects Publications Videos Downloads Toolkit

Contact Info

Email

Website

Research areas

Human language technologies

Artificial intelligence

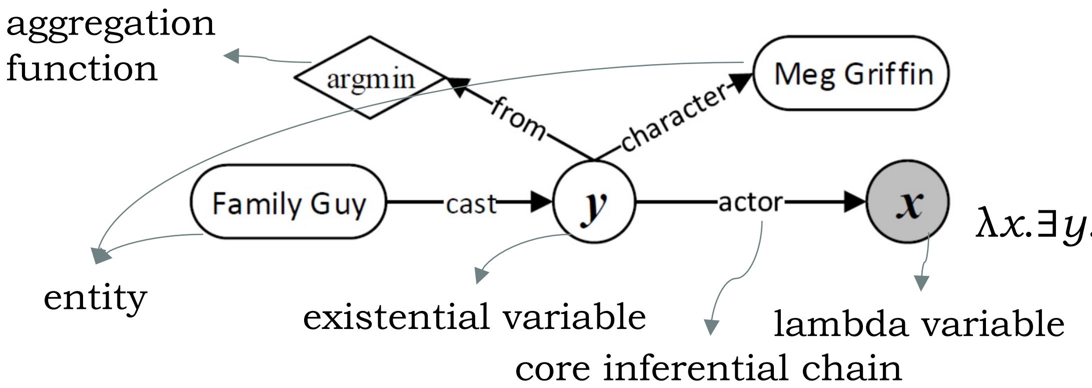
Partner Research Manager in the Deep Learning Group at Microsoft Research AI. IEEE Fellow.

From 2014 to 2017, I was Partner Research Manager in Business AI at Microsoft AI & Research and at Deep Learning Technology Center (DLTC) at Microsoft Research, Redmond. I lead the development of AI solutions to Predictive Sales and Marketing. I also work on deep learning for text and image processing (see our ACL/SIGIR2018 Tutorial, DeepLearning2017 Tutorial and IJCAI2016 Tutorial or *MS internal site*) and lead the development of AI systems for dialogue, machine reading comprehension (MRC), and question answering (QA).

https://www.microsoft.com/en-us/research/people/jfgao/

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Traditional methods' weakness:
  - **Generic** meaning representation → **Lexical gap** between language and KB
    - E.g., *daughter, number of people living in* ←→ `gender`, `parenthood`, `population`
  - Specific KB independent parser → annotated data sparseness
    - E.g., *What is the location of **ACL2014**? (missing in KB)*
- Motivation: **leverages KB more tightly** when parse a question
- Method:
  - 将λ-演算等价为一个**查询图**，语义解析则转化为查询图**生成**

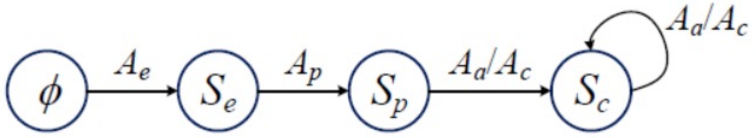


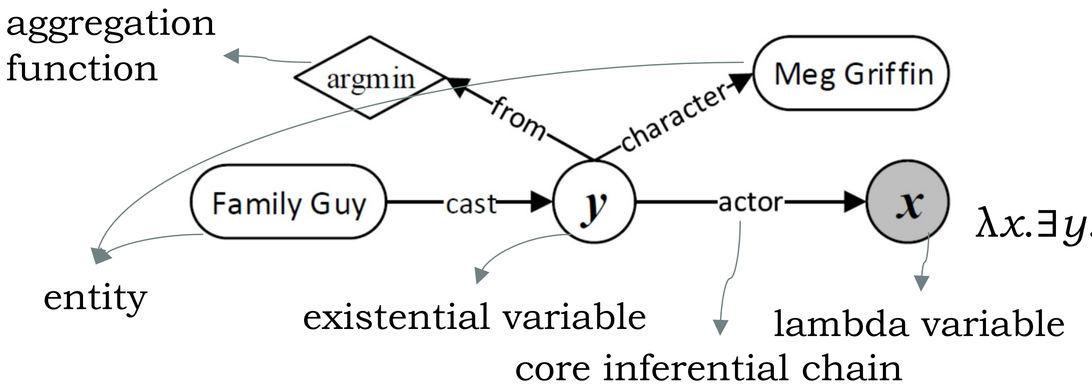
"Who first voiced Meg on Family Guy?"

$\lambda x.\exists y.\texttt{cast}(\text{FamilyGuy}, y) \wedge \texttt{actor}(y, x) \wedge \texttt{character}(y, \text{MegGriffin})$

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Method:
  - 图生成分三步：
    - 定位问题中的主题实体
    - 发现答案和主题实体之间的主要关系
    - 使用描述答案需要的属性的附加约束或问题中答案与其他实体之间的关系来扩展查询图
  - 图的中间状态：S=U{Ø,Se,Sp,Sc}
    - Se: 有主题节点的单一节点图，Sp:核心推理链，Sc:额外的约束
  - 图的生成动作：A=U{Ae,Ap,Ac,Aa}
    - Ae:选择一个实体节点，Ap:决定核心推理链，Ac和Aa:添加约束和聚合节点





"Who first voiced Meg on Family Guy?"

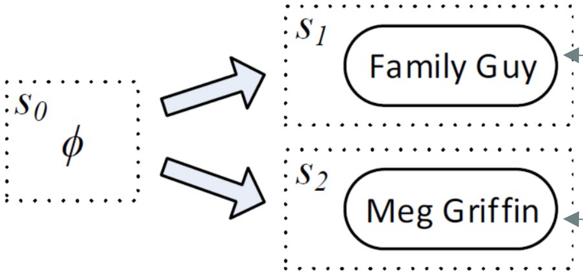
$\lambda x.\exists y.\texttt{cast}(\text{FamilyGuy},y) \wedge \texttt{actor}(y,x) \wedge \texttt{character}(y,\text{MegGriffin})$

核心推理链：主题实体和答案的关系

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Method:

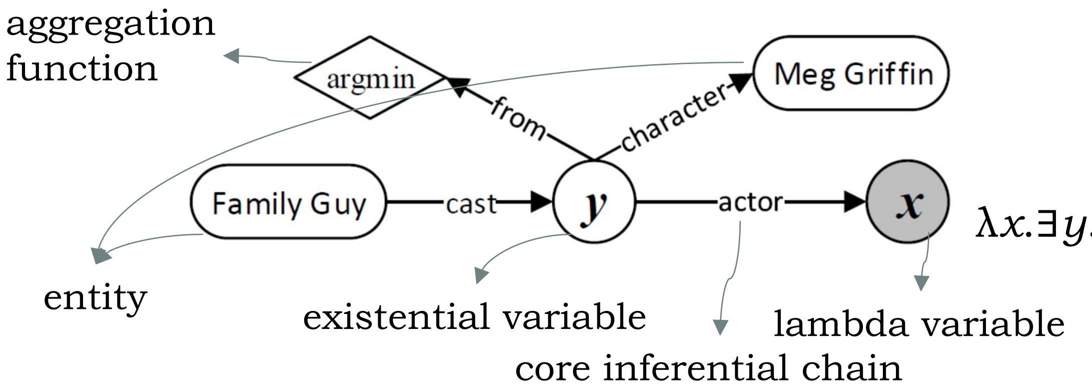
Step1: Linking Topic Entity



entity linking

Figure 4: Two possible topic entity linking actions applied to an empty graph, for question "Who first voiced [Meg] on [Family Guy]?"

"Who first voiced Meg on Family Guy?"



$\lambda x.\exists y.\mathtt{cast}(\text{FamilyGuy},y)\wedge\mathtt{actor}(y,x)\wedge\mathtt{character}(y,\text{MegGriffin})$

核心推理链：主题实体和答案的关系

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Method:

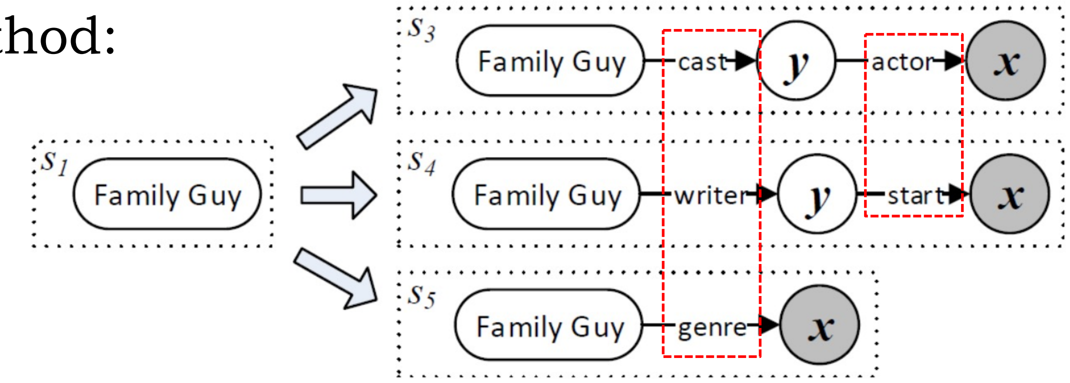


Figure 5: Candidate core inferential chains start from the entity `FamilyGuy`.

Step2: Identifying Core Inferential Chain

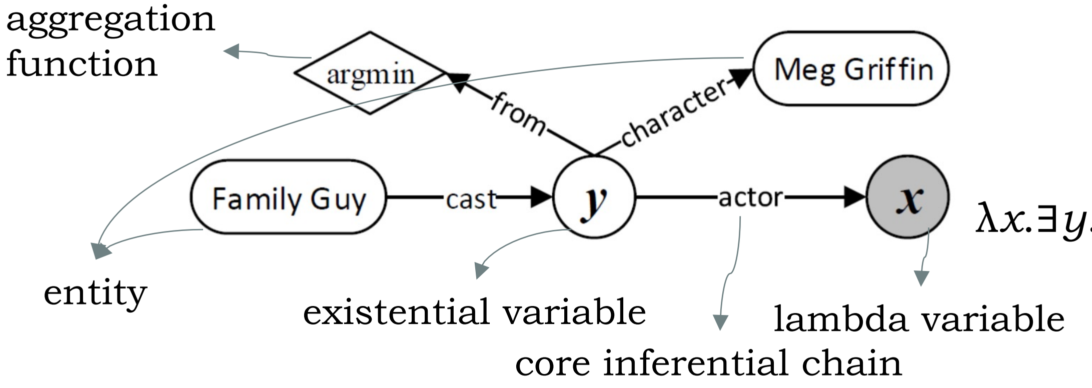
(1) 从e开始，扩展出**合法谓词**序列

(2) 用Siamese NN计算每种序列和问题的相似度

sim(<*Family Guy*, *s*>, `cast-actor`)

sim(<*Family Guy*, *s*>, `writer-start`)

sim(<*Family Guy*, *s*>, `genre`)



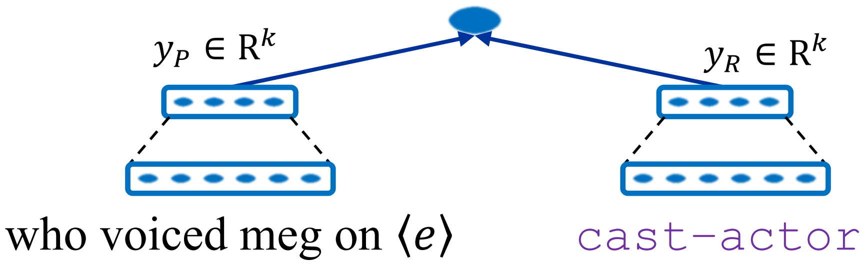
"Who first voiced Meg on Family Guy?"

$\lambda x.\exists y.\texttt{cast}(\text{FamilyGuy},y) \wedge \texttt{actor}(y,x) \wedge \texttt{character}(y,\text{MegGriffin})$

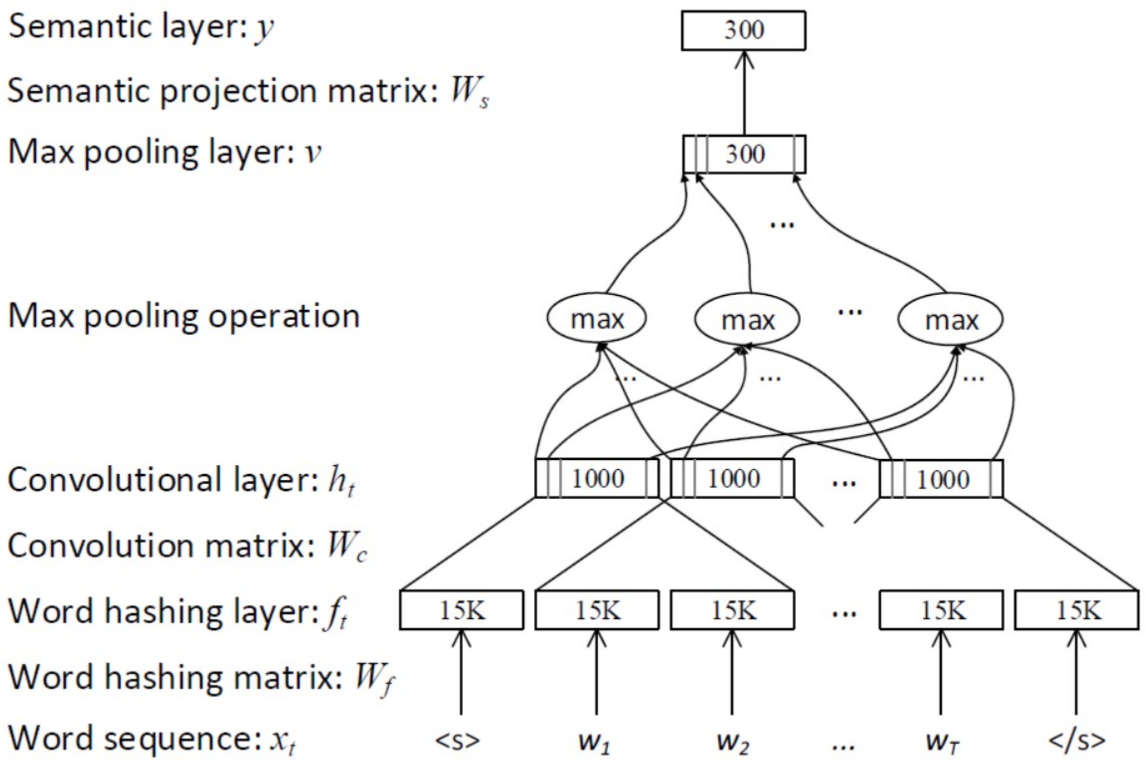
核心推理链：主题实体和答案的关系

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Method:
  - Input:
    - NL Model: *Who first voiced meg on <e>*
      - `Who` → `#-w-h, w-h-o, h-o-#`
    - Predicate Chain Model: `cast-actor`
  - Output:
    - continuous-space representation
  - sim(*s*, `cast-actor`): cosine-distance



Step2: Identifying Core Inferential Chain

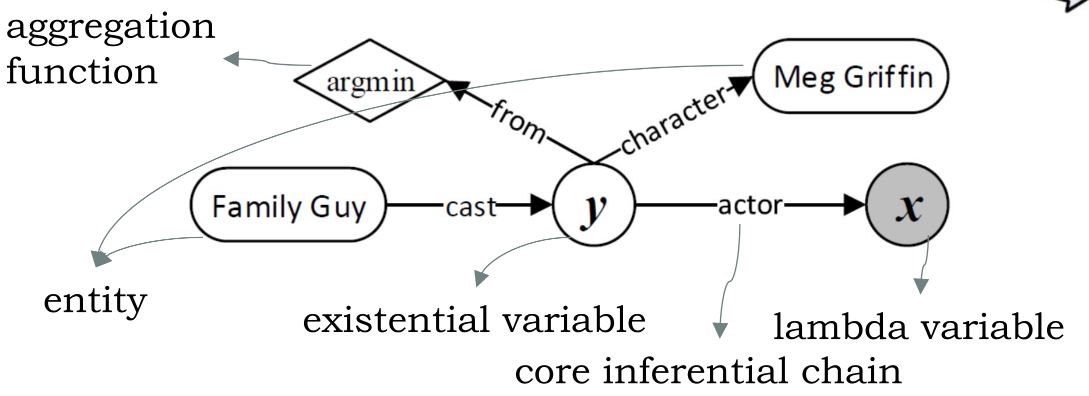


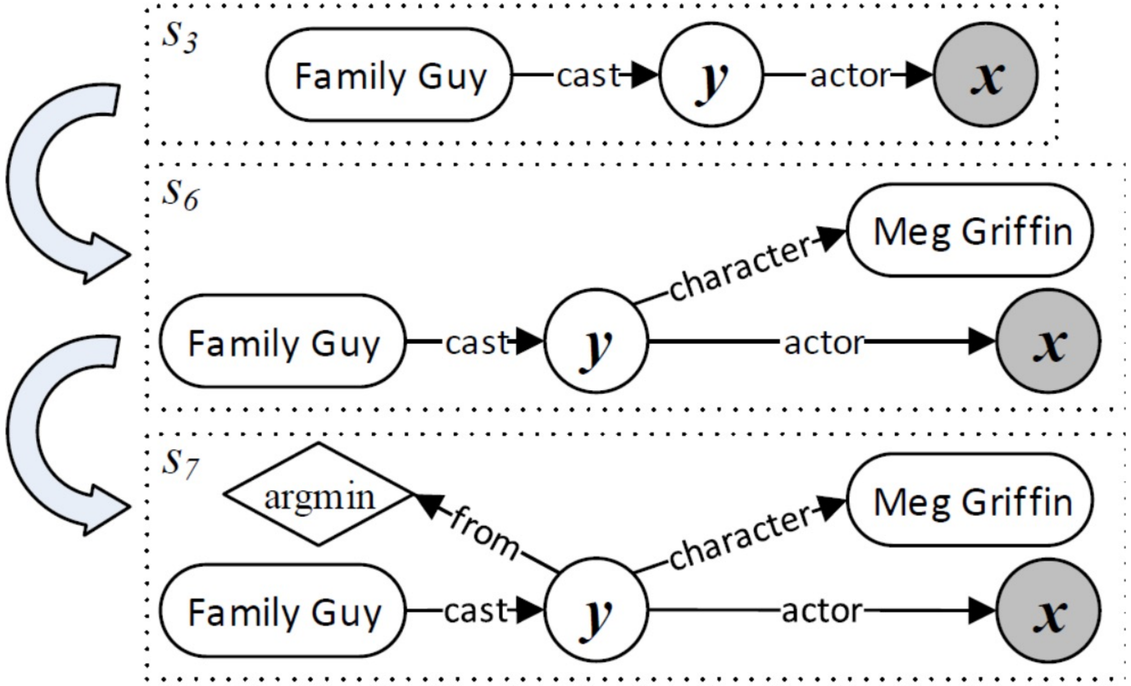
Goal: sim(<Family Guy, s>, cast-actor*)=?*

