CCL 2016 Tutorial T2B

基于深度学习的知识库问答

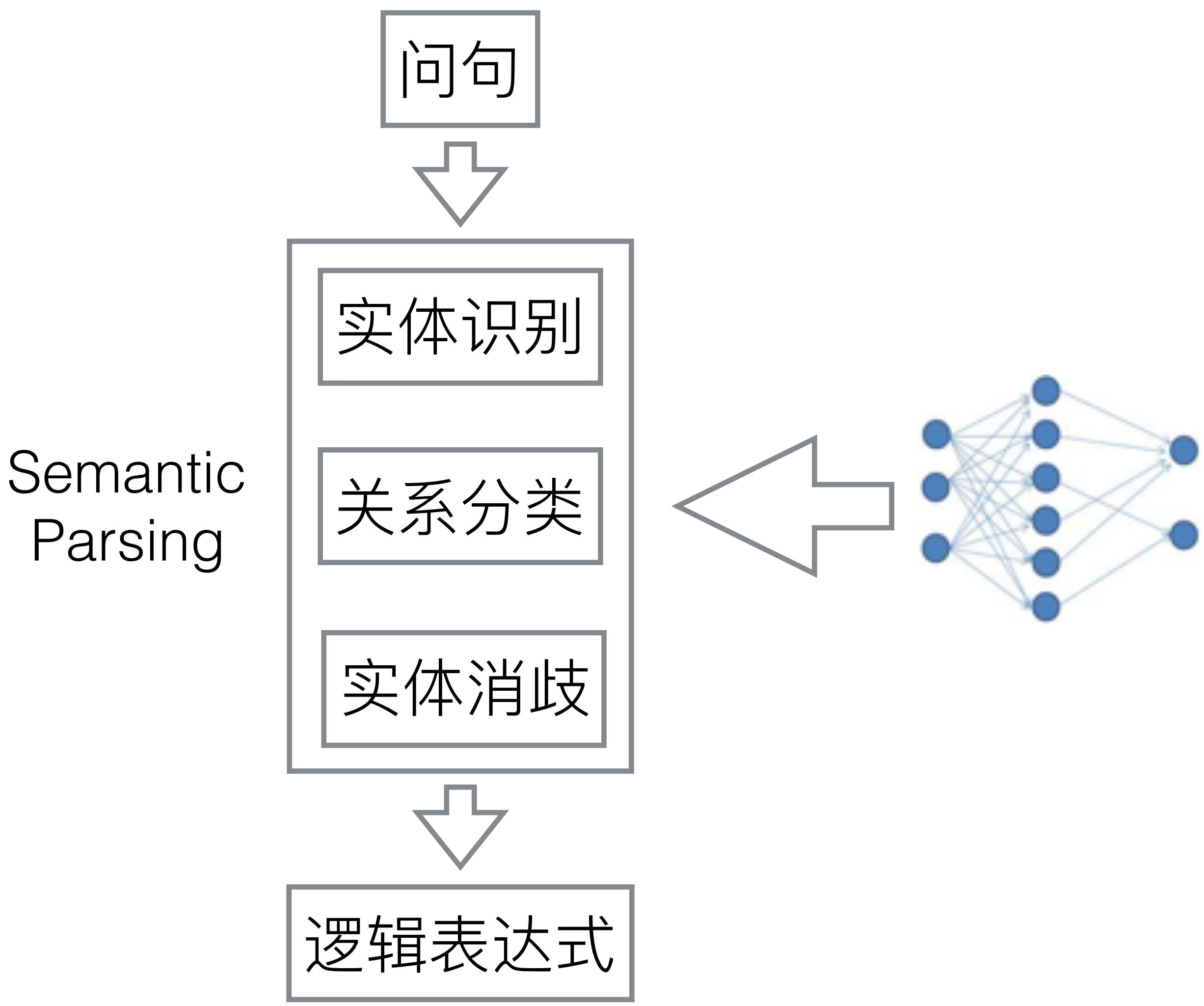
刘 康

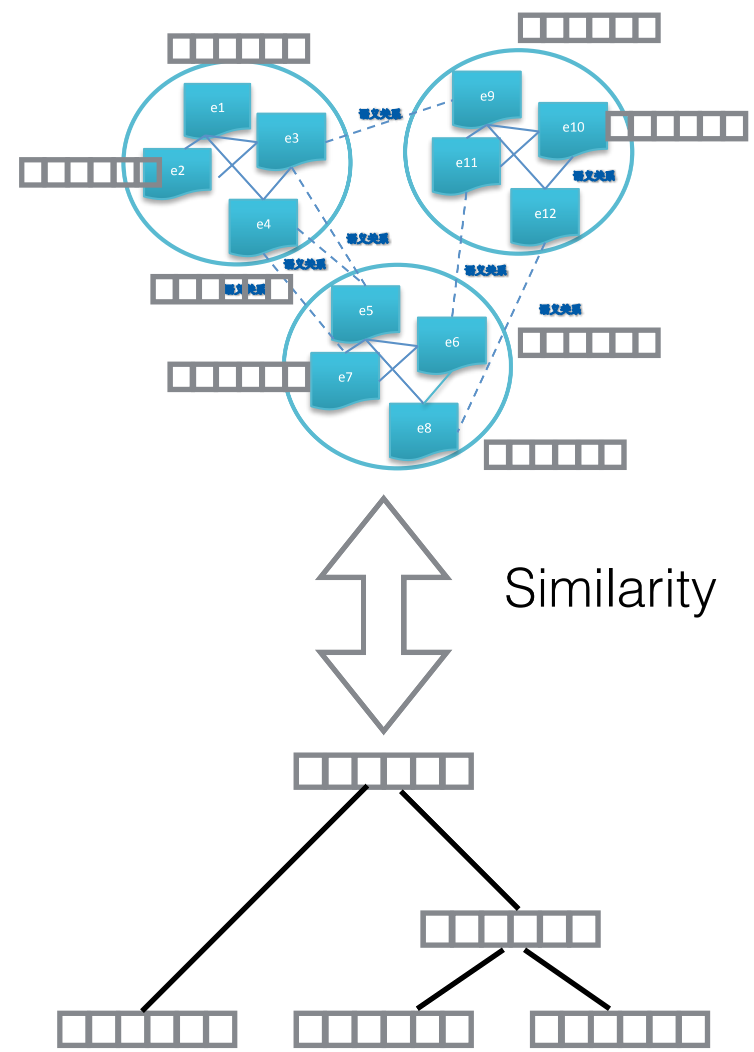
中国科学院自动化研究所
模式识别国家重点实验室

2016年10月14日

分类

- 利用深度学习对于传统问答方法的改进
- 基于深度学习的End2End模型

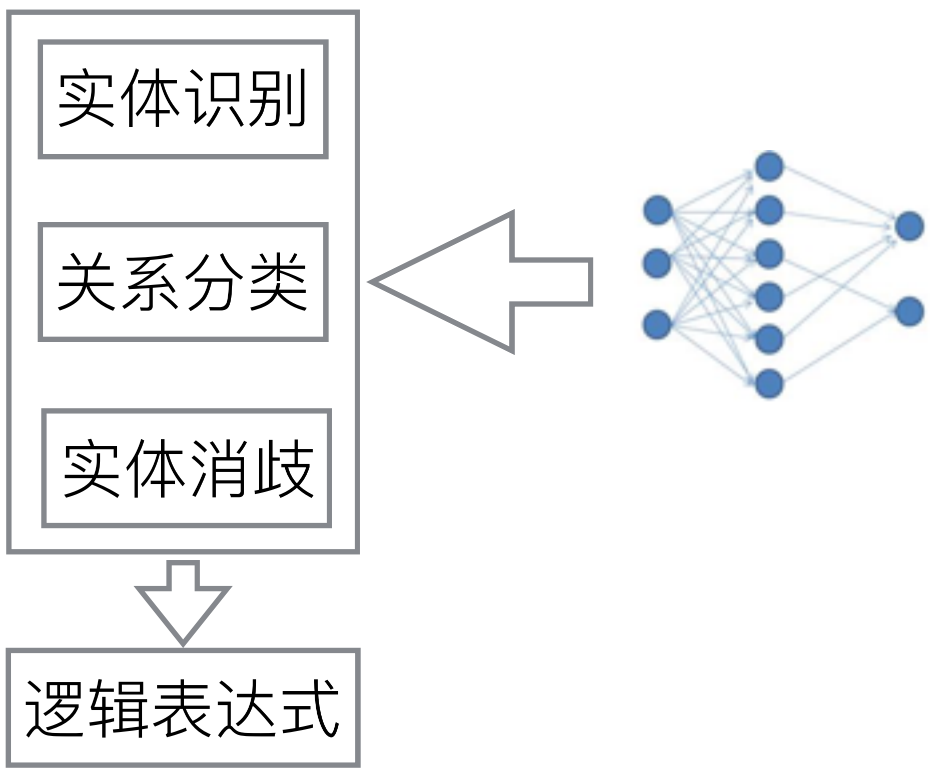




姚明的老婆的国籍是?

深度学习对于传统知识库问答方法的改进

利用深度学习对于传统问答方法的改进



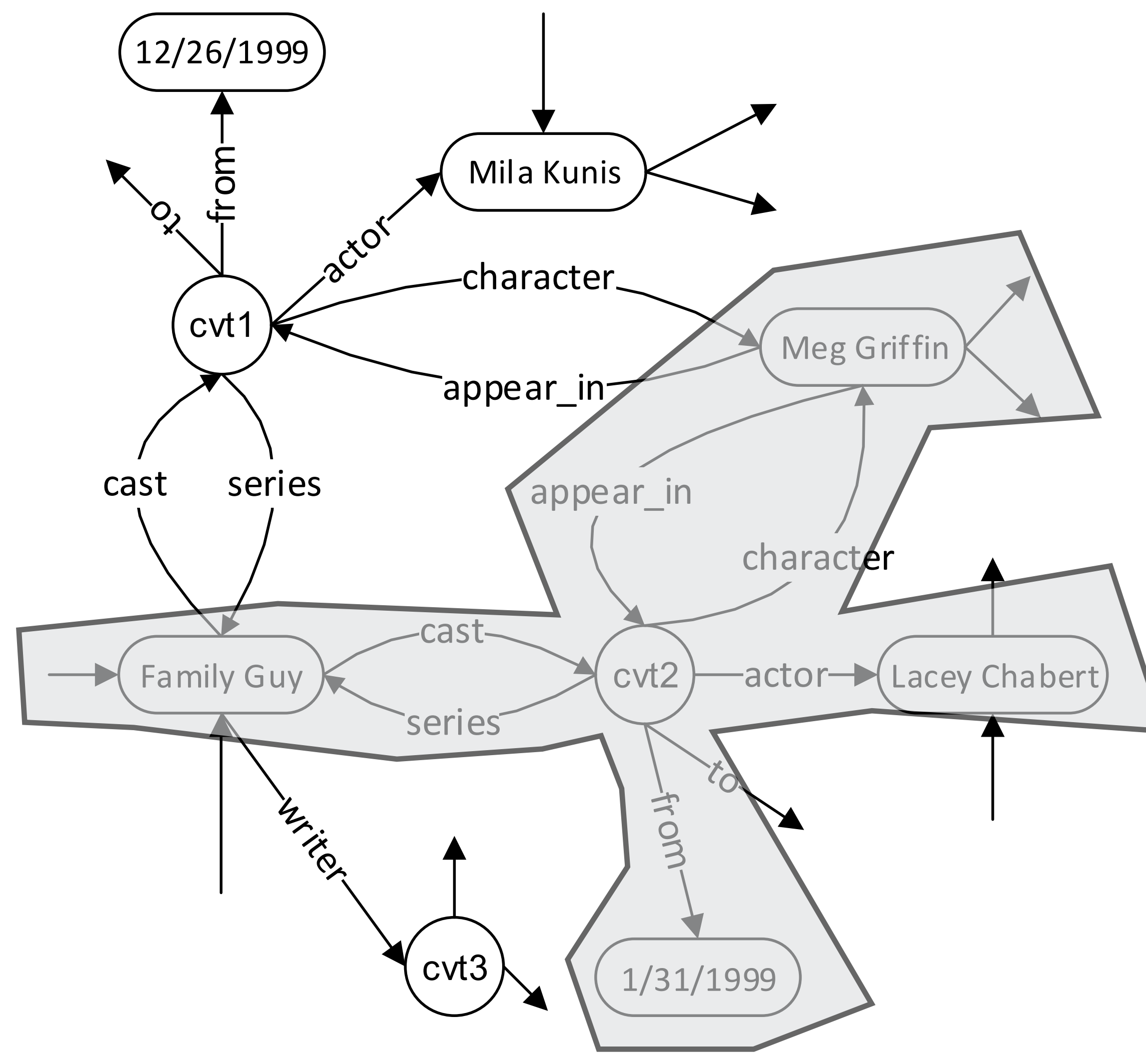
- DL-based关系模板识别
  - Yinh et al. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Bas, In Proceedings of ACL 2015 (**Outstanding Paper**)
  - Zeng et al. Relation Classification via Convolutional Deep Neural Network, in Proceedings of COLING 2014 (**Best Paper**)
  - Zeng et al, Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks, in Proceedings of EMNLP 2015
  - Xu et al. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths, In Proceedings of EMNLP 2015

Staged Query Graph Generation (Yinh et al. ACL 2015)

- 问答过程：基于结构化的问句语义表达（Lambda演算）在知识图谱匹配最优子图

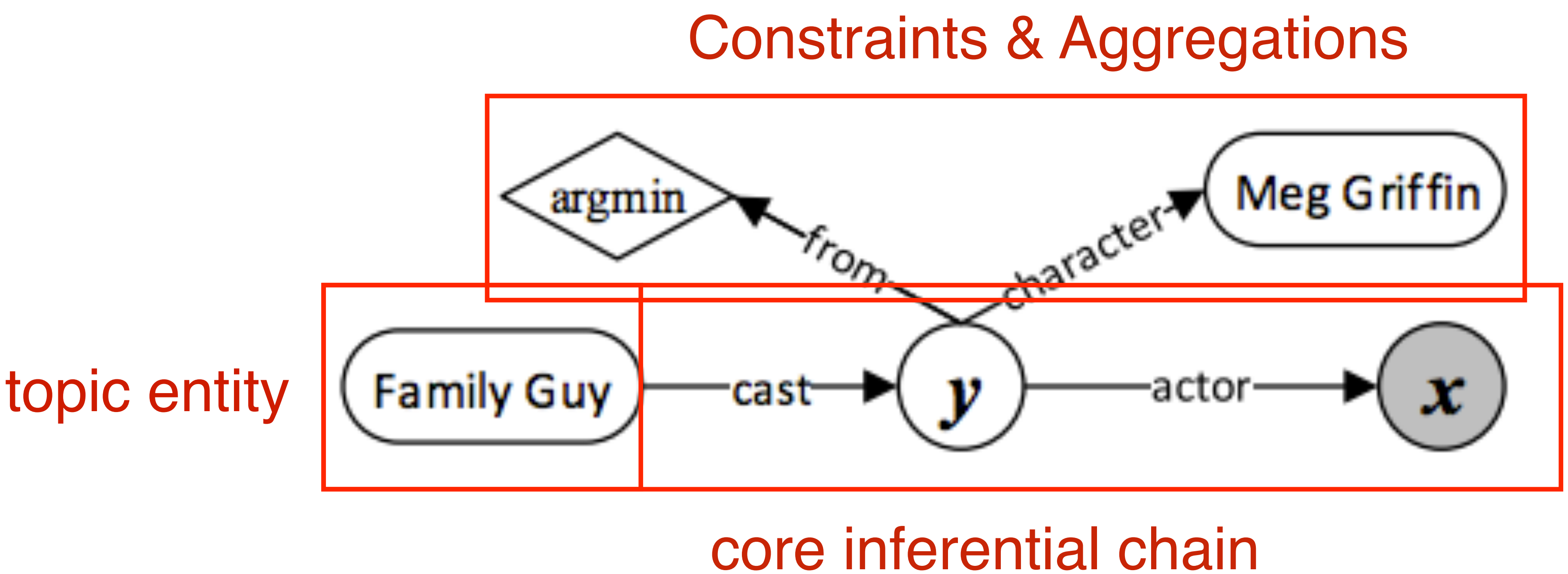
Who first voiced Meg on Family Guy?

$$\arg\min(\lambda x.Actor(x,Family\_Guy)$$
$$\wedge Voice(x,Meg\_Griffin),\lambda x.casttime(x))$$



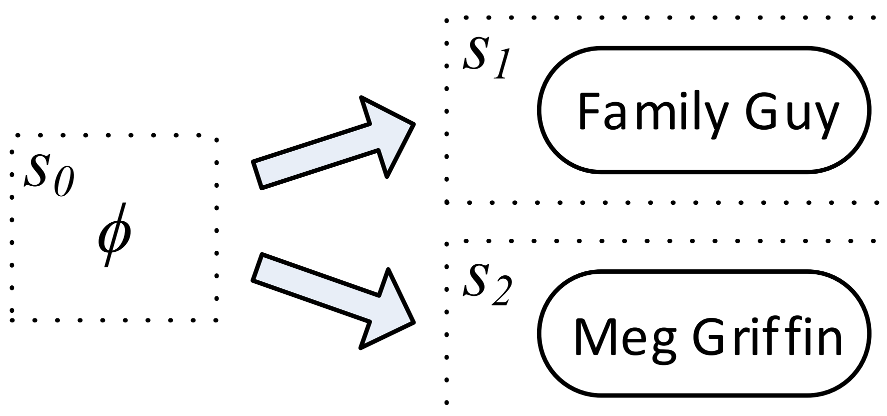
Query Graph

Who first voiced Meg on Family Guy?



Linking Topic Entity

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



Entity Linking

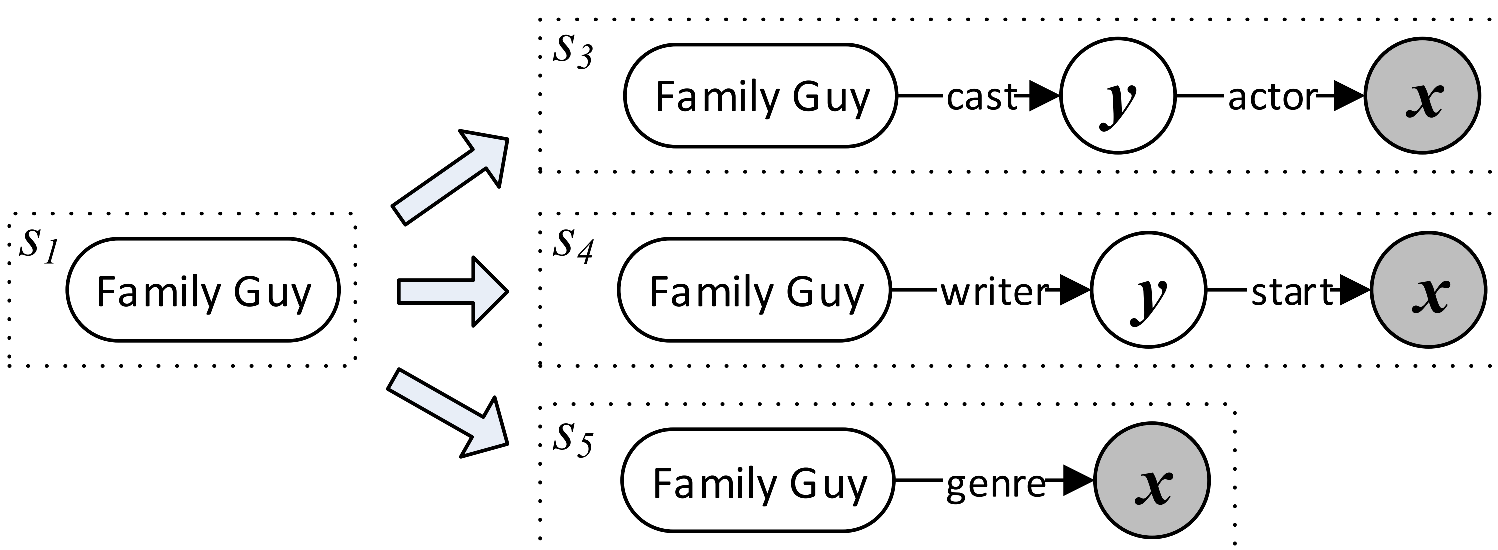
Step1： Lexicon-based candidate Generation
(anchor texts, redirect pages and others)