Semantic parsing via staged query graph generation: Question answering with knowledge base

Step3: Augmenting Constraints & Aggregations

- Method:
- 为了检索答案实体集，拥有核心推理链的图只可以被两种类型的动作扩展：Ac和Aa
  - Ac是**连接一个实体到变量节点可能方式**的集合，边表示合法谓词，如s6 所示
  - Aa**通过聚合函数在整个答案集中描述约束**，在一个变量节点中连接一个聚合节点，如s7所示
- 通过核心推理链找到y和x节点的邻居节点，以此获得全部的约束集。本工作做了基于规则的简化处理





*“Who **first** voiced **Meg** on Family Guy?”*

Semantic parsing via staged query graph generation: Question answering with knowledge base

- Method
  - Credit(*sentence*, *graph*) = log-linear(Step1, Step2, Step3)
- Experiment
  - WebQuestions数据集，包括5810个问答对，其中65%作为训练集，35%作为测试集

Pattern	Inferential Chain
what was <e> known for	people.person.profession
what kind of government does <e> have	location.country.form_of_government
what year were the <e> established	sports.sports_team.founded
what city was <e> born in	people.person.place_of_birth
what did <e> die from	people.deceased_person.cause_of_death
who married <e>	people.person.spouse_s people.marriage.spouse

Method	Prec.	Rec.	F_1
(Berant et al., 2013)	48.0	41.3	35.7
(Bordes et al., 2014b)	-	-	29.7
(Yao and Van Durme, 2014)	-	-	33.0
(Berant and Liang, 2014)	40.5	46.6	39.9
(Bao et al., 2014)	-	-	37.5
(Bordes et al., 2014a)	-	-	39.2
(Yang et al., 2014)	-	-	41.3
(Wang et al., 2014)	-	-	45.3
Our approach – STAGG	52.8	60.7	**52.5**

Recommended References

- Yih S W, Chang M W, He X, et al. Semantic parsing via staged query graph generation: Question answering with knowledge base. In *Proceedings of the Joint Conference of the 53rd Annual Meeting of the ACL and the 7th International Joint Conference on Natural Language Processing of the AFNLP*. 2015. (*Outstanding Paper*)
- Yao, Xuchen, and Benjamin Van Durme. "Information extraction over structured data: Question answering with freebase." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2014.
- Lukovnikov, Denis, et al. "Neural network-based question answering over knowledge graphs on word and character level." *Proceedings of the 26th international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017.
- Abujabal, Abdalghani, et al. "Automated template generation for question answering over knowledge graphs." *Proceedings of the 26th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 2017.
- Bao, Junwei, et al. "Constraint-based question answering with knowledge graph." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.
- Diefenbach, Dennis, et al. "Core techniques of question answering systems over knowledge bases: a survey." *Knowledge and Information systems* 55.3 (2018): 529-569.
- Unger, Christina, et al. "Template-based question answering over RDF data." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.
- Gao, Jianfeng , M. Galley , and L. Li . " [ACM Press The 41st International ACM SIGIR Conference - Ann Arbor, MI, USA (2018.07.08-2018.07.12)] The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, - SIGIR '18 - Neural Approaches to Conversational AI." (2018):1371-1374.
- Yih, Wen-tau, et al. "The value of semantic parse labeling for knowledge base question answering." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vol. 2. 2016.
- Yih, Wen-tau, Xiaodong He, and Christopher Meek. "Semantic parsing for single-relation question answering." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vol. 2. 2014.
- Antoine Bordes, et al. "Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing." AISTATS(2012)

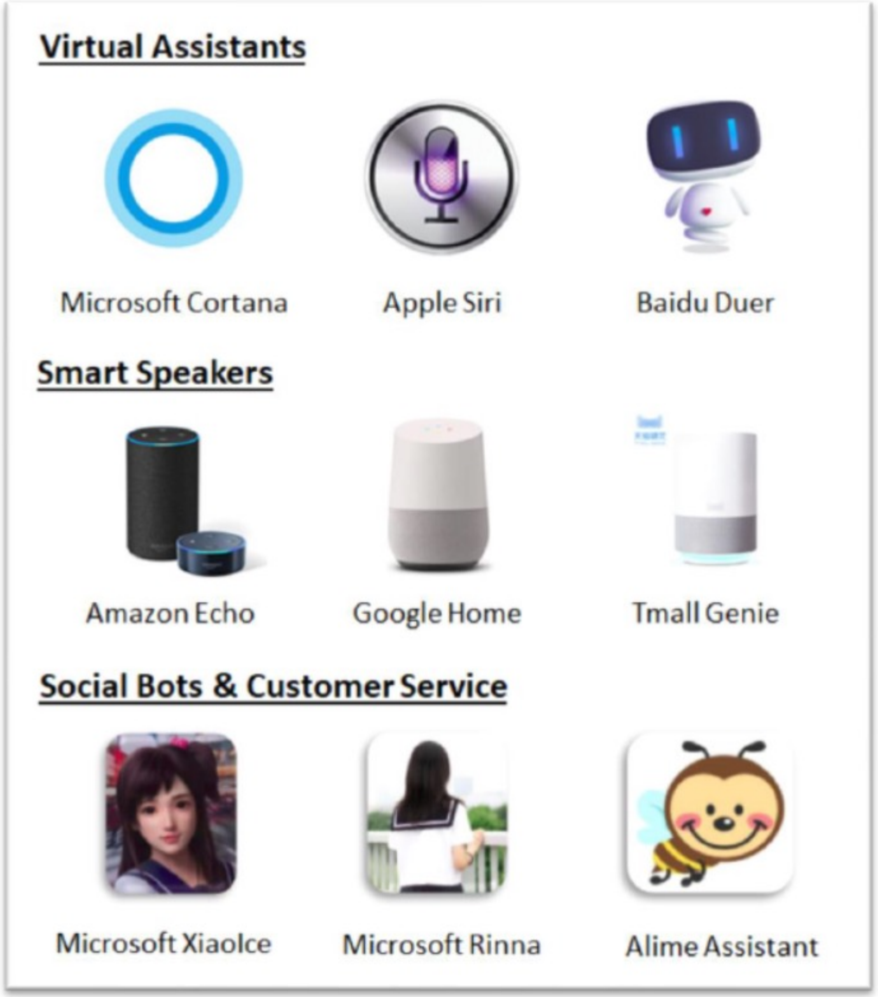
4. Generation-based Chatbot

问答系统

分析方法	答案形态	客观性	轮次	领域
检索	FAQ	事实性	单轮	开放
NLU	结构化数据	非事实	多轮	垂直
End-to-End	多文档抽取	情感性		
	单文档抽取			
	自然语言生成			

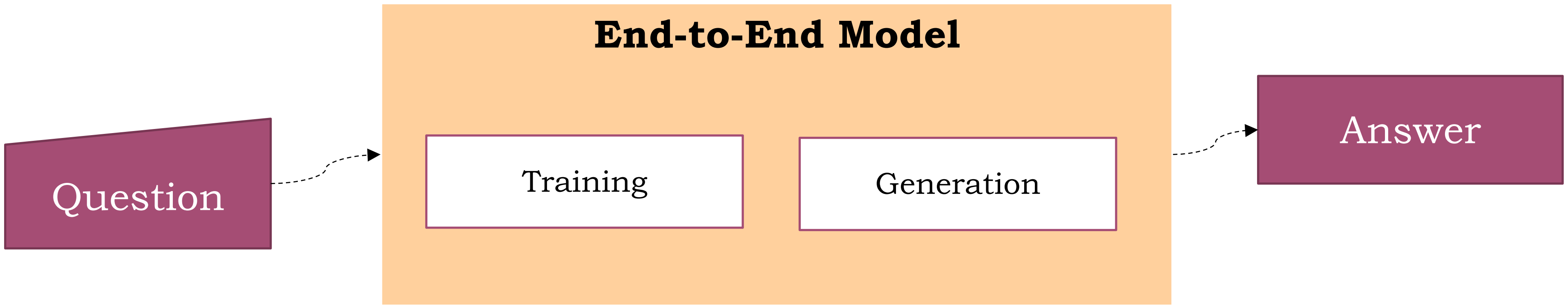
QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

4. Generation-based Chatbot



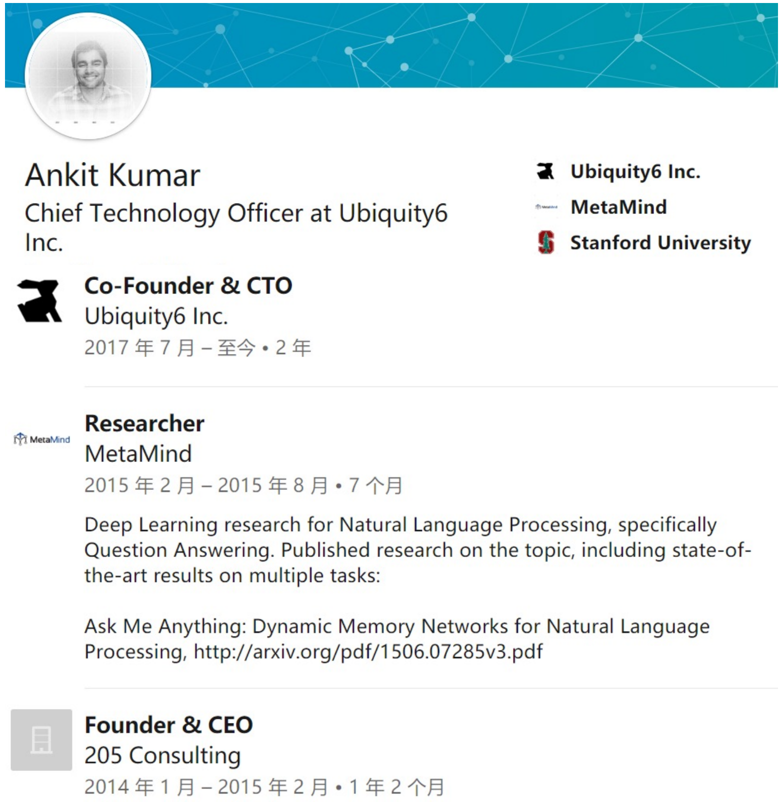
- 优点：领域开放，内容多样性，情感性
- 缺点：很难实现任务型对话

4. Generation-based Chatbot



Ask Me Anything: Dynamic Memory Networks for NLP

- 来自斯坦福博士毕业生 Richard Socher 的创业公司 Salesforce (Metamind)，第一作者是实习生**Ankit Kumar**，文章发表于ICML 2016 (Citation: 575)



Ankit Kumar
Chief Technology Officer at Ubiquity6 Inc.

Ubiquity6 Inc.
MetaMind
Stanford University

Co-Founder & CTO
Ubiquity6 Inc.
2017 年 7 月 – 至今 • 2 年

Researcher
MetaMind
2015 年 2 月 – 2015 年 8 月 • 7 个月

Deep Learning research for Natural Language Processing, specifically Question Answering. Published research on the topic, including state-of-the-art results on multiple tasks:

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, http://arxiv.org/pdf/1506.07285v3.pdf

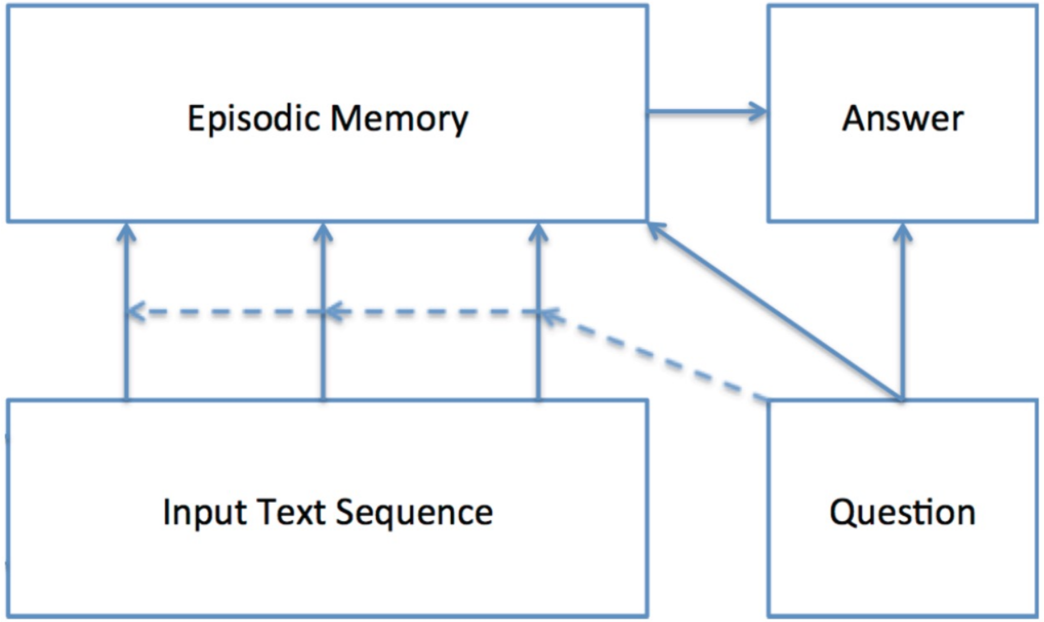
Founder & CEO
205 Consulting
2014 年 1 月 – 2015 年 2 月 • 1 年 2 个月

Richard Socher，被世界经济论坛誉为“实现改变现有自然语言处理与计算机视觉技术突破的、人工智能和深度学习领域的天才”
他还有两个身份，MetaMind创始人，Salesforce现任首席科学家

Ask Me Anything: Dynamic Memory Networks for NLP

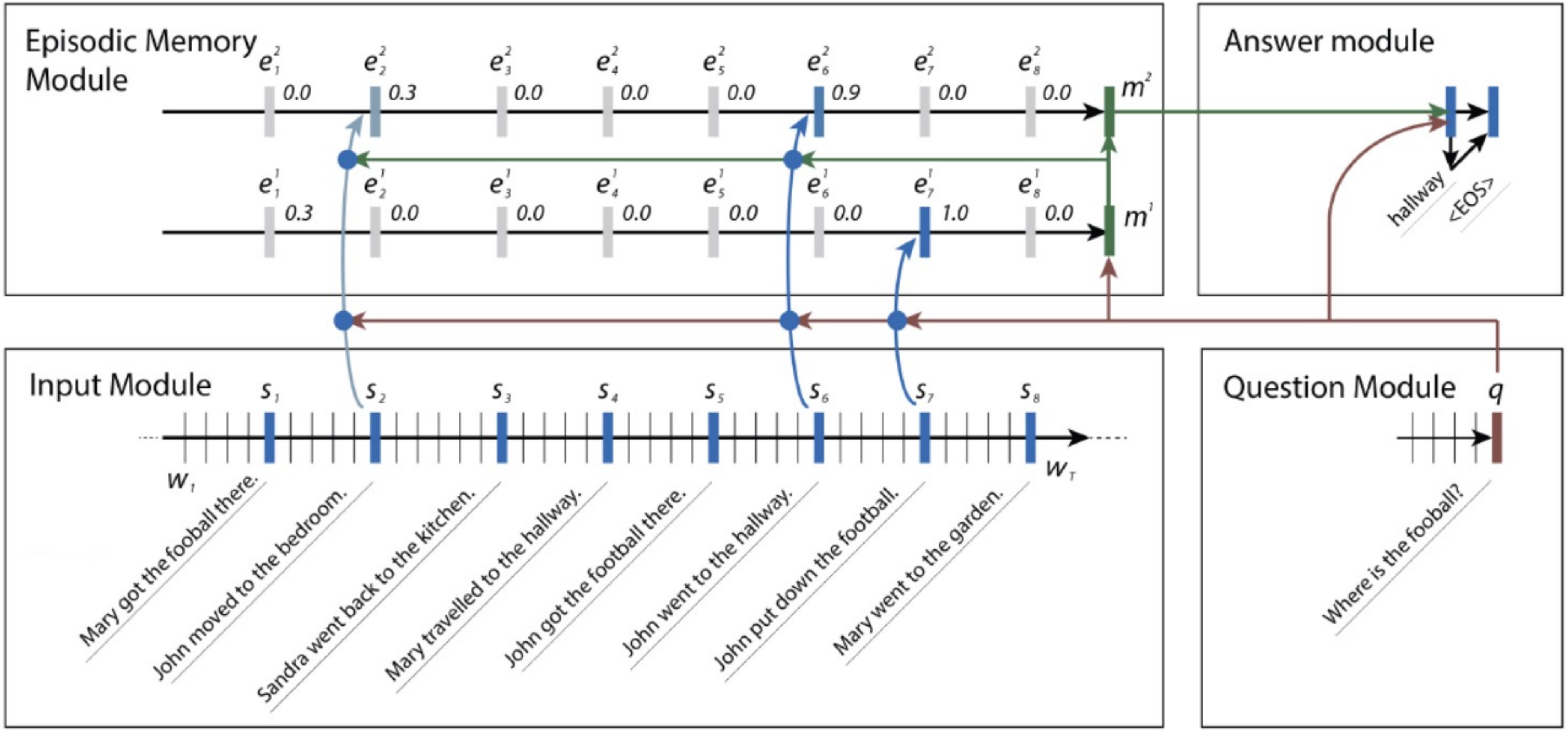
- DMN基本架构
- 将机器翻译、NER、情感分析、共指消解、词性标注看作问答，则也可以用DMN建模

I:	Jane went to the hallway.
I:	Mary walked to the bathroom.
I:	Sandra went to the garden.
I:	Daniel went back to the garden.
I:	Sandra took the milk there.
Q:	Where is the milk?
A:	garden
I:	It started boring, but then it got interesting.
Q:	What's the sentiment?
A:	positive
Q:	POS tags?
A:	PRP VBD JJ , CC RB PRP VBD JJ .