Step2: Selecting entities by using
Structured Learning (Top 10)

This example comes from Yinh's Slides in 2015

Identify Core Inferential Chain

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



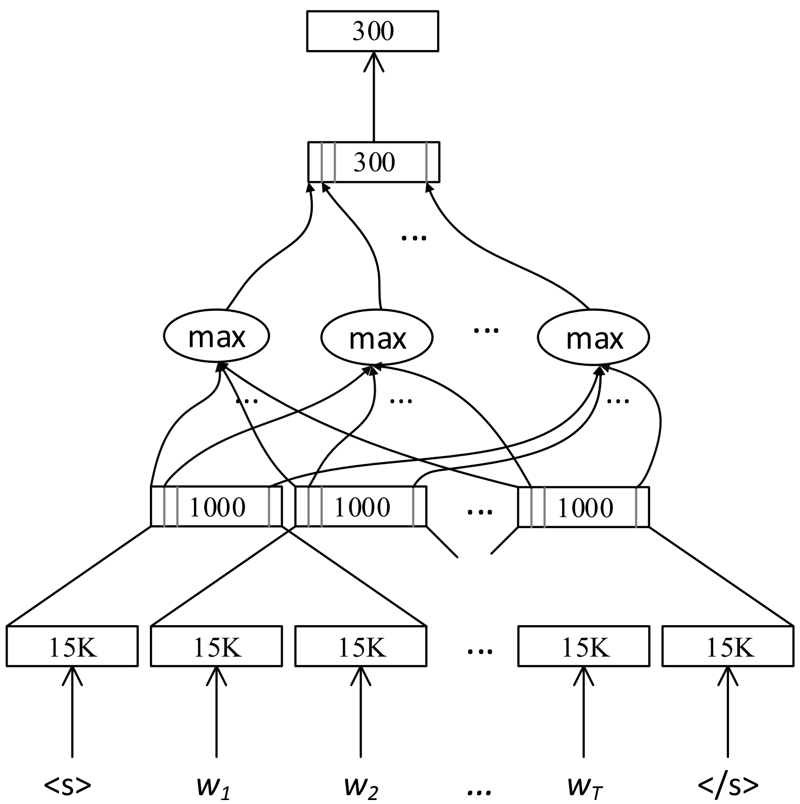
2 steps if y is a CVT-node
1 step if y is a non-CVT-node

Who first voiced Meg on Family Guy?

$\{cast\!-\!actor,\ writer\!-\!start,\ genre\}$

This example comes from Yinh's Slides in 2015

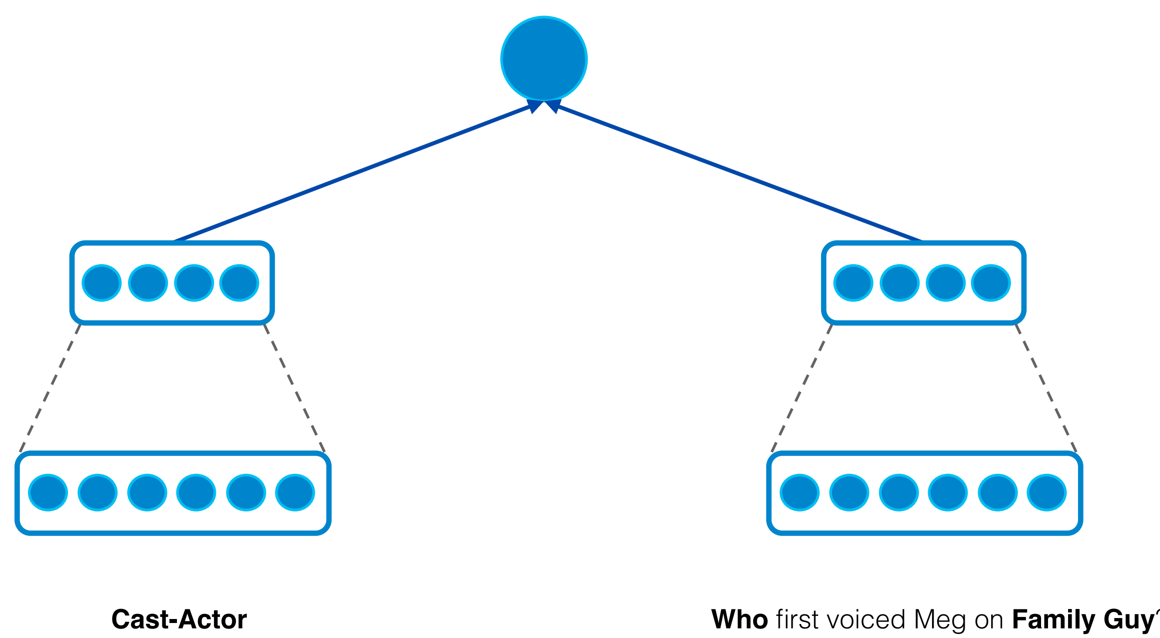
Using CNN



word hash: “cat” —> “#ca”,”cat”, “at#”