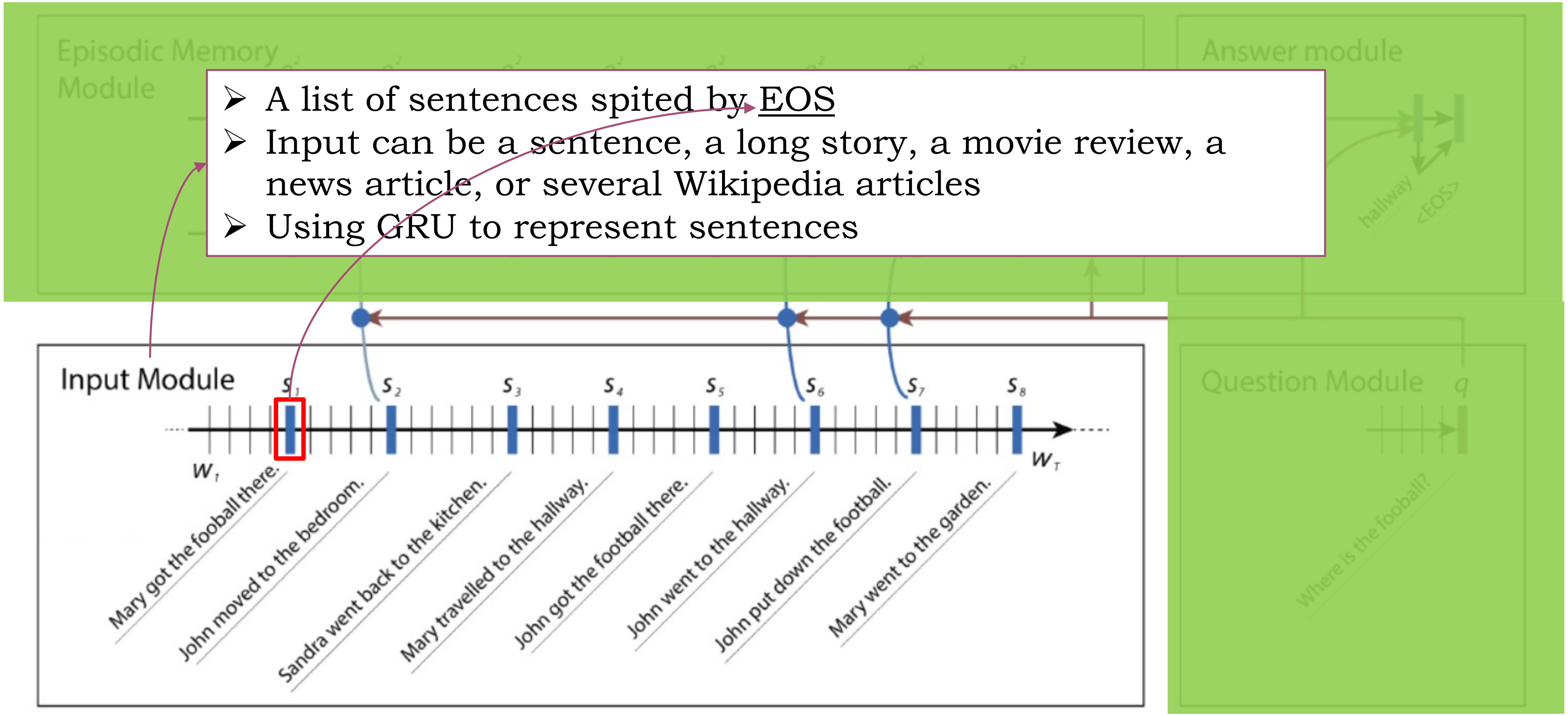
DMN流程：
(1) 表示输入和问题
(2) 根据问题迭代地检索相关facts
(3) 由memory模块根据facts和问题进行推理，给出相关的信息
(4) 由回答模块将相关信息转化成答案

Ask Me Anything: Dynamic Memory Networks for NLP

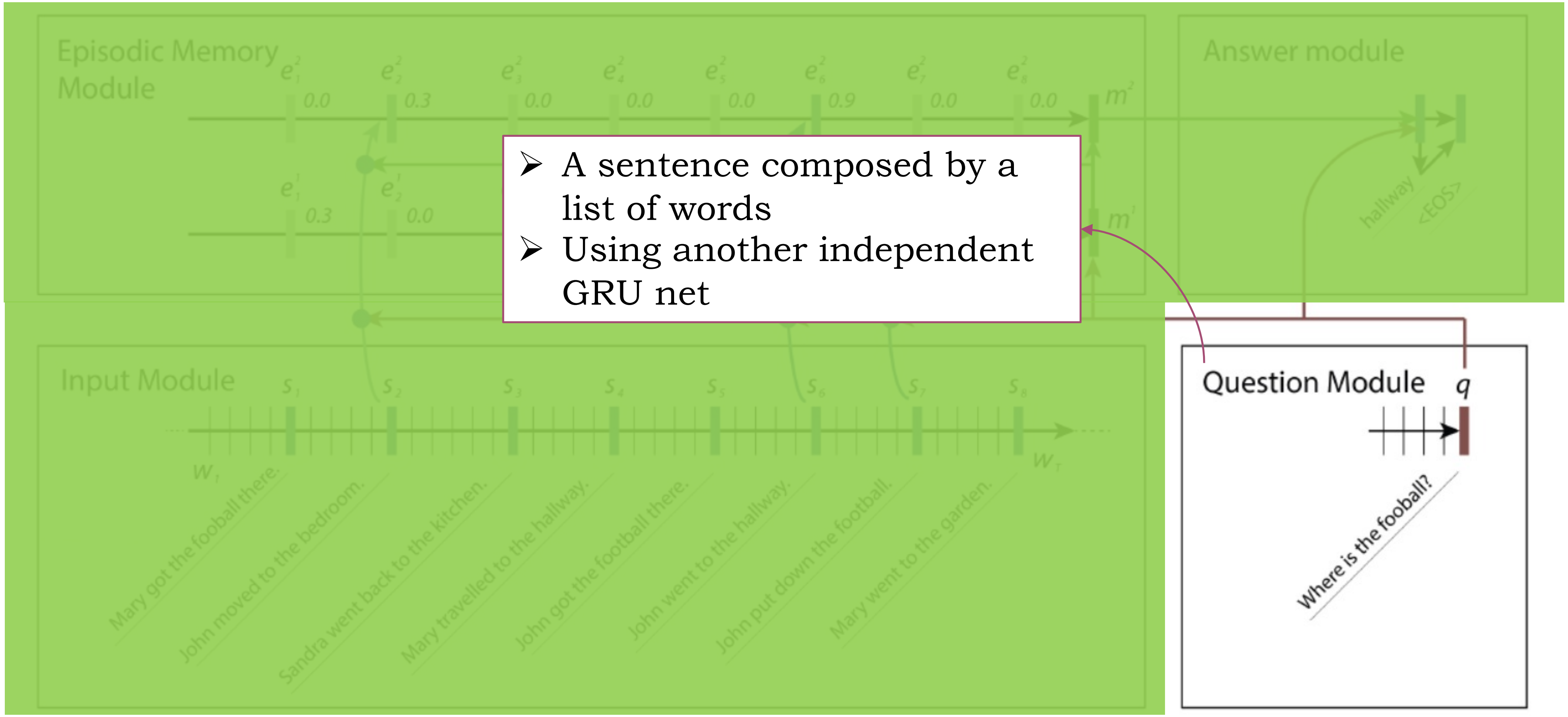


Ask Me Anything: Dynamic Memory Networks for NLP

- A list of sentences spited by EOS
- Input can be a sentence, a long story, a movie review, a news article, or several Wikipedia articles
- Using GRU to represent sentences

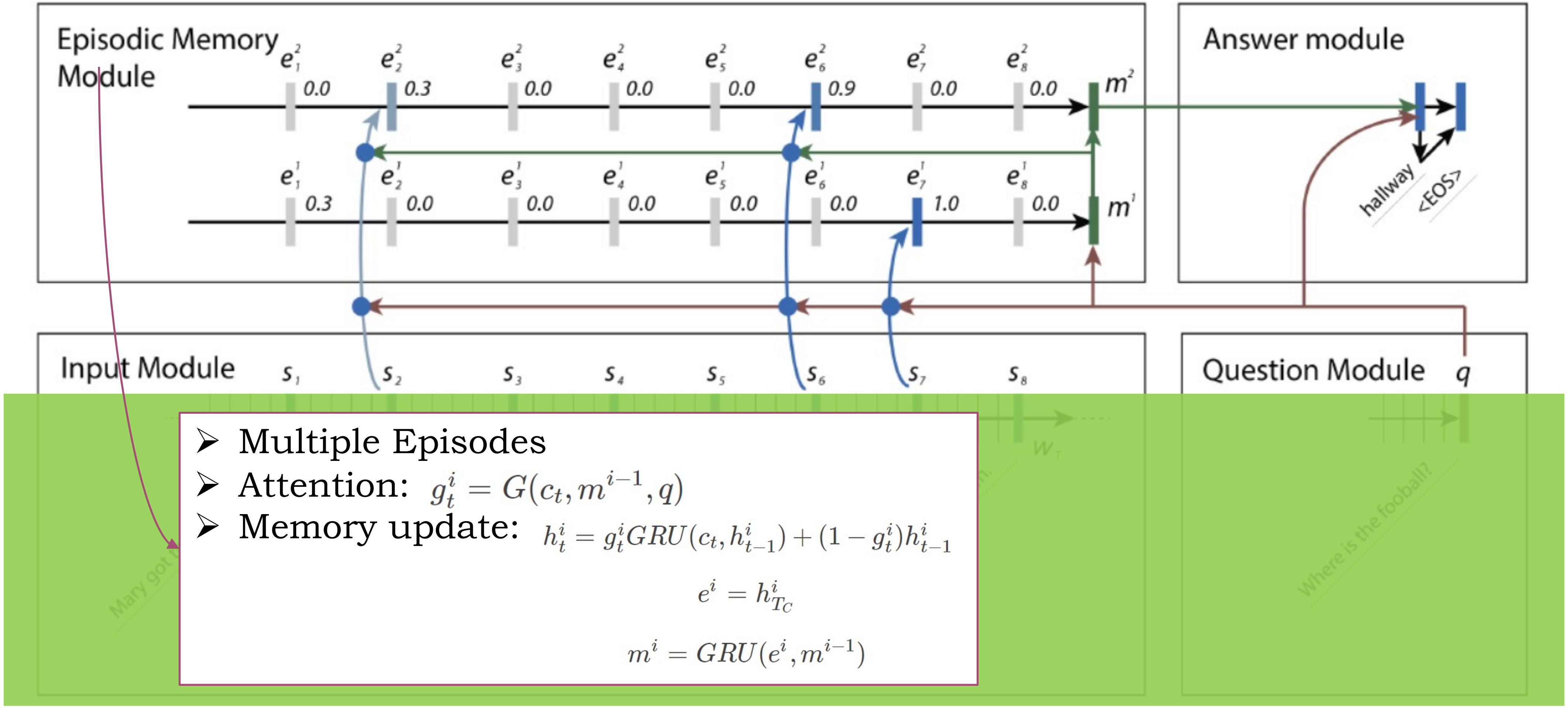


Ask Me Anything: Dynamic Memory Networks for NLP



- A sentence composed by a list of words
- Using another independent GRU net

Ask Me Anything: Dynamic Memory Networks for NLP

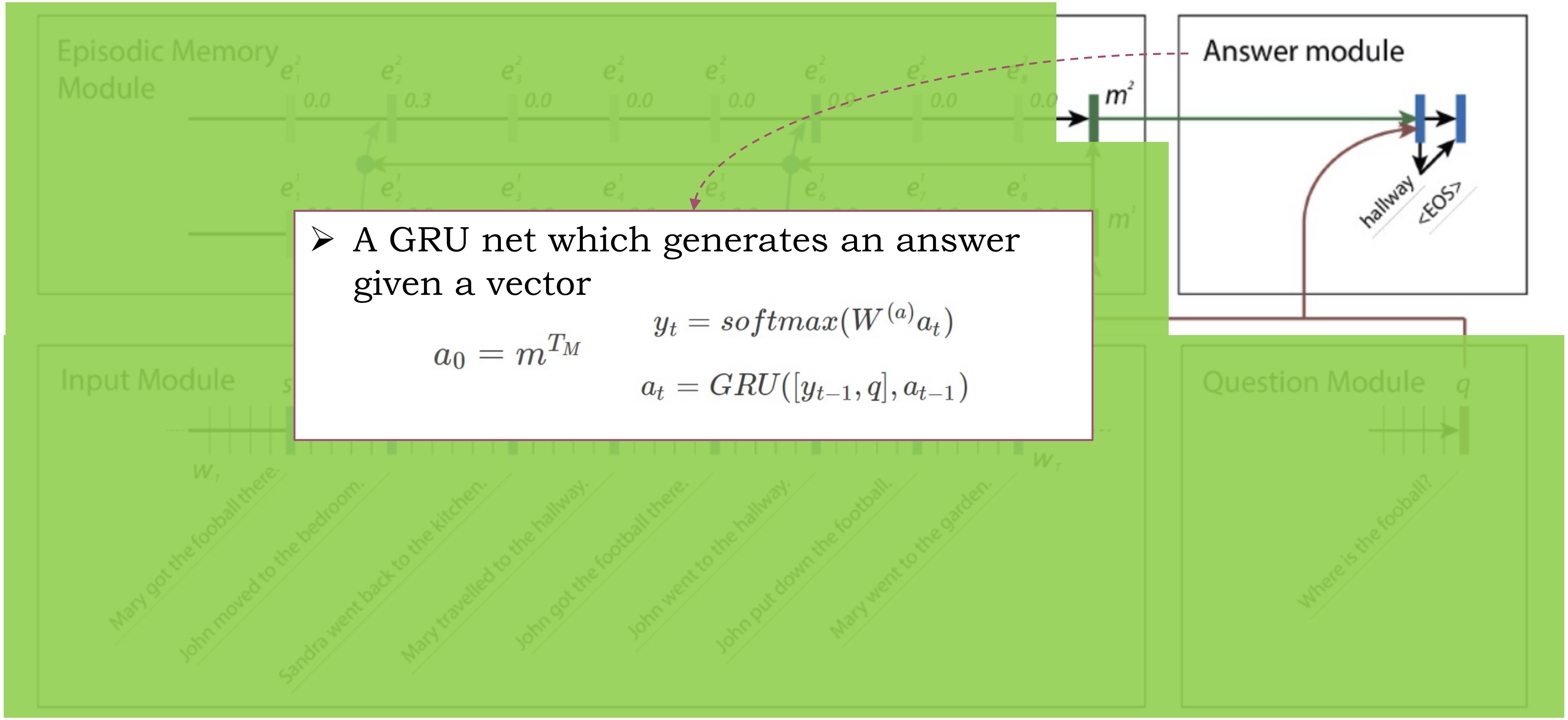


- Multiple Episodes
- Attention: $g_t^i = G(c_t, m^{i-1}, q)$
- Memory update: $h_t^i = g_t^i GRU(c_t, h_{t-1}^i) + (1 - g_t^i) h_{t-1}^i$

$$e^i = h^i_{T_C}$$

$$m^i = GRU(e^i, m^{i-1})$$

Ask Me Anything: Dynamic Memory Networks for NLP



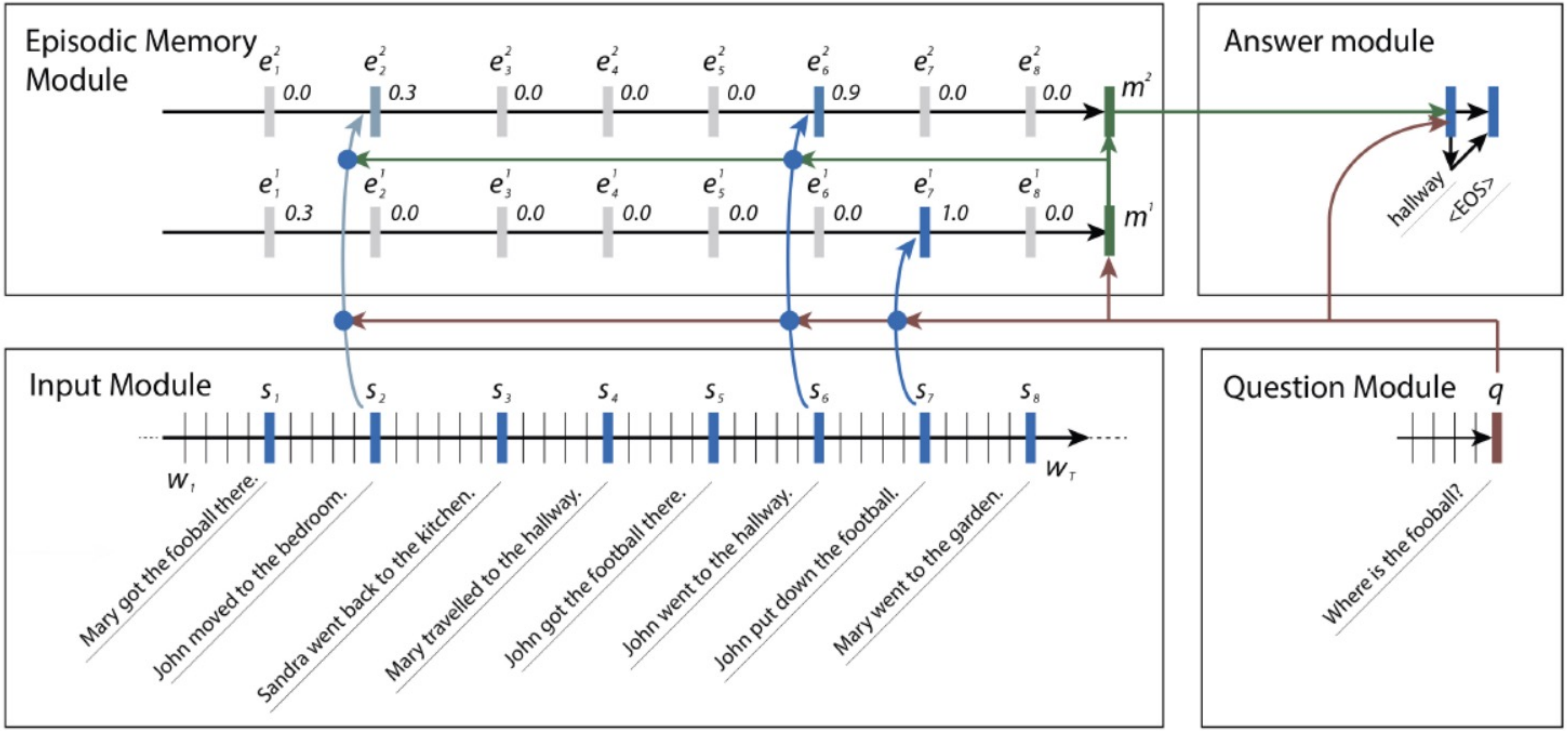
- A GRU net which generates an answer given a vector

$$a_0 = m^{T_M}$$

$$y_t = softmax(W^{(a)} a_t)$$

$$a_t = GRU([y_{t-1}, q], a_{t-1})$$

Ask Me Anything:
Dynamic Memory Networks for NLP



Ask Me Anything: Dynamic Memory Networks for NLP

Task	MemNN	DMN
1: Single Supporting Fact	100	100
2: Two Supporting Facts	100	98.2
3: Three Supporting Facts	100	95.2
4: Two Argument Relations	100	100
5: Three Argument Relations	98	99.3
6: Yes/No Questions	100	100
7: Counting	85	96.9
8: Lists/Sets	91	96.5
9: Simple Negation	100	100
10: Indefinite Knowledge	98	97.5
11: Basic Coreference	100	99.9
12: Conjunction	100	100
13: Compound Coreference	100	99.8
14: Time Reasoning	99	100
15: Basic Deduction	100	100
16: Basic Induction	100	99.4
17: Positional Reasoning	65	59.6
18: Size Reasoning	95	95.3
19: Path Finding	36	34.5
20: Agent's Motivations	100	100
Mean Accuracy (%)	93.3	**93.6**

QA accuracies on the bAbI dataset

Facts	Episode 1	Episode 2	Episode 3
Yesterday Julie traveled to the school.			
Yesterday Marie went to the cinema.			
This morning Julie traveled to the kitchen.			
Bill went back to the cinema yesterday.			
Mary went to the bedroom this morning.			
Julie went back to the bedroom this afternoon.			
[done reading]			

Question: *Where was Mary before the Bedroom?*
Answer: *Cinema.*
Darker colors mean that the **attention weight** is higher

Task	Binary	Fine-grained
MV-RNN	82.9	44.4
RNTN	85.4	45.7
DCNN	86.8	48.5
PVec	87.8	48.7
CNN-MC	88.1	47.4
DRNN	86.6	49.8
CT-LSTM	88.0	51.0
DMN	**88.6**	**52.1**

sentiment analysis on Stanford Sentiment Treebank

Model	Acc (%)
SVMTool	97.15
Sogaard	97.27
Suzuki et al.	97.40
Spoustova et al.	97.44
SCNN	97.50
DMN	**97.56**

POS Tagging on WSJ-PTB

Recommended References

- Sukhbaatar, Sainbayar, Jason Weston, and Rob Fergus. "End-to-end memory networks." *Advances in neural information processing systems*. 2015.
- Junwei Bao, Nan Duan, Ming Zhou, and Tiejun Zhao. 2014. Knowledge-based question answering as machine translation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 967–976, Baltimore, Maryland, June. Association for Computational Linguistics.
- Wei Wu Chen Xing Yu Wu Ming Zhou Yalou Huang Wei-Ying Ma. Hierarchical Recurrent Attention Network for Response Generation. In Thirty-Second AAAI Conference on Artificial Intelligence (AAAI'18). February 2018
- Wei Wu (wuwei) Yu Wu Dejian Yang Can Xu Zhoujun Li. Neural Response Generation with Dynamic Vocabularies. In Thirty-Second AAAI Conference on Artificial Intelligence (AAAI'18). February 2018.
- Qiu, Minghui, et al. "Alime chat: A sequence to sequence and rerank based chatbot engine." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2017.
- Cui, Lei, et al. "Superagent: A customer service chatbot for e-commerce websites." *Proceedings of ACL 2017, System Demonstrations* (2017): 97-102.
- Bordes, Antoine, Y-Lan Boureau, and Jason Weston. "Learning end-to-end goal-oriented dialog." *arXiv preprint arXiv:1605.07683* (2016).
- Serban, Iulian V., et al. "Building end-to-end dialogue systems using generative hierarchical neural network models." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- Yang, Yi, Wen-tau Yih, and Christopher Meek. "Wikiqa: A challenge dataset for open-domain question answering." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015.
- Kumar, Ankit, et al. "Ask me anything: Dynamic memory networks for natural language processing." *International conference on machine learning*. 2016.

5. Machine Comprehension

问答系统

- 分析方法
  - 检索
  - NLU
  - End-to-End
- 答案形态
  - FAQ
  - 结构化数据
  - 多文档抽取
  - 单文档抽取
  - 自然语言生成
- 客观性
  - 事实性
  - 非事实
  - 情感性
- 轮次
  - 单轮
  - 多轮
- 领域
  - 开放
  - 垂直

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

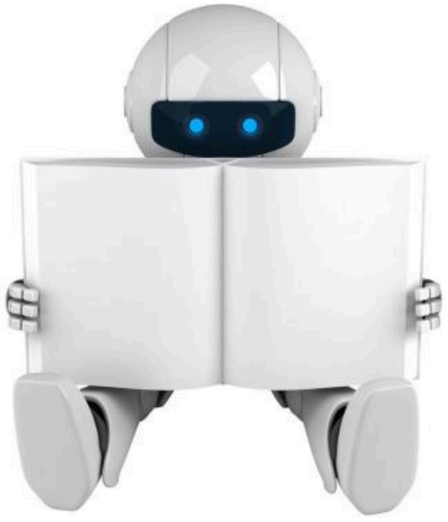
5. Machine Comprehension

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She's now in Miami. She went to Miami to visit some friends. But she wanted some time to herself at the beach, so she went there first. After going swimming and laying out, she went to her friend **Ellen**'s house. **Ellen** greeted **Alyssa** and they both had some lemonade to drink. **Alyssa** called her friends **Kristen** and **Rachel** to meet at **Ellen**'s house. The girls traded stories and caught up on their lives. It was a happy time for everyone. The girls went to a restaurant for dinner. The restaurant had a special on catfish. **Alyssa** enjoyed the restaurant's special. **Ellen** ordered a salad. **Kristen** had soup. **Rachel** had a steak. After eating, the ladies went back to **Ellen**'s house to have fun. They had lots of fun. They stayed the night because they were tired. **Alyssa** was happy to spend time with her friends again.

(a) **Question:** What city is Alyssa in?
Answer: Miami

(b) **Question**: What did Alyssa eat at the restaurant?
Answer: catfish

(c) **Question**: How many friends does Alyssa have in this story?
Answer: 3



- 需要对文本的深入理解：词法、句法、语义、篇章
- 需要融合知识
- 需要有推理能力

Squad : The **S**tanford **Qu**estion **A**nswering **D**ataset

- 根据维基百科，人工产生一些问题，并且在原文中标出问题的答案

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".

What causes precipitation to fall?
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

- 维基百科文档
- 每个文档有4-8个段落
- 每个段落包含有3-5个问题
- 总共107,785个问题-答案对
- 基于事实性的问题比较多

Squad : The **S**tanford **Qu**estion **A**nswering **D**ataset

SQuAD | Home | Explore 2.0 | Explore 1.1

SQuAD2.0

The Stanford Question Answering Dataset

What is SQuAD?

Stanford **Qu**estion **A**nswering **D**ataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.

New **SQuAD2.0** combines the 100,000 questions in SQuAD1.1 with over 50,000 new, unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering. SQuAD2.0 is a challenging natural language understanding task for existing models, and we release SQuAD2.0 to the community as the successor to SQuAD1.1. We are optimistic that this new dataset will encourage the development of reading comprehension systems that know what they don't know.

Explore SQuAD2.0 and model predictions

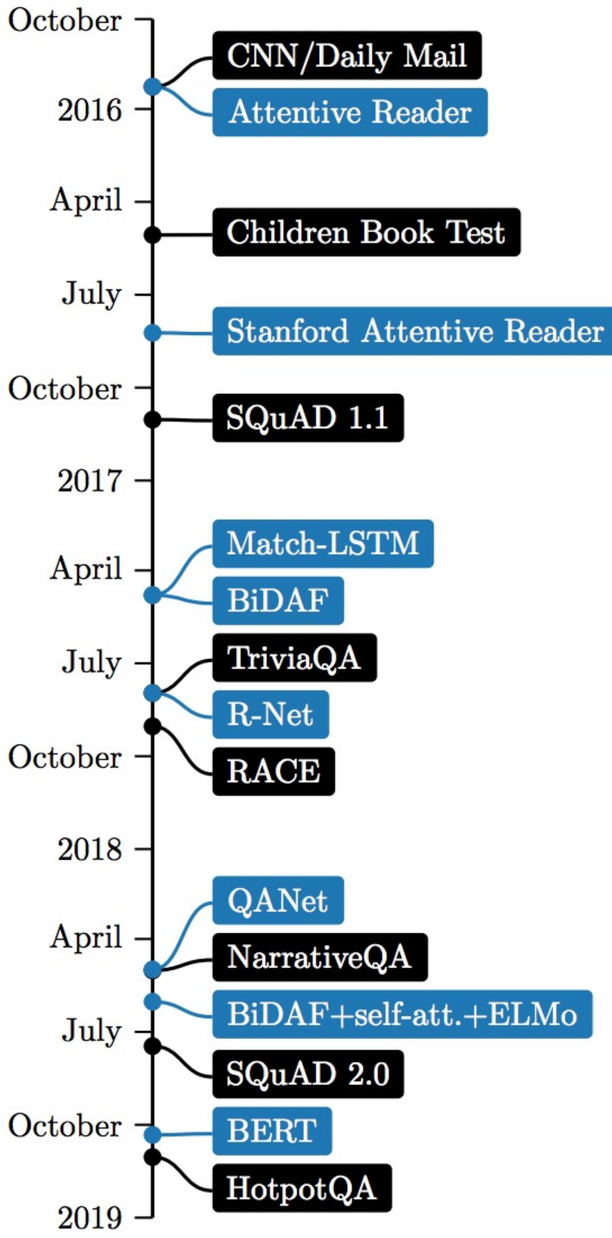
Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph. How will your system compare to humans on this task?

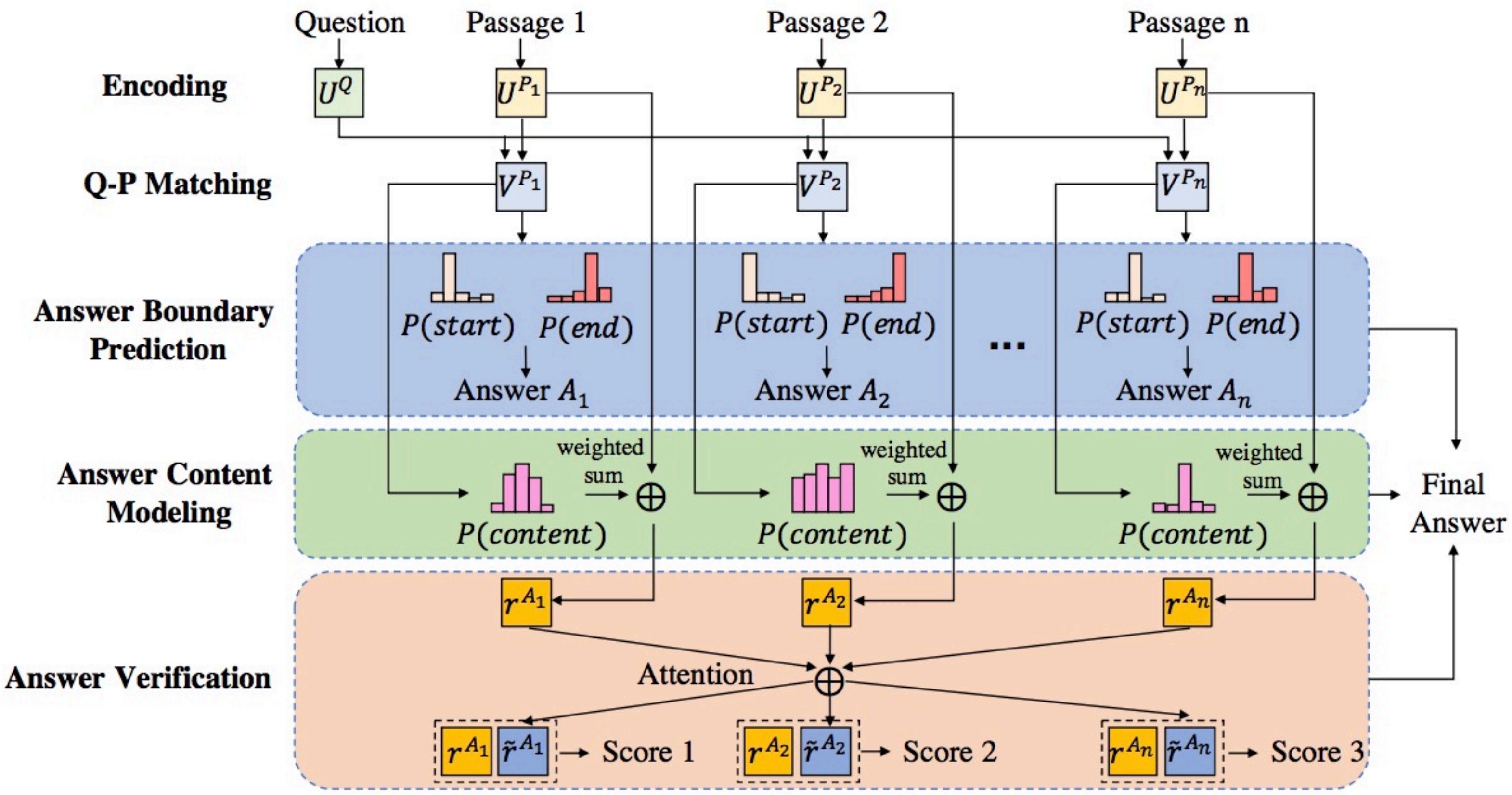
Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Mar 20, 2019	BERT + DAE + AoA (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	**87.147**	**89.474**
2 Mar 15, 2019	BERT + ConvLSTM + MTL + Verifier (ensemble) *Layer 6 AI*	86.730	89.286
3 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert	86.673	89.147
4 May 21, 2019	XLNet (single model) *XLNet Team*	86.346	89.133
5	SemBERT(ensemble)	86.166	88.886

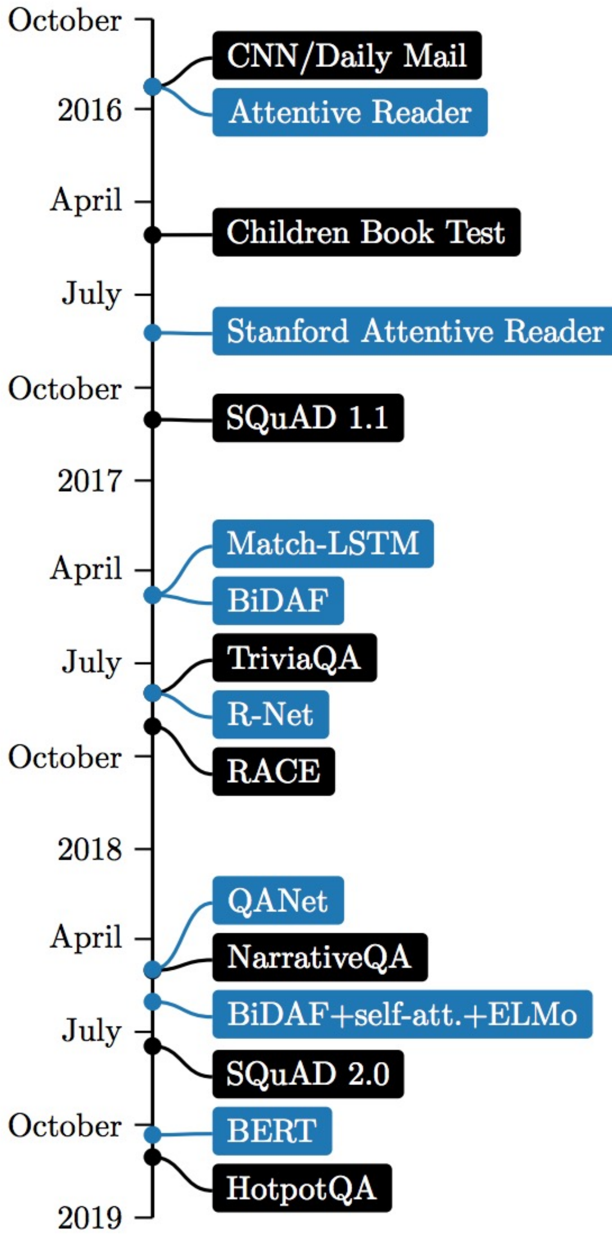
Other Dataset

数据集	语料来源	问题规模	文档规模	答案形式
CoQA	多源	127k	8000+	人工编写
SQuAD v2 (斯坦福)	Wikipedia	150k	500+	文中片段（span）
NarrativeQA (Google DeepMind)	Wikipedia等	46765	1572	问题和答案人工编写
RACE (CMU)	中学英语试题	100k	28k	单选题
MS MARCO (微软)	Bing搜索	100k	200k	摘要（不在原文）
SQuAD (斯坦福)	Wikipedia	100k	500+	文中片段（span）
CBT (Facebook)	儿童故事	687k	108	多选题
CNN|Daily Mail (Google DeepMind)	新闻	1.4M	200k+	实体完形填空