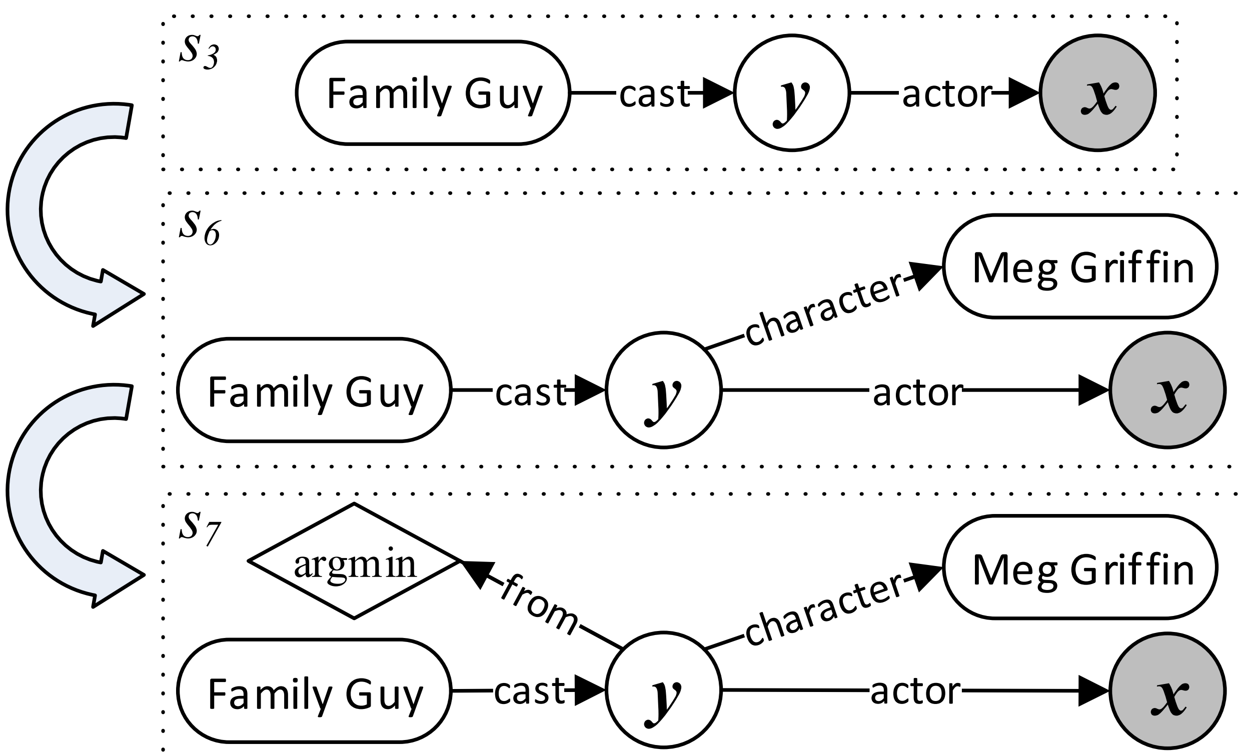
“cat” —>[0….1..1….0….1….0]

$$p(r_1 | q) = \frac{\exp(\cos(e_{r_1}, e_q))}{\sum_r \exp(\cos(e_r, e_q))}$$



Argument Constraints

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



Using rules to add constraints on the core inferential chain

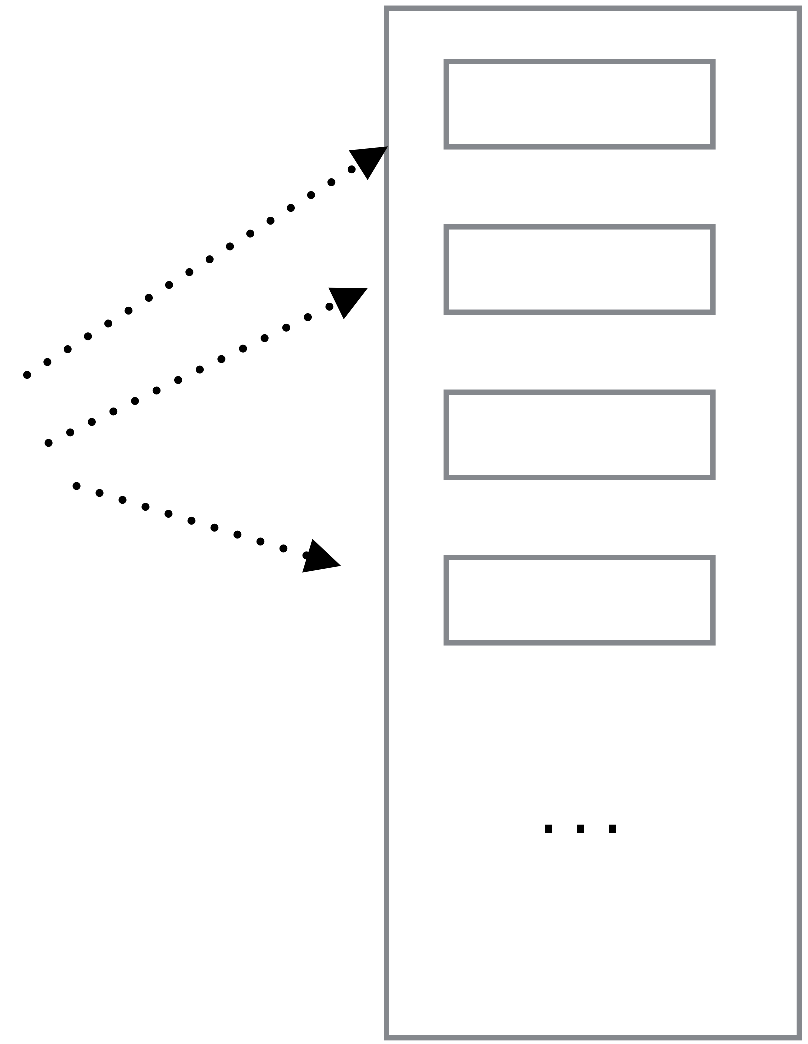
If x is a entity, it can be added as entity node

If x is such keywords, like “first”, “latest”, it could be added as aggregation constraints.