NN-based Architecture

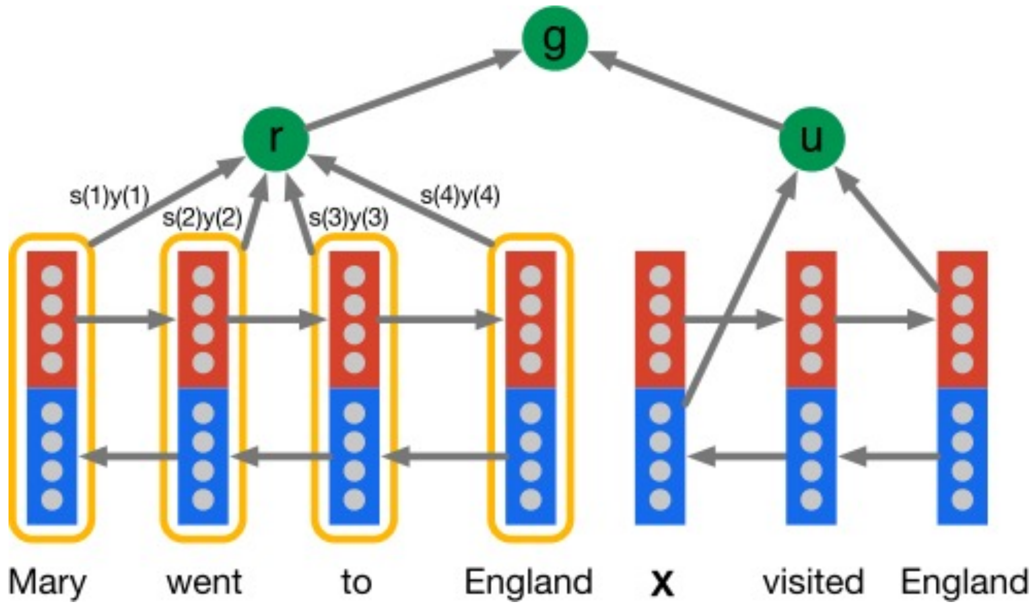




- 表示层：问题编码，文档编码
- 交互层：问题答案互相表示
- 预测层：分类预测起止位置（抽取式）
- 交叉验证层：答案级别的attention，对答案聚合加权

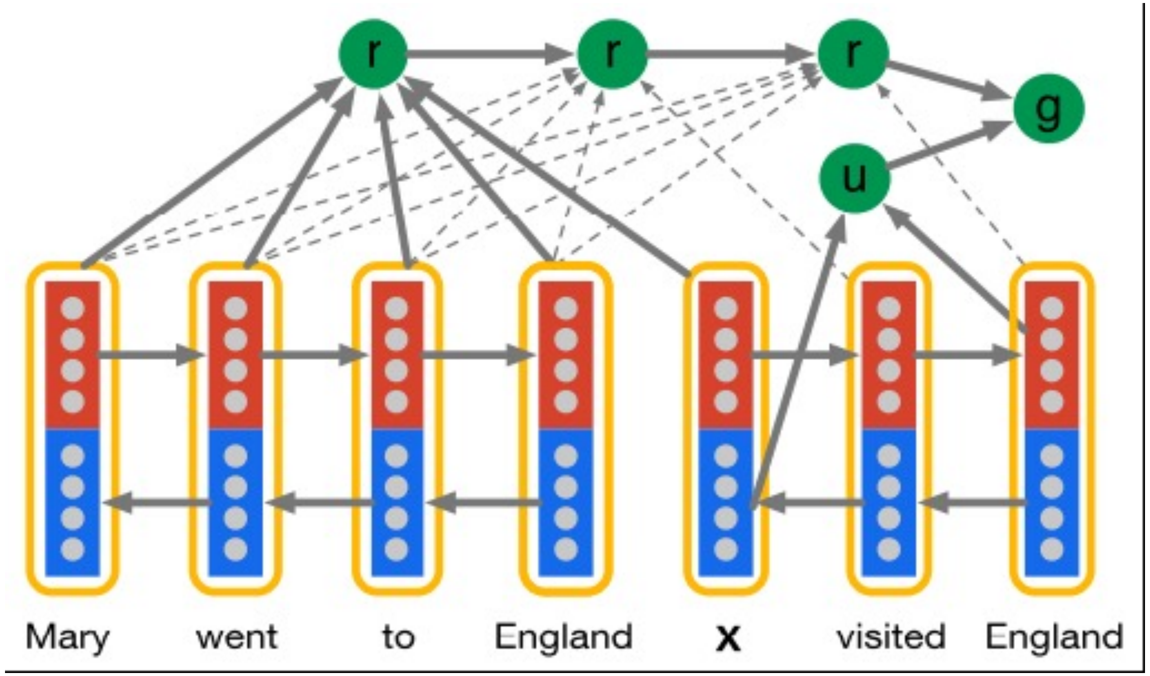
Attentive Reader

- 在表示背景文档的时候，针对不同的问题，文档的表示也随之动态的变化，这就是注意力机制（Attention Mechanism）。
- 在整个问句表示完毕，才进行关注过程。



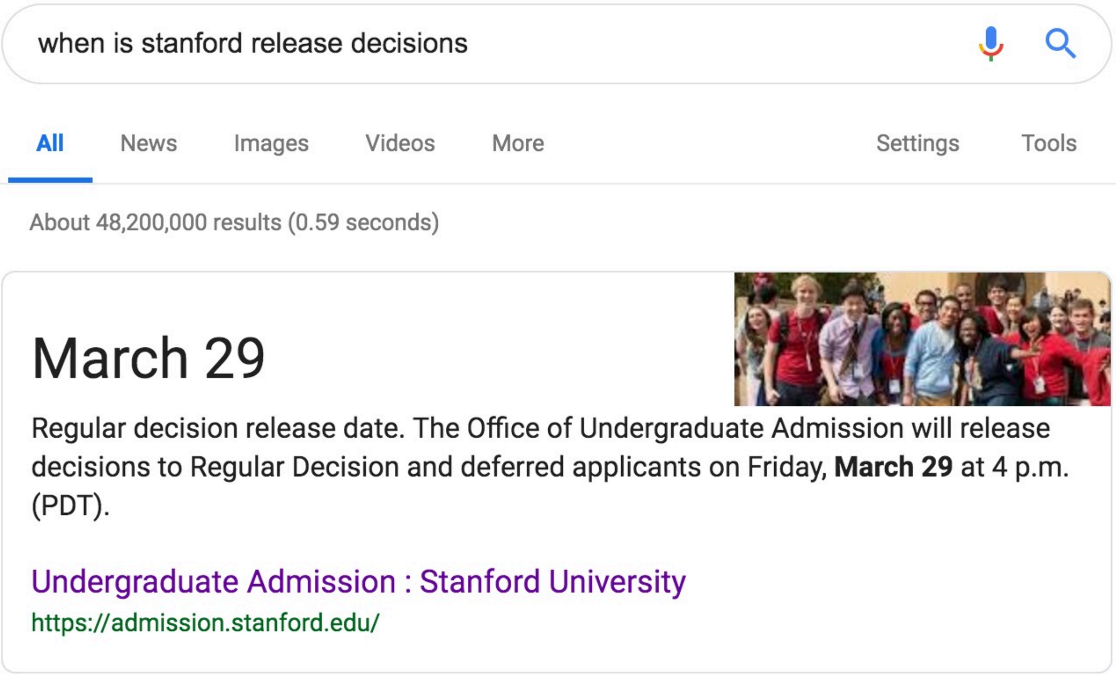
Impatient Reader

- 在表示背景文档的时候，关注不仅仅来源于整个问题的信息，还来源于问题中每个词的信息。
- 每处理问句中的一个单词，就进行关注。



利用双向LSTM去建模document和query，query利用正向LSTM的最后一个状态和逆向LSTM的第一个状态去表示。表示为U的过程中，每个状态都去动态分配文章权重，进而改变文章的整体表示（多个r），最终利用最后一个状态作为文章的表示。

IR-based + MC QA



- 阅读理解模型模型相对比较复杂，无法线上直接取代基于检索的问答系统。
- 目前基于检索的问答系统，基本都是段落级别的ranking，不能精确的抽取答案句或实体答案。
- 阅读理解模型和检索的问答系统互补，将检索的问答系统得到的最优段落作为阅读理解系统的输入，得到精准的主干结果。

Leveraging Knowledge Bases in LSTMs for Improving Machine Reading

- Yang, Bishan, and Tom Mitchell. "Leveraging Knowledge Bases in LSTMs for Improving Machine Reading." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017 (Citation: 40)

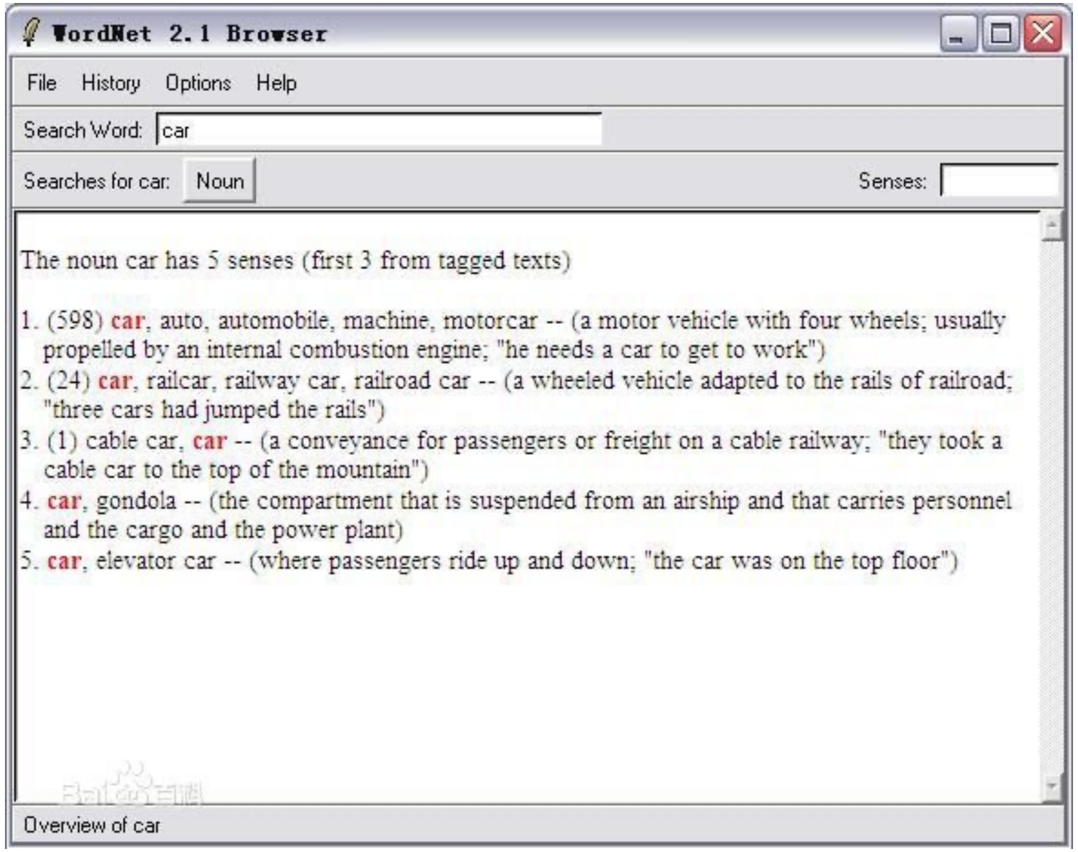
I am working on solving exciting problems at LAER AI, a startup Igor Labutov and I co-founded, focusing on building next-generation semantic search technologies for enterprise.

Tom Mitchell，美国卡内基梅隆大学（CMU）计算机科学学院机器学习系主任、教授，美国工程院、艺术与科学院院士，美国科学促进会（AAAS）、国际人工智能协会（AAAI）Fellow，他在机器学习、人工智能、认知神经科学等领域卓有建树，撰写了机器学习方面最早的教科书之一《机器学习》（1977年），是机器学习领域的著名学者

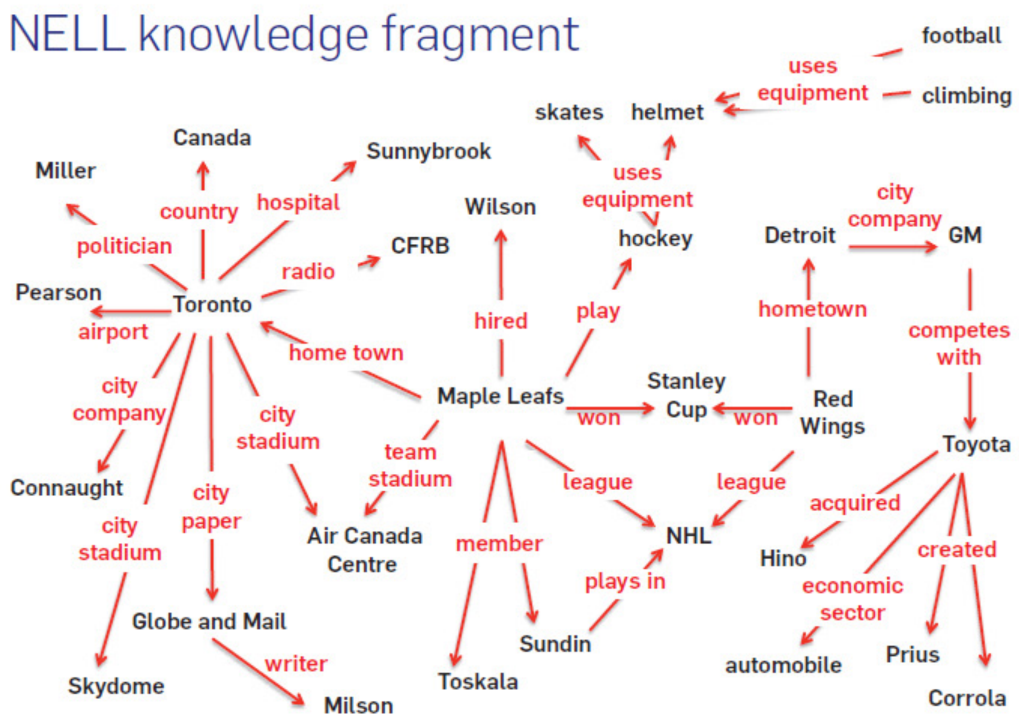
Leveraging Knowledge Bases in LSTMs for Improving Machine Reading

- 提出通过更好地利用外部知识库的方法解决机器阅读问题（本文关注事件和实体抽取）
- 传统方法中的问题：
  - 用离散特征表示知识库的知识存在了特征生成效果差而且特征工程偏特定任务：本文选择用连续向量表示方法来表示知识库
  - 传统神经网络端到端模型使得大部分背景知识被忽略：论文基于 BiLSTM 网络提出扩展网络 KBLSTM，结合 attention 机制在做任务时有效地融合知识库中的知识
- 问题：
  - 需不需要知识库？
  - “Maigretleft viewers in tears.”利用背景知识和上下文我们可以知道Maigret指一个电视节目
  - “Santiago is charged withmurder.”如果过分依赖知识库就会错误地把它看成一个城市
  - 因此：根据上下文判断知识库哪些知识是相关的也很重要

Leveraging Knowledge Bases in LSTMs for Improving Machine Reading



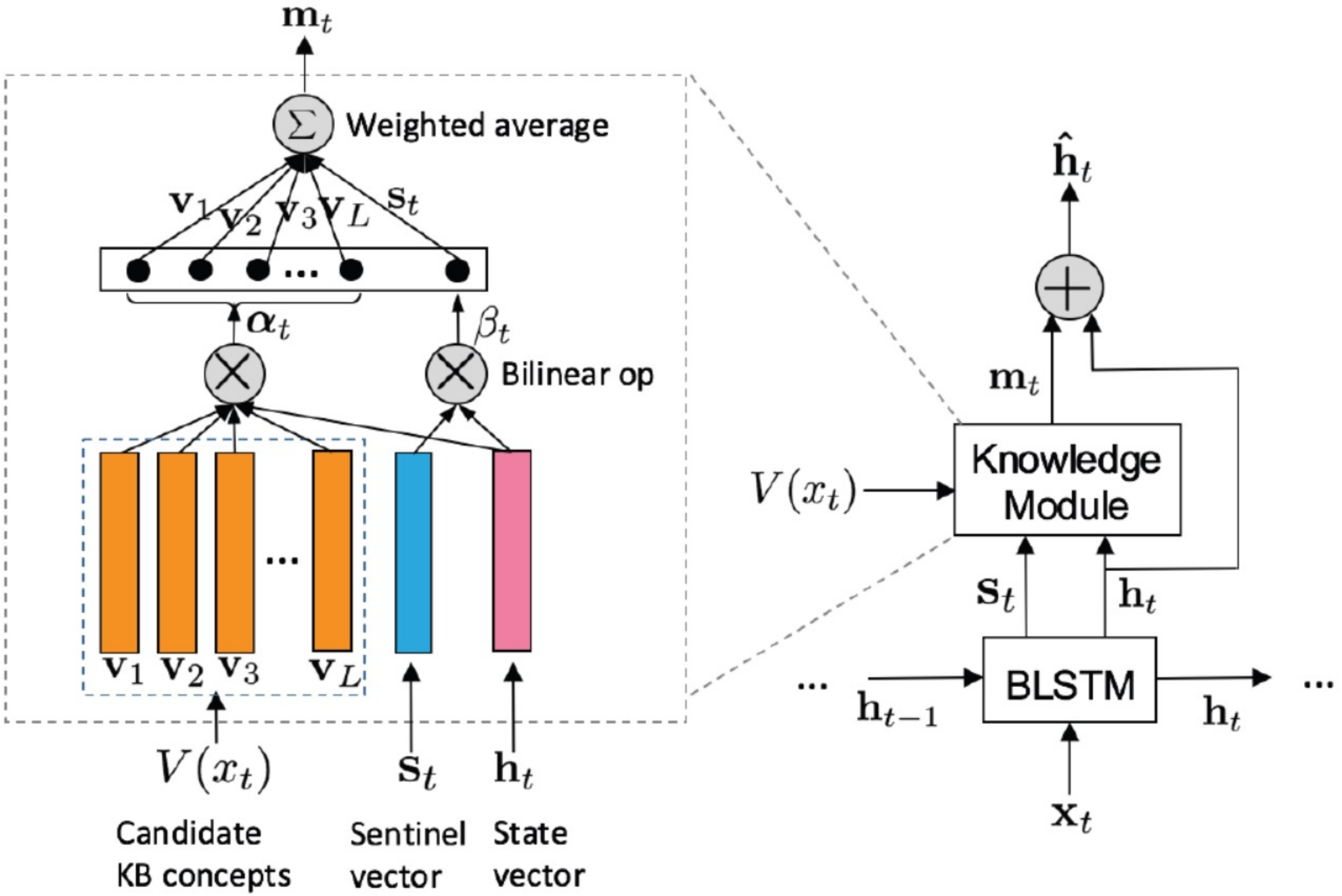
WordNet



NELL

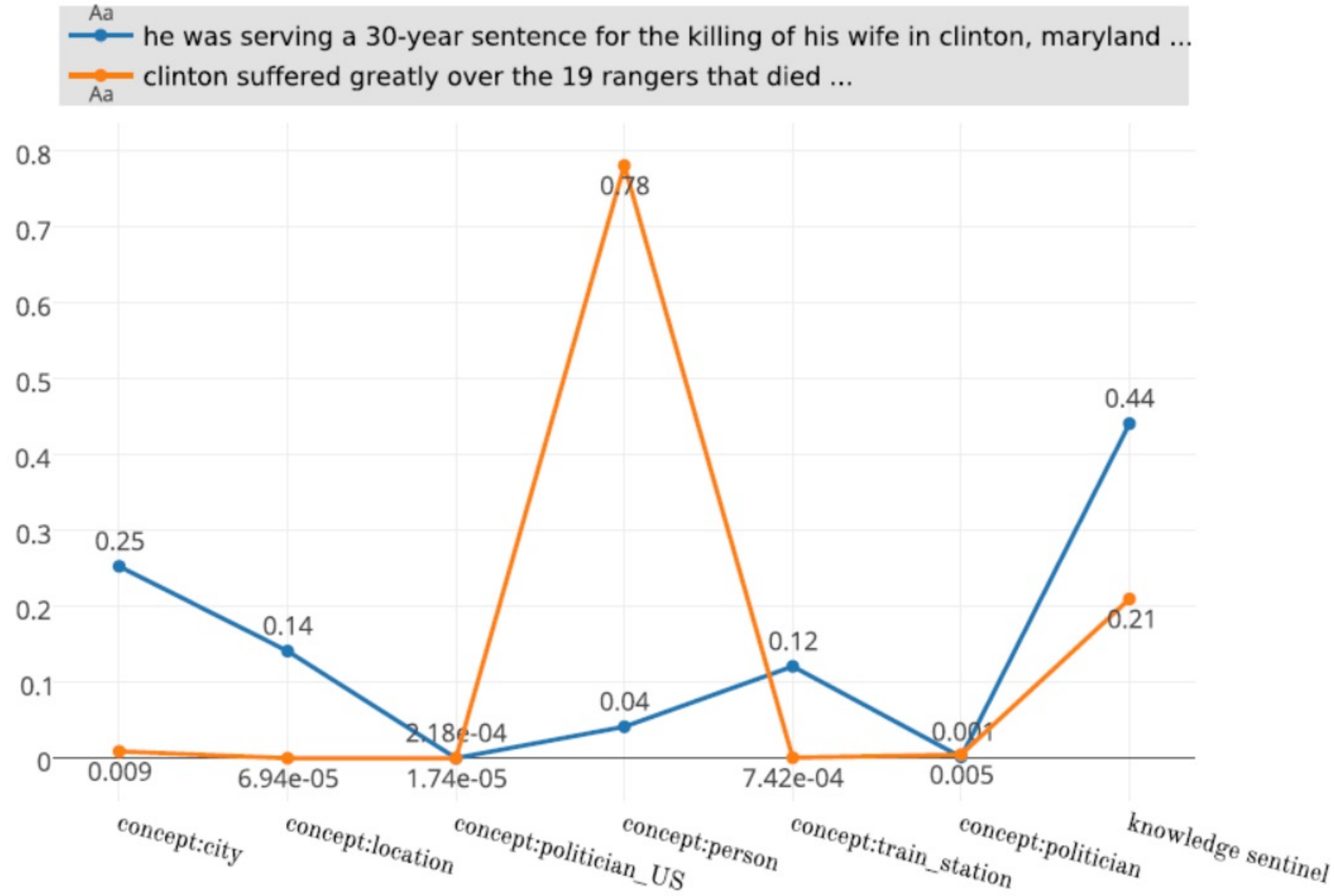
Leveraging Knowledge Bases in LSTMs for Improving Machine Reading

- KBLSTM（Knowledge-aware Bidirectional LSTMs）
  - As each time step t, the knowledge module retrieves a set of candidate KB concepts V (x_t) that are related to the current input x_t, and then computes a knowledge state vector m_t that integrates the embeddings of the candidate KB concepts v_1; v_2; …, v_L and the current context vector s_t



Leveraging Knowledge Bases in LSTMs for Improving Machine Reading

- 例一：He killing of his wife in clinton…
- 例二：Clinton suffered greatly over the…



Leveraging Knowledge Bases in LSTMs for Improving Machine Reading

Model	P	R	F1
BiLSTM	83.5	86.4	84.9
BiLSTM-CRF	87.3	84.7	86.0
BiLSTM-Fea	86.1	84.7	85.4
BiLSTM-Fea-CRF	87.7	86.1	86.9
KBLSTM	87.8	86.6	87.2
KBLSTM-CRF	**88.1**	**87.8**	**88.0***

Table 1: Entity extraction results on the ACE2005 test set with gold-standard mention boundaries.

Model	P	R	F1
Ratinov and Roth (2009)	82.0	84.9	83.4
Durrett and Klein (2014)	85.2	82.8	84.0
BiLSTM-CNN	82.5	82.4	82.5
BiLSTM-CNN+emb	85.9	86.3	86.1
BiLSTM-CNN+emb+lexicon	86.0	86.5	86.2
BiLSTM	84.9	86.3	85.6
BiLSTM-CRF	85.3	86.6	85.9
BiLSTM-Fea	85.2	86.4	85.8
BiLSTM-Fea-CRF	85.2	86.8	86.0
KBLSTM	**86.3**	86.2	86.2
KBLSTM-CRF	86.1	**86.8**	**86.4***

Table 4: Entity extraction results on the OntoNotes 5.0 test set.

Recommended References

- Parikh, Soham, et al. "Eliminet: A model for eliminating options for reading comprehension with multiple choice questions." *arXiv preprint arXiv:1904.02651* (2019).
- Xu, Hu, et al. "Review Conversational Reading Comprehension." *arXiv preprint arXiv:1902.00821* (2019).
- Xu, Hu, et al. "BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis." *arXiv preprint arXiv:1904.02232* (2019).
- Yang, Bishan, and Tom Mitchell. "Leveraging knowledge bases in lstms for improving machine reading." *arXiv preprint arXiv:1902.09091* (2019).
- Liu, Xiaodong, et al. "Stochastic answer networks for machine reading comprehension." *arXiv preprint arXiv:1712.03556*(2017).
- Pan, Boyuan, et al. "Memen: Multi-layer embedding with memory networks for machine comprehension." *arXiv preprint arXiv:1707.09098* (2017).
- Cheng, Jianpeng, Li Dong, and Mirella Lapata. "Long short-term memory-networks for machine reading." *arXiv preprint arXiv:1601.06733* (2016).

6. Dialogue Management

问答系统

- 分析方法
  - 检索
  - NLU
  - End-to-End
- 答案形态
  - FAQ
  - 结构化数据
  - 多文档抽取
  - 单文档抽取
  - 自然语言生成
- 客观性
  - 事实性
  - 非事实
  - 情感性
- 轮次
  - 单轮
  - **多轮**
- 领域
  - 开放
  - 垂直

QA is not a single atom task but a assembling application which is composed by a group of technical components and has various forms. Moreover, different forms have very different technical roadmaps.

6. Dialogue Management

- 自然语言理解（NLU）
  - 将用户输入转化为抽象的语义表达，包含用户意图和一系列槽值集合
- 对话状态追踪(DST)
  - 在 NLU基础上，状态追踪模块(DST)更新内部状态s，包括用户目标，对话进度和用户需求的槽值对等
- 对话策略（Policy Learning）
  - 根据对话策略(policy)决定下一时刻的动作（action）
- 自然语言生成(NLG)
  - 将策略选择的动作作为输入，转换生成自然语言

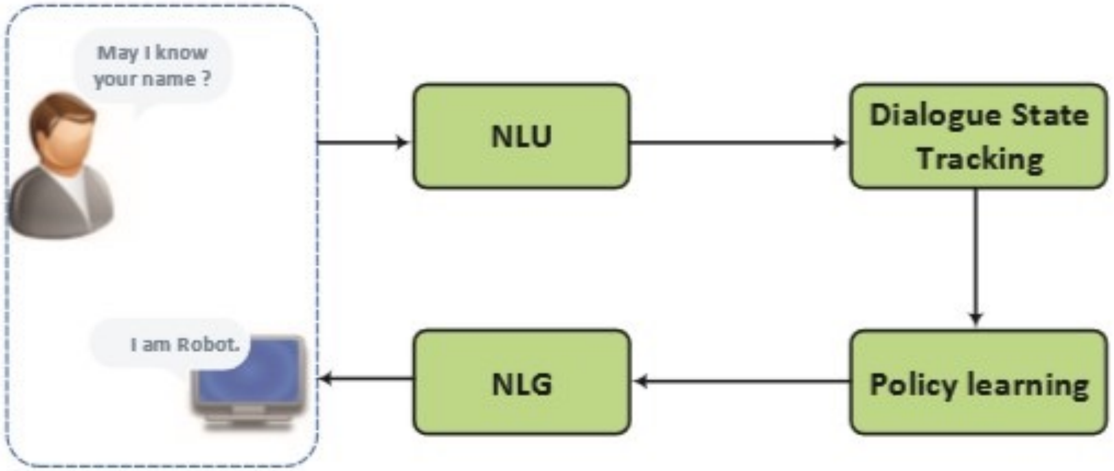
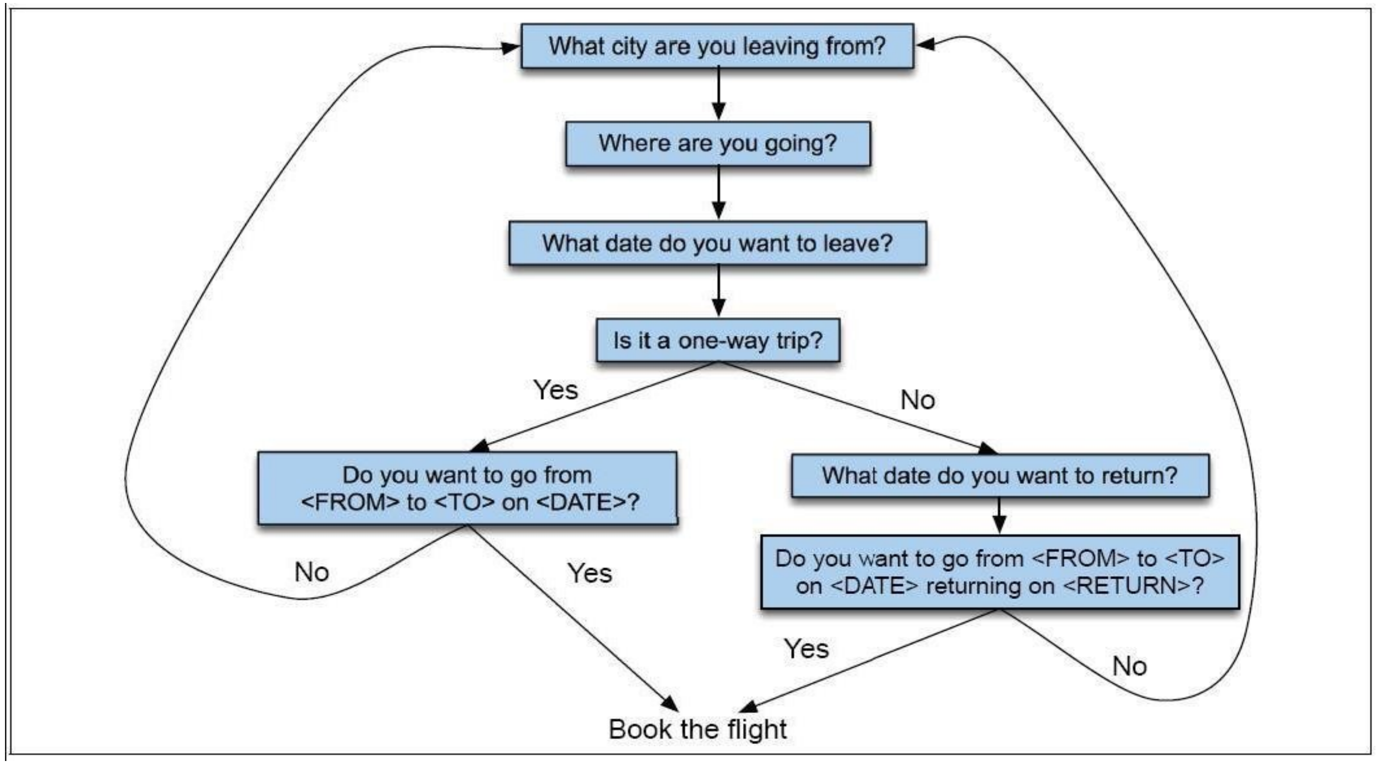


Figure 1: Traditional Pipeline for Task-oriented Systems.

Finite Status Machine

- 考虑一个订票系统，它有如下可能的状态:
- Ask the departure city
- Ask for a destination city
- Ask for a time
- Ask whether the trip is round-trip or not



Slot-filling

使用框架的结构指导对话过程

用户可以一次回答多个系统问题
- 问答过程就是一个槽-值填充的过程
- 当所有槽的值都填满了，则可以信息系统查询

Slot	Question
ORIGIN	What city are you leaving from?
DEST	Where are you going?
DEPT DATE	What day would you like to leave?
DEPT TIME	What time would you like to leave?
AIRLINE	What is your preferred airline?

- Rule-based Learning
  Warm Start
- Supervised Learning
  针对规则产生的行为进行有监督学习
- Deep Reinforcement Learning
  - Q-Learning
  - 表现要好于基于规则和有监督的方法

A Survey on Dialogue Systems: Recent Advances and New Frontiers

- NLU:
  - Intent detection: classification problem
  - Slot filling: sequence labeling problem

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- Zhao, Tiancheng, and Maxine Eskenazi. *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 2016. (Citation: 91)

Welcome! I am Tiancheng Zhao (赵天成). I received my Ph.D. degree in May 2019, from Language Technologies Institute (LTI) , **Carnegie Mellon University**, advised by Maxine Eskenazi. Also, I am a member of Dialog Research Center (Dialrc) . Before becoming a PhD student at Carnegie Mellon University, I graduated from the Master of Language Technologies (MLT) program at LTI in 2016, advised by Maxine Eskenazi and and Alan W Black. Prior to that, I obtained my bachelor degree in Electrical Engineering from University of California, Los Angeles and worked on speech signal processing, advised by Abeer Alwan.

Maxine Eskenazi

Principal Systems Scientist, Carnegie Mellon University

Research Goals

To create intelligent agents (using spoken dialogue architectures, automatic speech recognition and synthesis) using knowledge of the speech signal and of human cognition. To confront research with real human users and, in turn, provide a real benefit to those users. This endeavor implies studying groups of users, input conditions and speaking styles, the manner in which humans and systems can entrain to one another, and how we can assess the systems we build, often profiting from the wisdom of the crowd.