This example comes from Yinh’s Slides in 2015

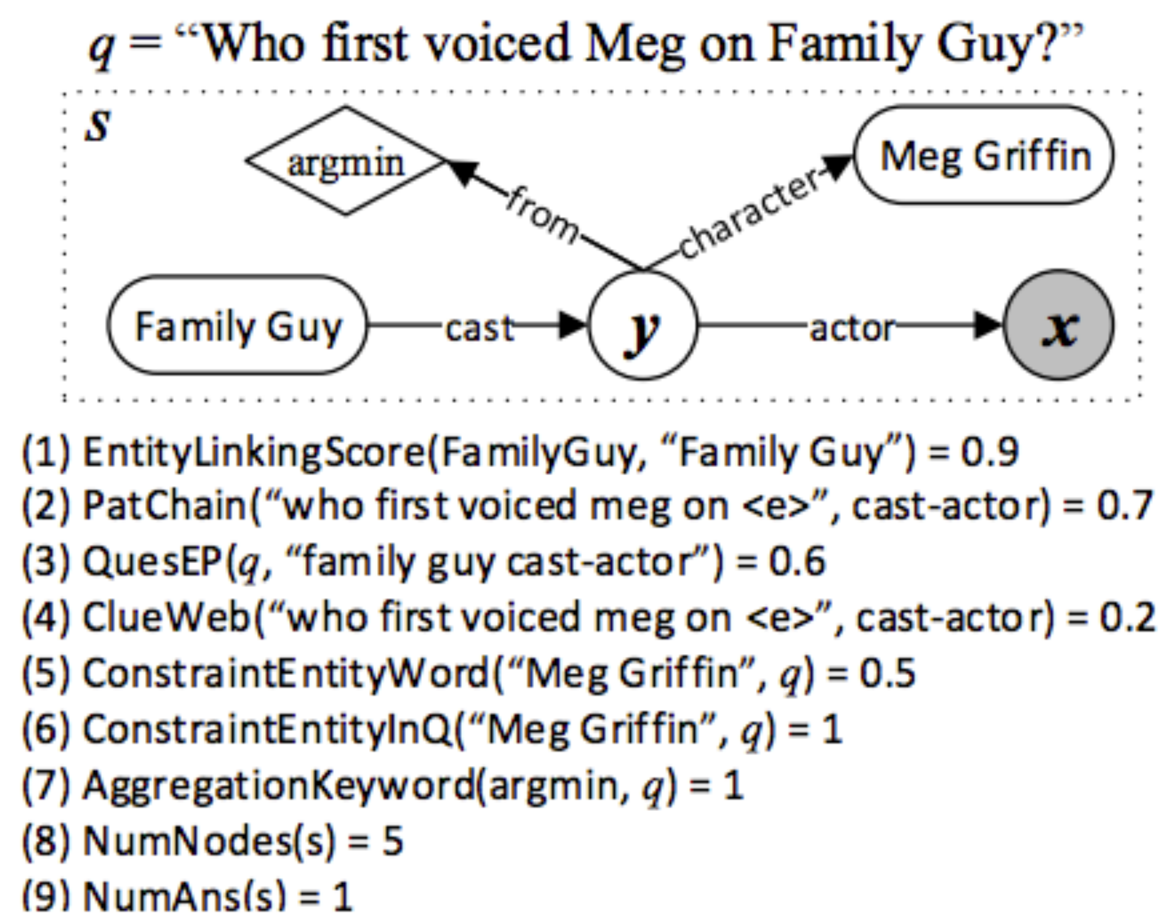
Ranking

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



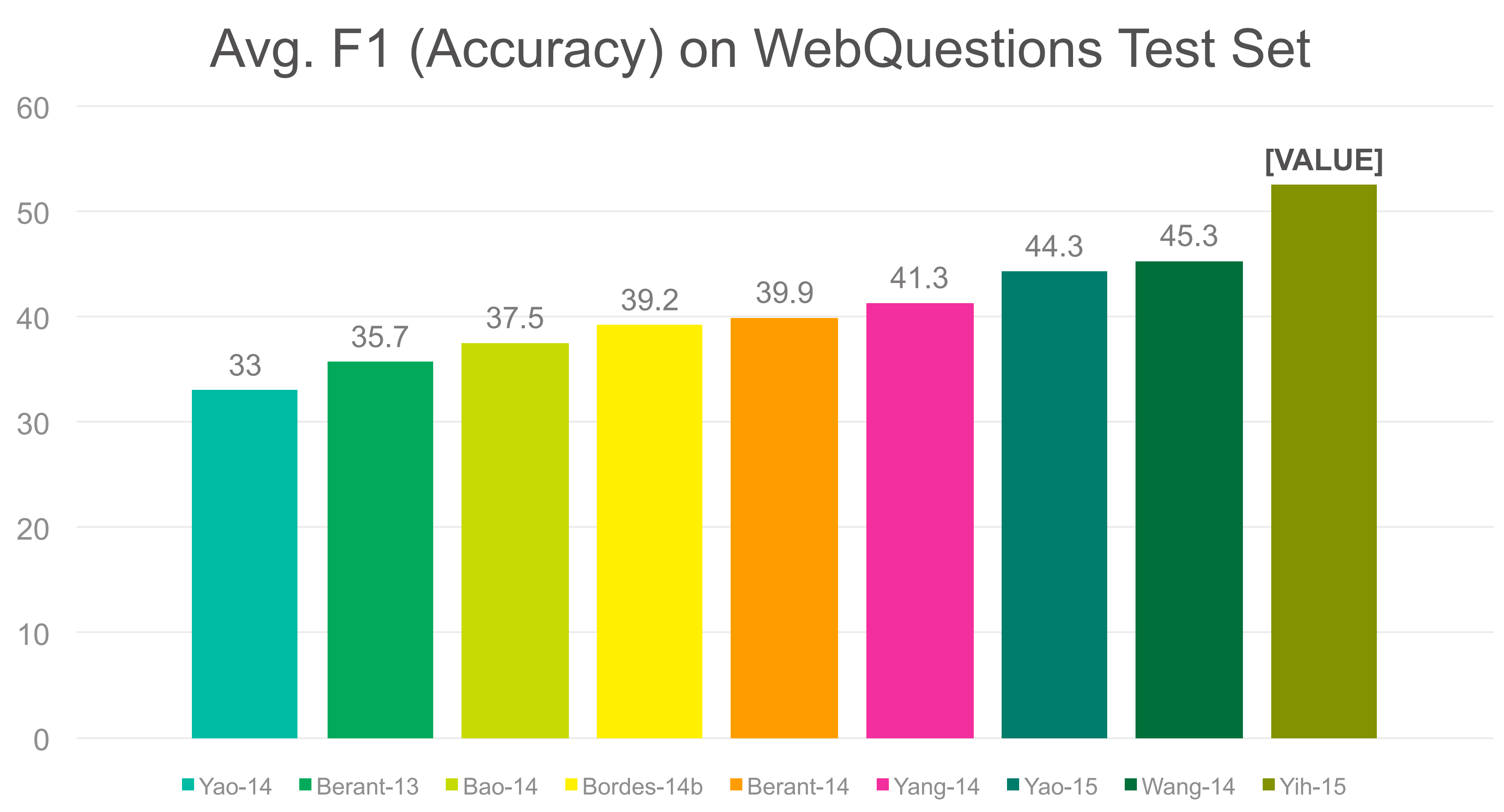
Ranking

- Log Linear Model
- Main Features:
  - Topic Entity: Entity Linking Score
  - Core Inferential Chain: Relation Matching Score (NN-based model)
  - Constraints: Keyword and entity matching



Results

- Bechmark: WebQuestions
  - 5,810 Q-A Pairs from google query log
  - Download: http://nlp.stanford.edu/software/sempre/



Key: Relation Classification

Who first voiced Meg on Family Guy?

{*cast−actor, writer−start, genre*}

Feature Representation

Labeled Training Data

Traditional Feature Representation

- 传统特征提取需要NLP预处理+人工设计的特征

The [haft]$_{e1}$ of the [axe]$_{e2}$ is made of yew wood

Component-Whole(e1,e2)

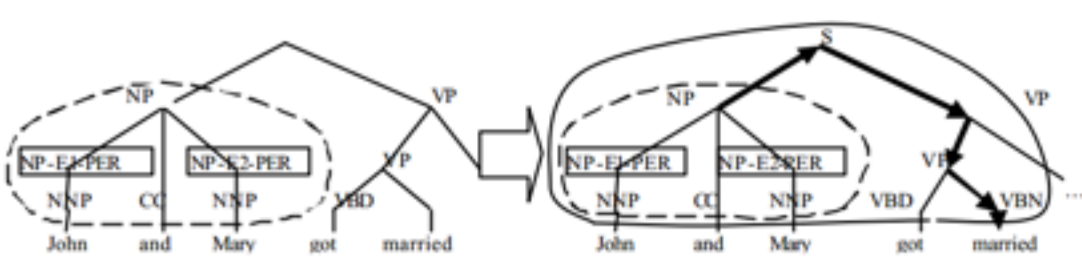
Traditional Features

Words： *haft*$_{m11}$, *of*$_{b1}$, *the*$_{b2}$, *axe*$_{21}$

Entity Type： *OBJECT*$_{m1}$, *PRODUCT*$_{m2}$

Parse Tree： *OBJECT-NP-PP-PRODUCT*

Kernel Feature:



- 问题1：对于缺少NLP处理工具和资源的语言，无法提取文本特征
- 问题2：NLP处理工具引入的“错误累积”

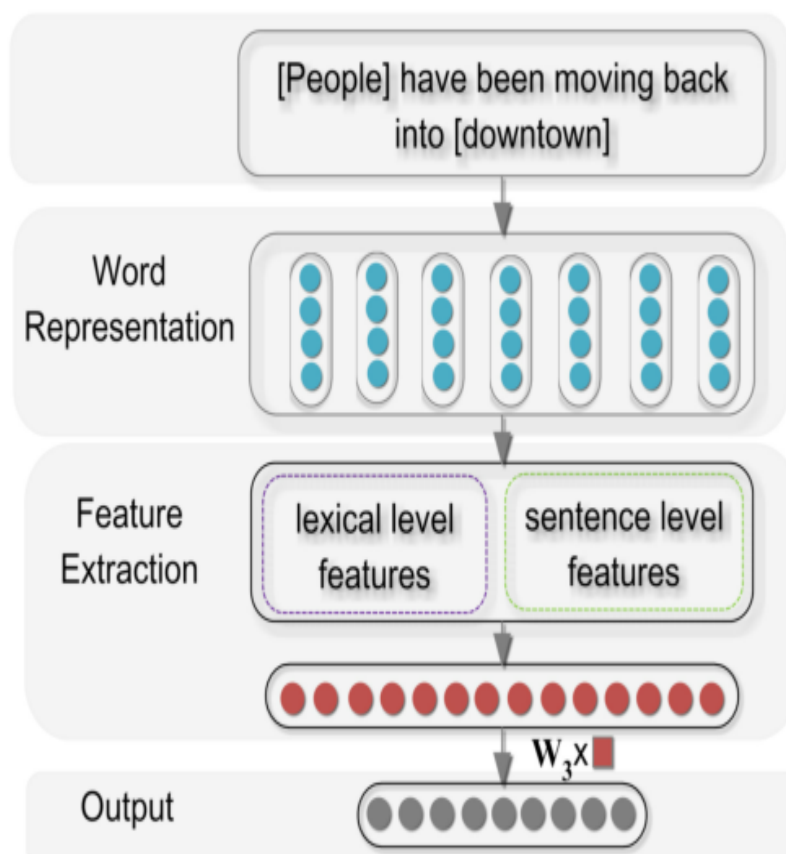
Relation Identification based on Deep Convolutional Neural Network (Zeng et al COLING 2014)

The [haft]$_{e1}$ of the [axe]$_{e2}$ is made of yew wood.

Component-Whole(e1,e2)