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- 不同于chatbot，任务驱动QA有一个明确的使用目的
  - 天气查询，机票预定，航班查询
- 用户在使用任务驱动对话系统的时候，通过多轮人机对话，向电脑阐述自己的需求。而电脑在理解了用户的所求之后，通过对于后端数据库的查询和修改，来实现用户要求的功能
- 假如用户的阐述不够清楚，或者用户的需求比较复杂，系统可以主动询问，澄清的方式来帮助用户找到满意的结果

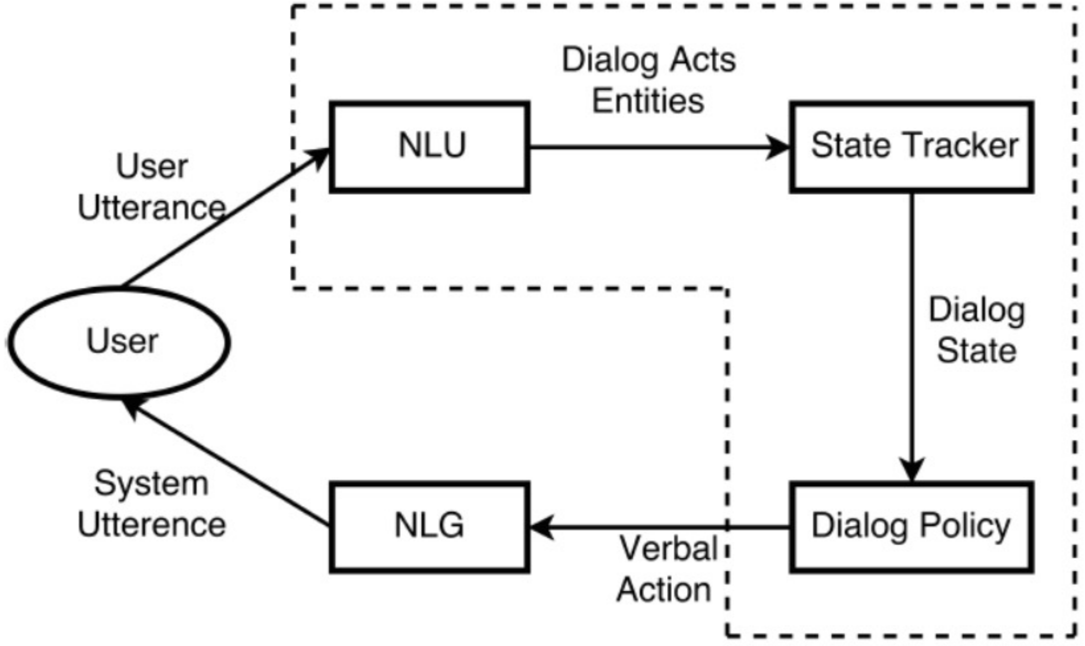


Figure 1: Conventional pipeline of an SDS. The proposed model replaces the modules in the dotted-line box with one end-to-end model.

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- 任务驱动对话系统需要同时解决数个AI核心问题，传统做法是对这些问题逐一分解解决
- 挑战：
  - 模块领域迁移困难
  - 上游模块的错误会传递到下游
- 端到端学习常见的做法是设计出可以整体可微的模型，然后利用 反向传播把输出端的 梯度传递到整个神经网络，已达到联合优化

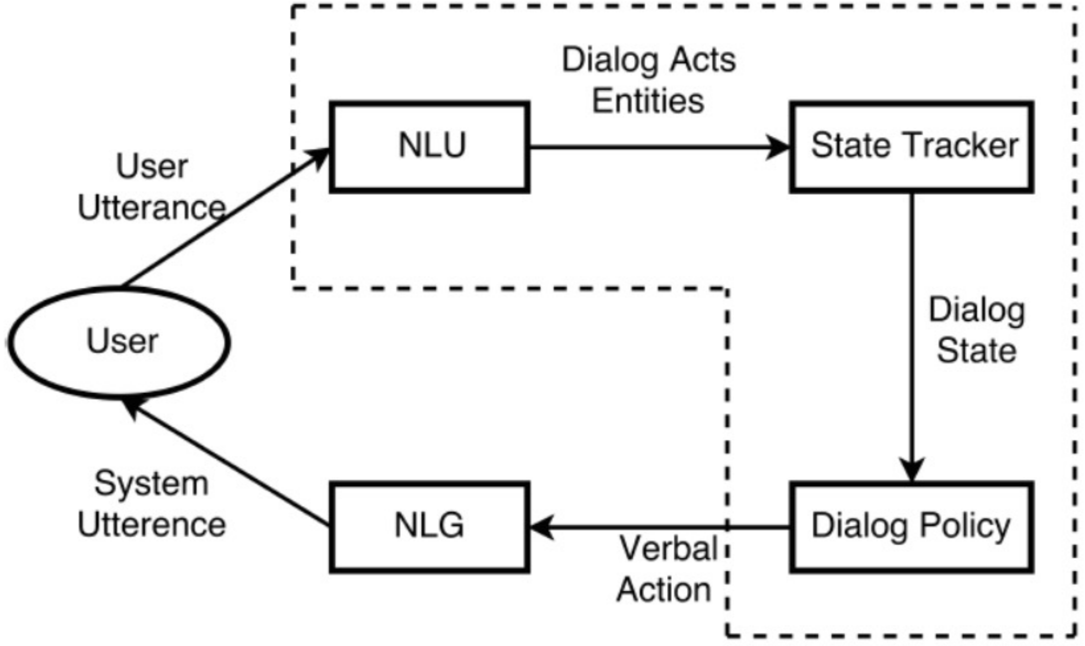


Figure 1: Conventional pipeline of an SDS. The proposed model replaces the modules in the dotted-line box with one end-to-end model.

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- 介绍：将对话看作是部分可观测的马尔可夫决策过程（POMDP），设计一个端对端的系统代替了NLU, DST和Policy learning三个模块，面向猜人名（20Q）的特定任务。将状态追踪（插槽动作）和对话策略（语言动作）均看作action联合训练。语言动作包括用户进行询问和产生guess，插槽动作修改用于数据库查询的假设向量h，其中h是一维度大小等于问题个数的向量。

- 输入：当前时刻的环境观测状态o。当下agent选择的action（one-hot），action后对用户的观测（用户的自然语言表示bag-of-bigram），action后对数据库的观测（满足查询条件的人数)向量并联构成。

- 中间件：利用LSTM作为状态追踪器。产生的状态b表示被输入进策略网络（MLP），策略网络被用来评估Q函数。每次评估一种Q函数。当用户有新输入时，产生在言语动作集上的Q函数，否则产生在插槽动作集上的Q函数。

$$A_{mask}(s) = A_h \quad \text{new inputs from the user} \quad (6)$$
$$= A_v \quad \text{otherwise} \quad (7)$$

- 输出：训练得到语言动作和插槽动作的Q函数，测试时根据Q函数选择action。

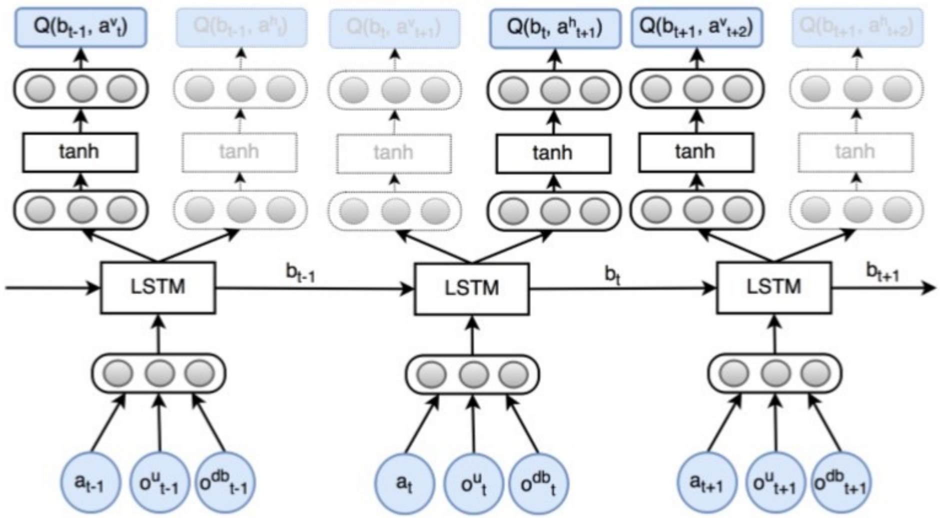


Figure 3: The network takes the observation o_t at turn t. The recurrent unit updates its hidden state based on both the history and the current turn embedding. Then the model outputs the Q-values for all actions. The policy network in grey is masked by the action mask

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- 优化使用三个技巧：

✓ 减少参数数量：由于假设向量中每个槽都相似（槽值都是yes/no/unknown），因此只需要考虑last question，使用策略网络后修改对应槽即可。

✓ 动作掩码：根据用户是否有新输入来确定训练语言动作网络还是修改假设向量网络。

✓ Reward sharping：来自用户的reward只有在对话结束出现（猜对or猜错），故利用数据库的查询结果作为reward加快训练。

$$\phi(s_t) = P_{max}(1 - \frac{d_t}{D}) \quad \text{if } d_t > 0 \qquad (10)$$

$$\phi(s_t) = 0 \quad \text{if } d_t = 0 \qquad (11)$$

$$\bar{R}(s,a,s') = R(s,a,s') + F(s,a,s') \qquad (8)$$

$$F(s,a,s') = \gamma\phi(s') - \phi(s) \qquad (9)$$

实验：

✓ 用户模拟器根据均匀采样的方式选取它想的人名，且有5%的可能性将某属性看作unknown。

✓ 用户回答问题语言是yes/no/unknown三种形式之一的自然语言。（意味着agent要猜测用户回答是属于哪一类型，因此插槽并不一定总是正确的）

Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning

- 数据集：100个人名，分别有6种属性，每个属性有若干种值。Agent有32种言语动作，包括31个问题和1个猜测答案；3种槽值用来修改数据库的查询向量h。将对话过程看作是填槽过程，一共有31个槽（属性个数），每个槽有yes/no/unknow三种值。
- 验证了对话策略的有效性和状态追踪表示的有效性。
- 奖励：猜对人名+30，任务失败（根据假设h未找到任何人；对话轮数达到100轮；已经猜测了10次）-30，每猜错一次-5。
- 缺点：1）任务背景过于简化，agent只会提出答案是yes/no的问题，且用户没有提问的场景；4）奖励函数提前指定，可能并不准确，且单纯RL收敛速度慢。

Attribute	Q_a	Example Question
Birthday	3	Was he/she born before 1950?
Birthplace	9	Was he/she born in USA?
Degree	4	Does he/she have a PhD?
Gender	2	Is this person male?
Profession	8	Is he/she an artist?
Nationality	5	Is he/she a citizen of an Asian country?

Table 1: Summary of the available questions. Q_a is the number of questions for attribute a

Sys: Is this person male?
User: Yes I think so.
Sys: Is this person an artist?
User: He is not an artist.
...
Sys: I guess this person is Bill Gates.
User: Correct.

	Win Rate (%)	Avg Turn
Baseline	68.5	12.2
RL	85.6	21.6
Hybrid-RL	90.5	19.22

Table 3: Performance of the three systems

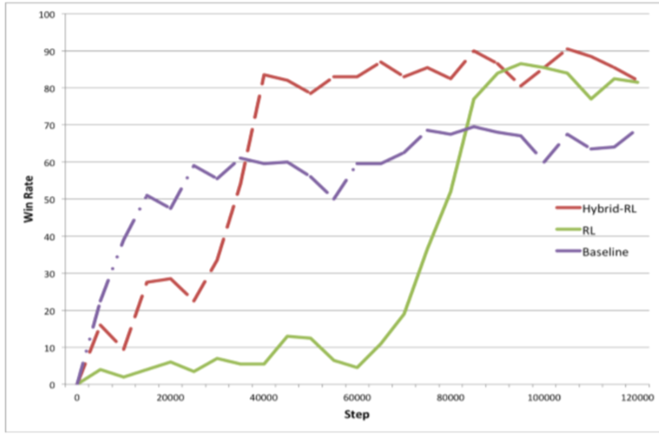


Figure 4: Graphs showing the evolution of the win rate during training.

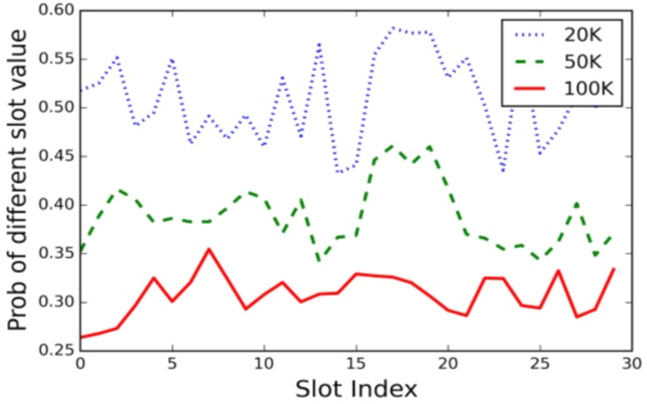


Figure 5: Performance of retrieving similar true dialog states using learned dialog state embedding.

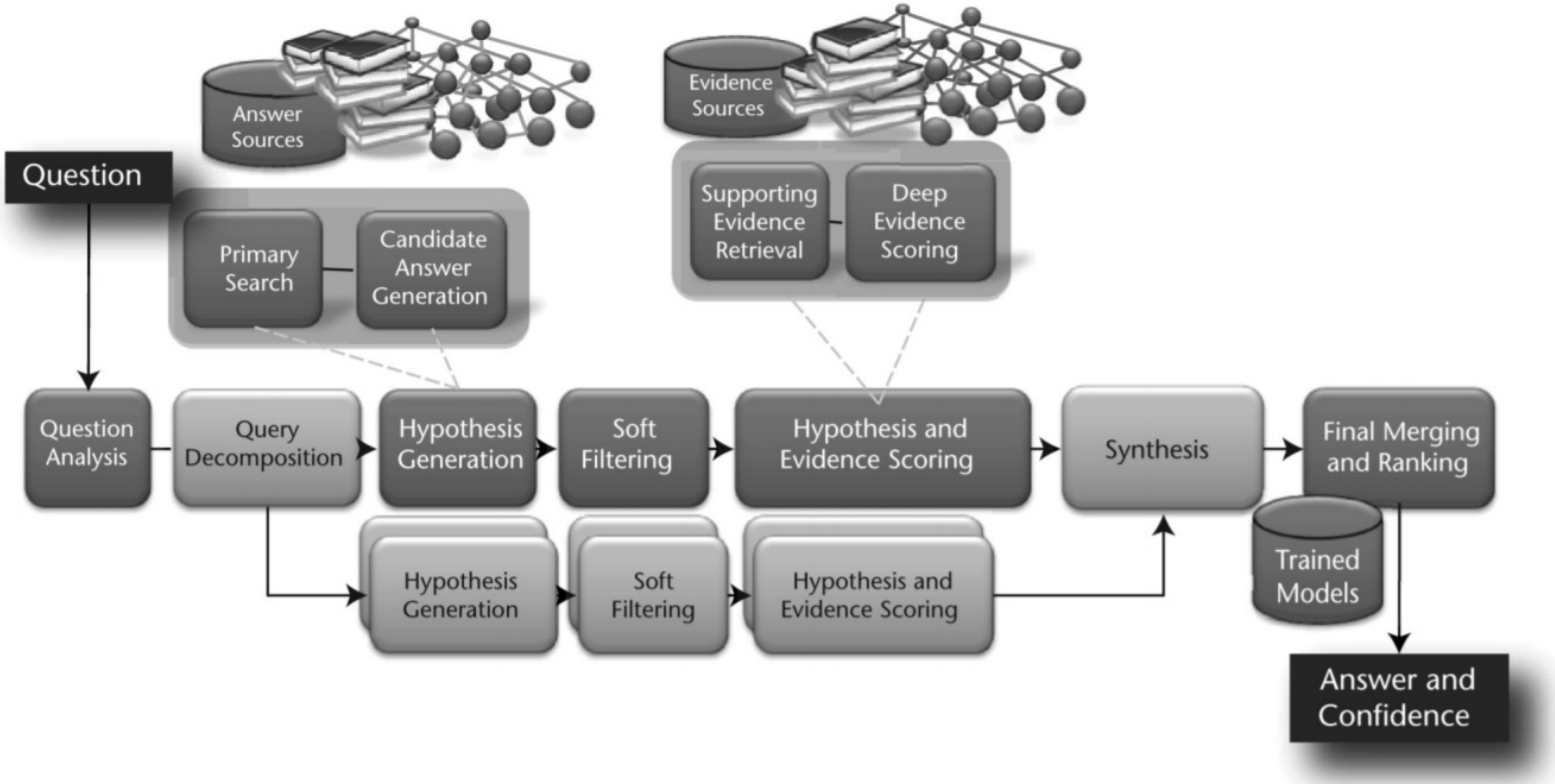
Recommended References

- Chen, Hongshen, et al. "A survey on dialogue systems: Recent advances and new frontiers." *ACM SIGKDD Explorations Newsletter* 19.2 (2017): 25-35.
- Zhao, Tiancheng, and Maxine Eskenazi. "Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning." *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 2016.
- Sarikaya, Ruhi, et al. "An overview of end-to-end language understanding and dialog management for personal digital assistants." *2016 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2016.
- Mrkšić, Nikola, et al. "Multi-domain Dialog State Tracking using Recurrent Neural Networks." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Vol. 2. 2015.
- Bordes, Antoine, Y-Lan Boureau, and Jason Weston. "Learning end-to-end goal-oriented dialog." *arXiv preprint arXiv:1605.07683* (2016).

94

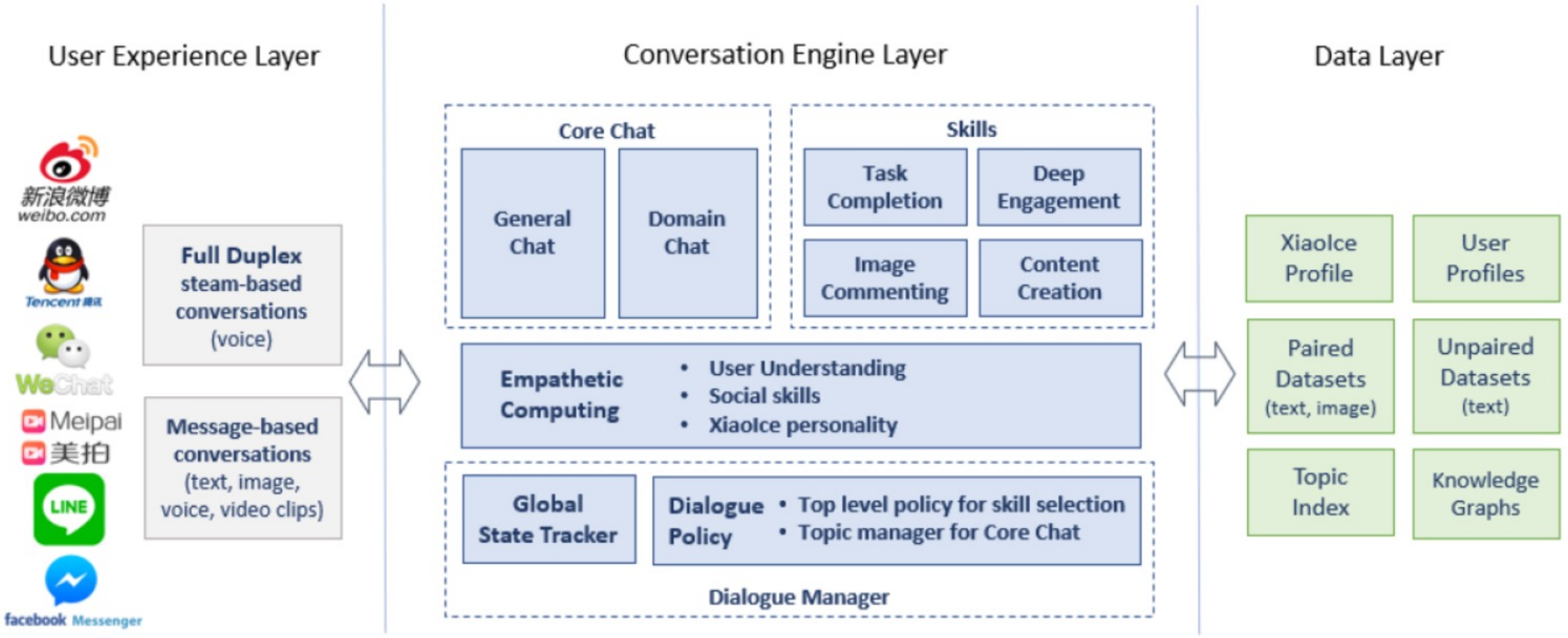
Products & Startups

IBM Watson (2013)



http://www.patwardhans.net/papers/GliozzoBPM13.pdf
https://researcher.watson.ibm.com/researcher/view_group_subpage.php?id=2159

Microsoft XiaoIce (2018)



https://arxiv.org/pdf/1812.08989.pdf

AliMe (2018)

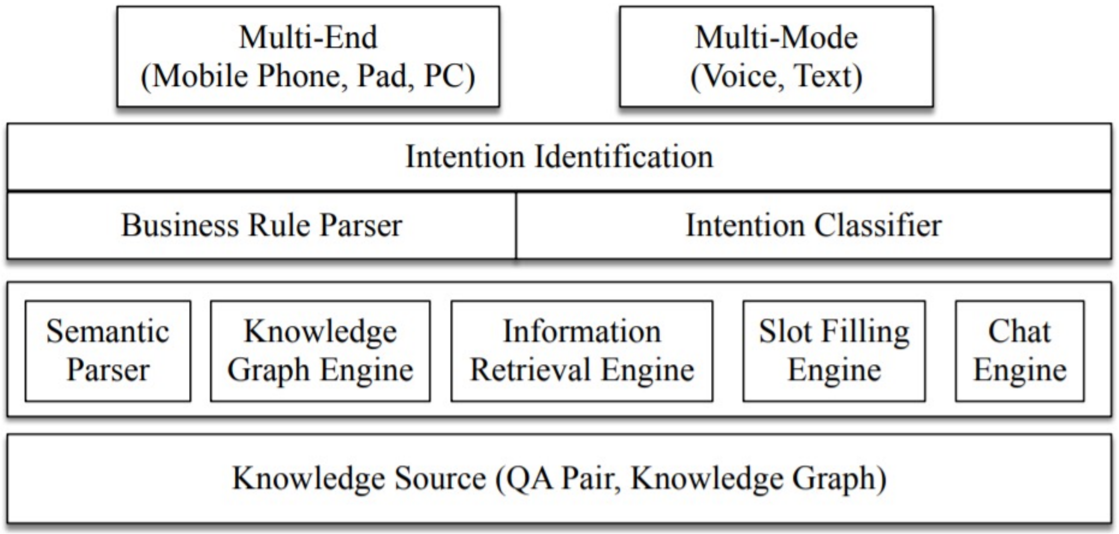


Figure 1: The overall architecture of *AliMe Assist*.

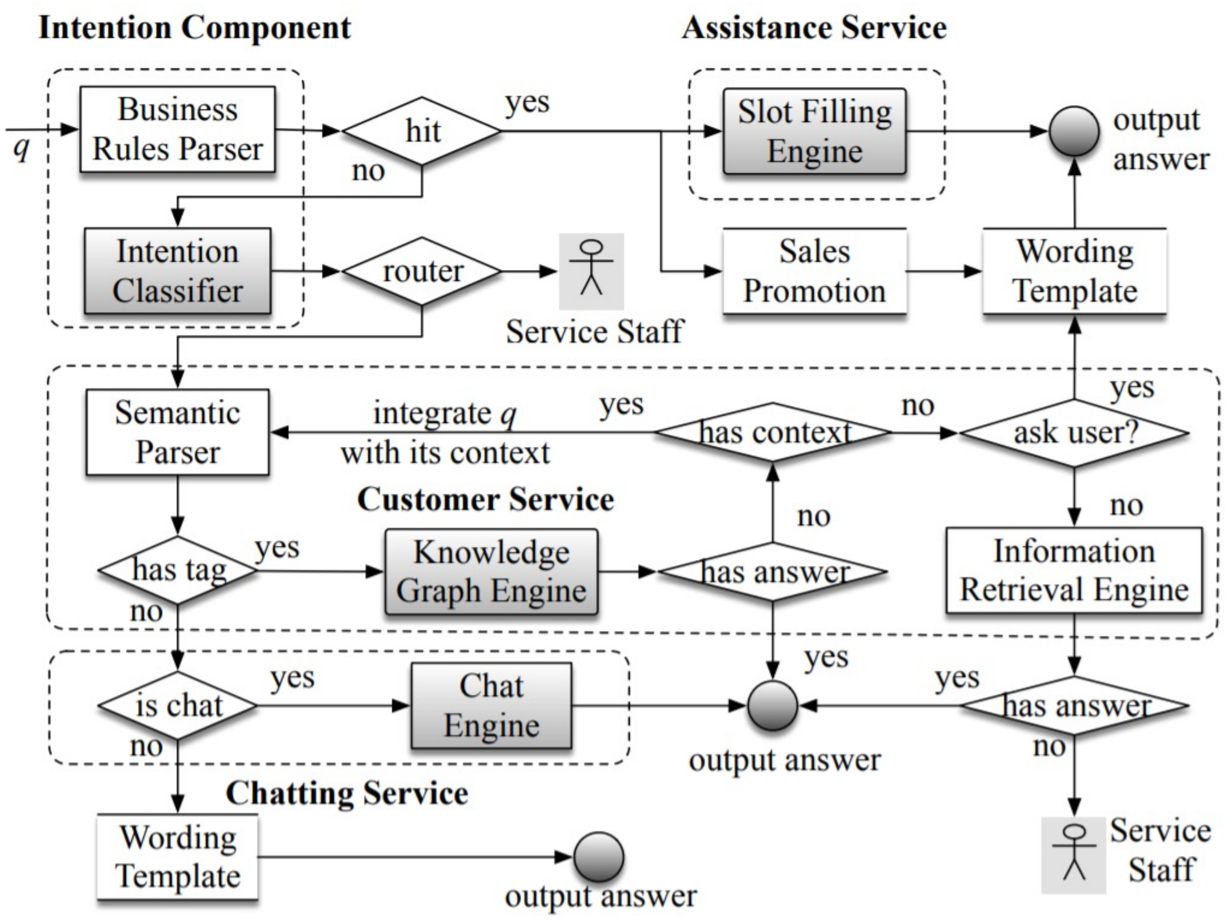


Figure 2: The overall processing flow of *AliMe Assist*

https://arxiv.org/pdf/1801.05032.pdf

DuerOS (2018)

小度对话核心系统：真正能听清听懂的对话系统
对话服务（DuerOS Conversational Service）
语音识别
ASR
语音播报
TTS
屏幕展示
View
自然语言理解
NLU
对话状态控制
DST
对话管理
DM
自然语言生成
NLG
搜索
Search
知识图谱
Knowledge graph
需求图谱
Query/click graph
网页图谱
Web graph
地理信息图谱
Map graph
用户画像
User profile
技能框架（DuerOS Bot Framework）
原生技能
音乐
电影
电视剧
短视频
资讯
问答
天气
有声书
地图
......
第三方技能
购物
打车
新闻
休闲
智能家居
聊天
电台
游戏
FAQ
......
技能开发工具

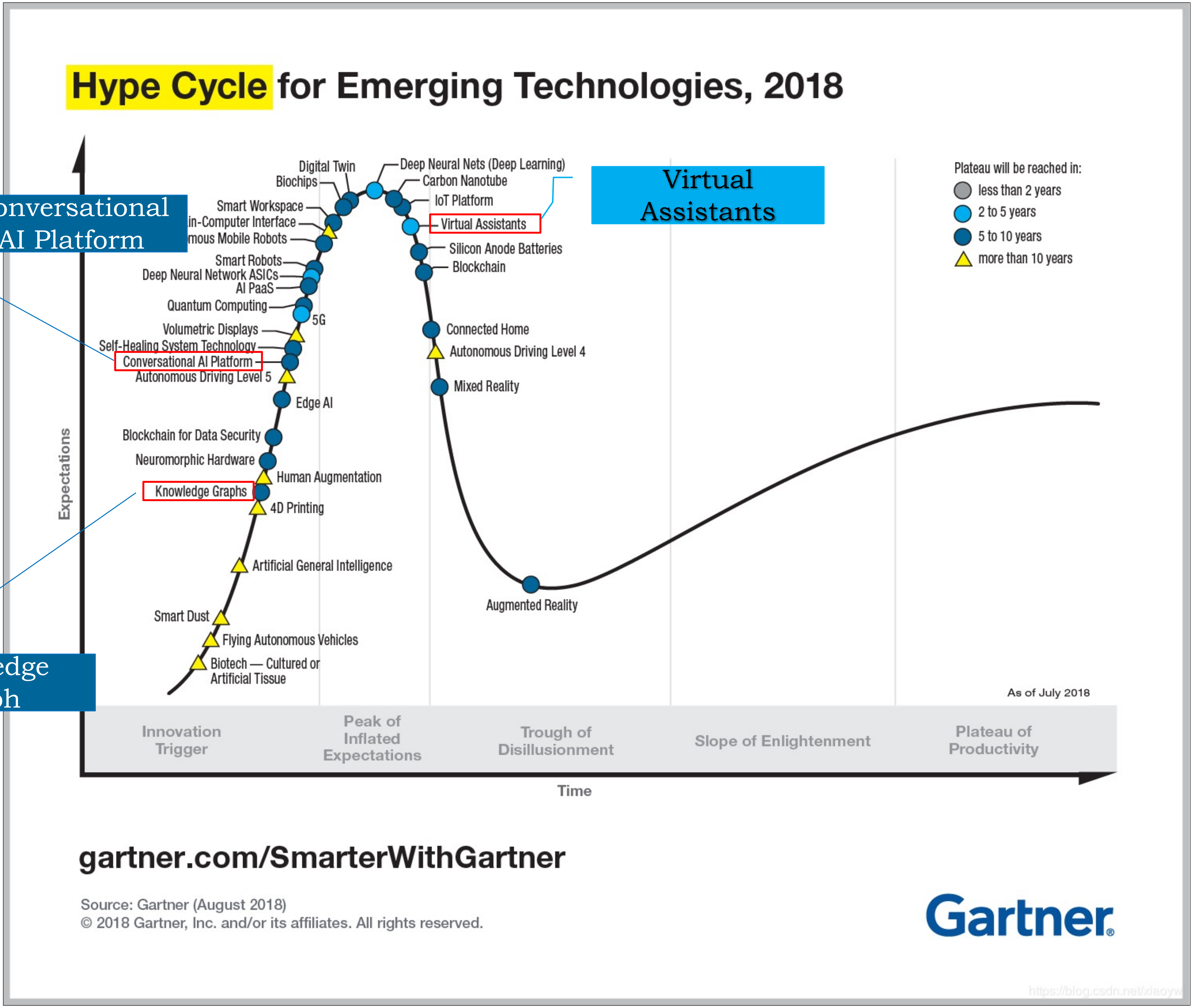
Looking forward

QA system is a good career plan

Conversational AI Platform

Knowledge Graph

Virtual Assistants



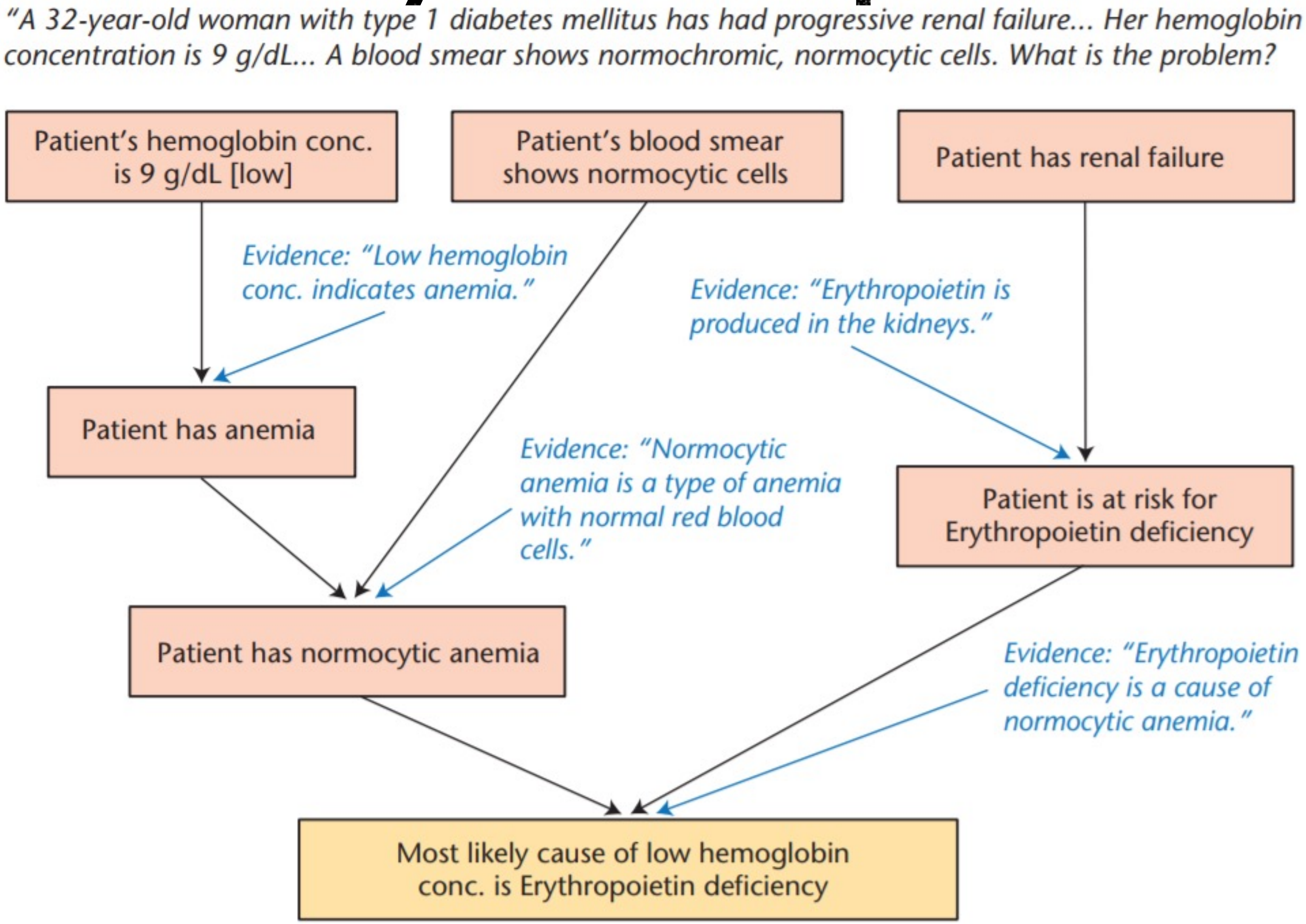
Future in 5-10 years

周明博士，微软亚洲研究院副院长
国际计算语言学协会（ACL20）主席

- **问答和阅读理解，使得搜索引擎更加精准**
- 语音识别和神经机器翻译，使得口语机器翻译会完全实用
- 用户画像，推动信息服务和广告更加自然、友好和个性化
- **聊天、问答和对话技术，推动自然语言对话达到实用**
- 对话技术和知识图谱，使得智能客服与人工客服更加完美结合
- 自然语言生成技术，使得自动写诗、作曲、自动生成新闻甚至小说会流行起来
- 人机对话的进步推动语音助手、物联网、智能硬件、智能家居的普及
- NLP+，就是 NLP 在金融、法律、教育、医疗等垂直领域得到广泛应用

Question in Reality is Complex

- QA is a precision first task in task-oriented applications



Tech. Trends

- Knowledge + NN (as human beings)
  - Huang, Xiao, et al. "Knowledge graph embedding based question answering." *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 2019.
- Transferring leaning QA
  - Kratzwald, Bernhard, and Stefan Feuerriegel. "Adaptive document retrieval for deep question answering." *arXiv preprint arXiv:1808.06528* (2018).
- Commonsense based QA
  - Commonsense knowledge aware conversation generation with graph attention （IJCAI 2018）
- Unsupervised QA
  - Lewis, Patrick, Ludovic Denoyer, and Sebastian Riedel. "Unsupervised Question Answering by Cloze Translation." *arXiv preprint arXiv:1906.04980* (2019).

Q&A