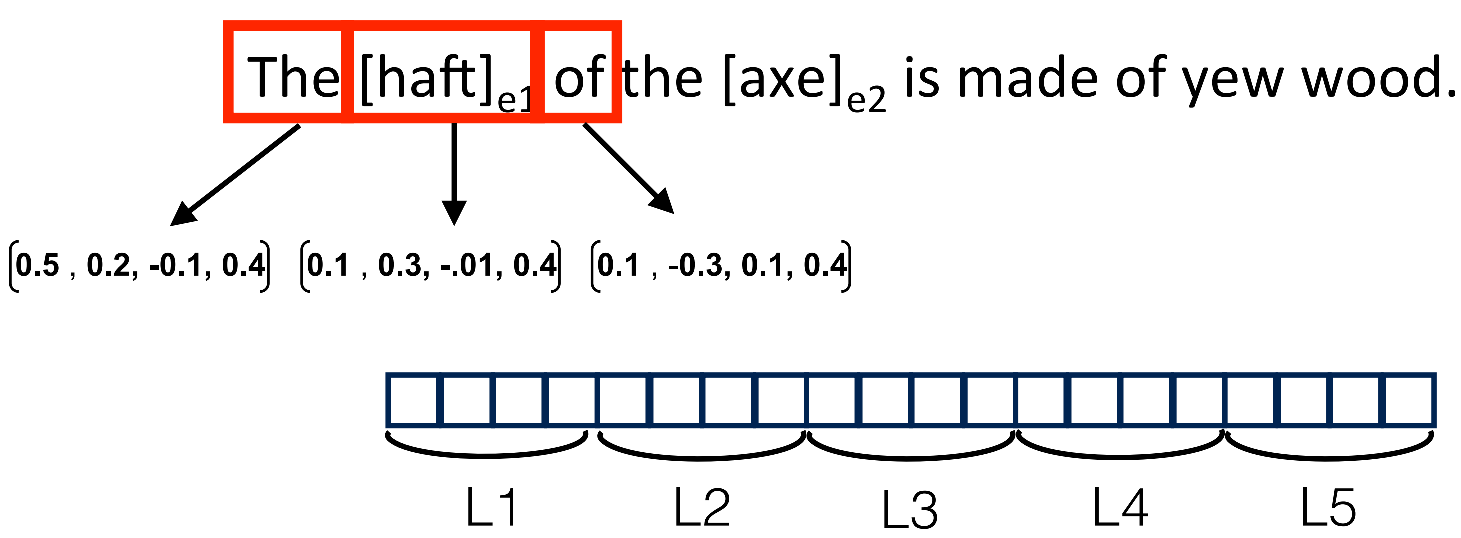
Lexical Level Features:
捕捉词本身的语义信息

Sentence Level Features:
捕捉所在句子的上下文信息

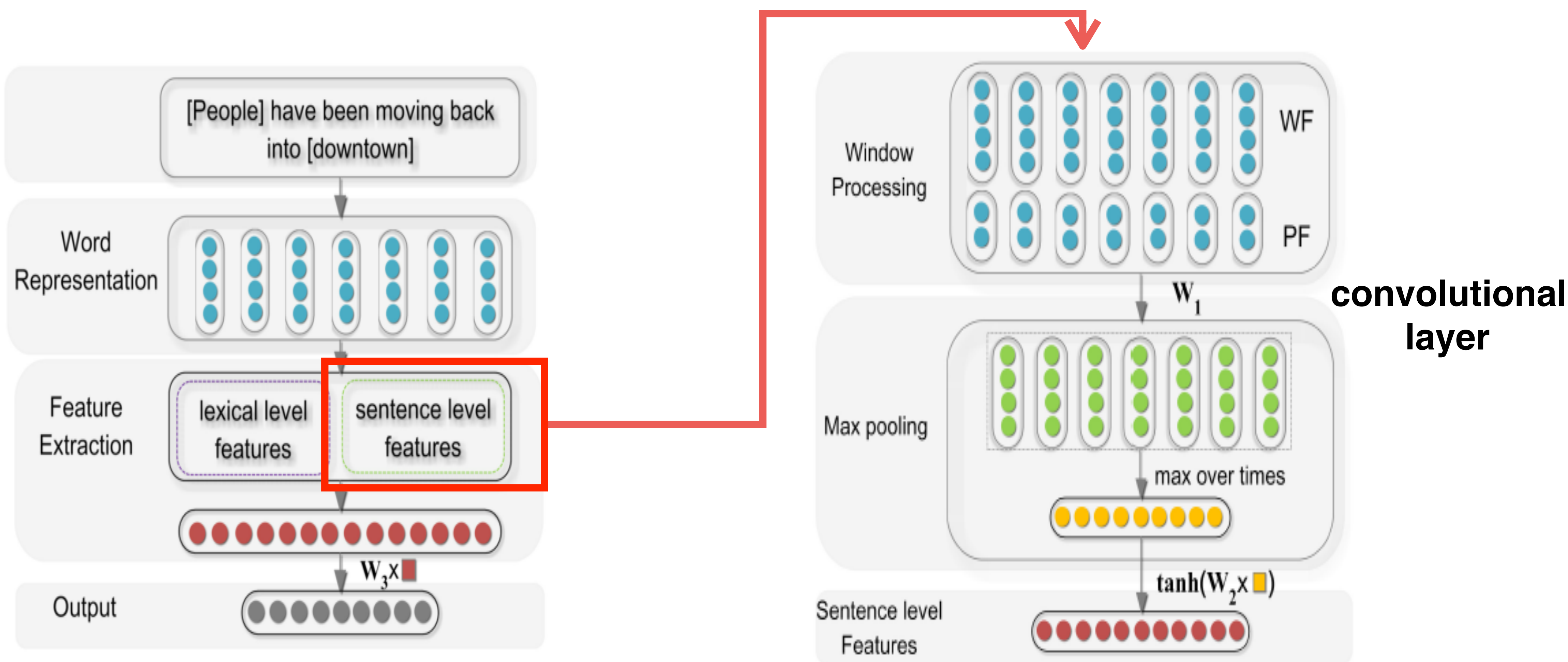


Lexical Level Features

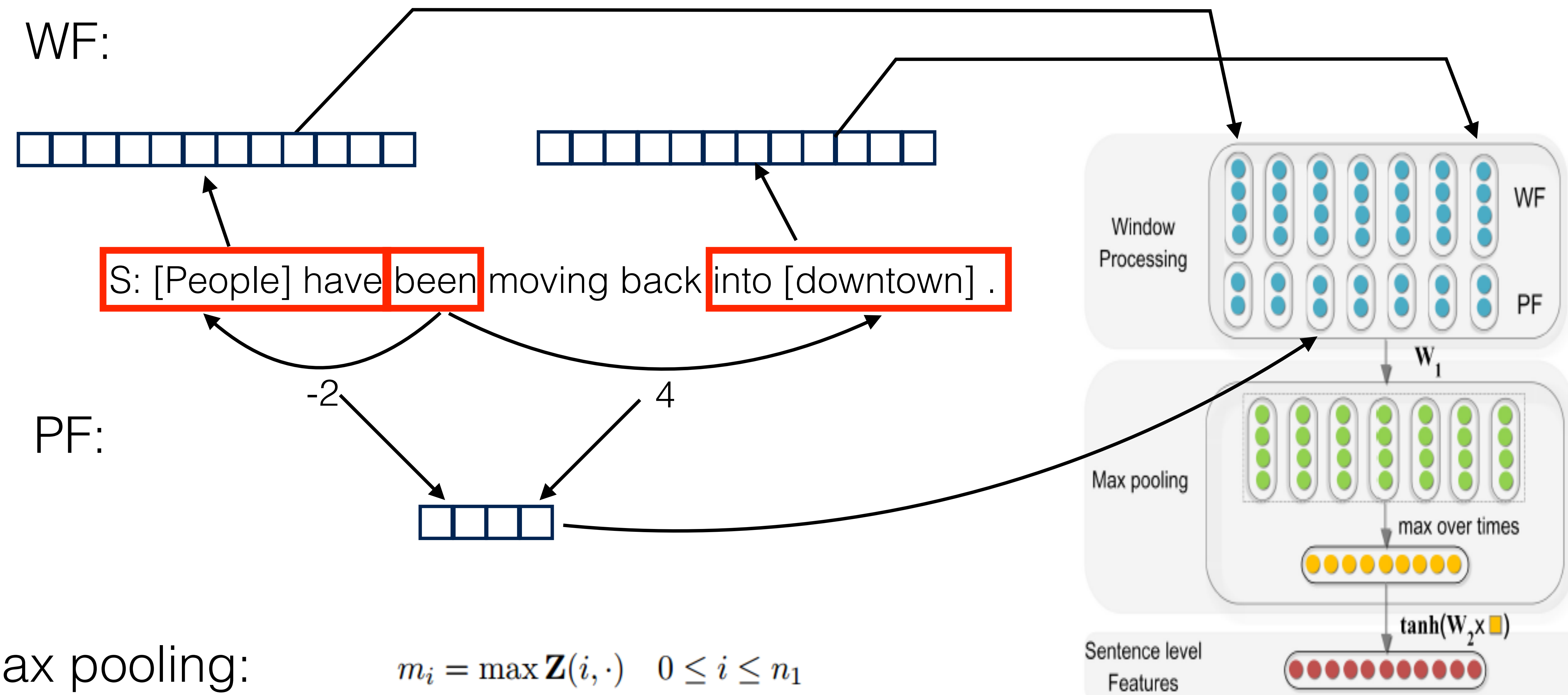
Features	Remark
L1	Noun 1
L2	Noun 2
L3	Left and right tokens of noun 1
L4	Left and right tokens of noun 2
L5	WordNet hypernyms of nouns



Sentence Level Features



Sentence Level Features

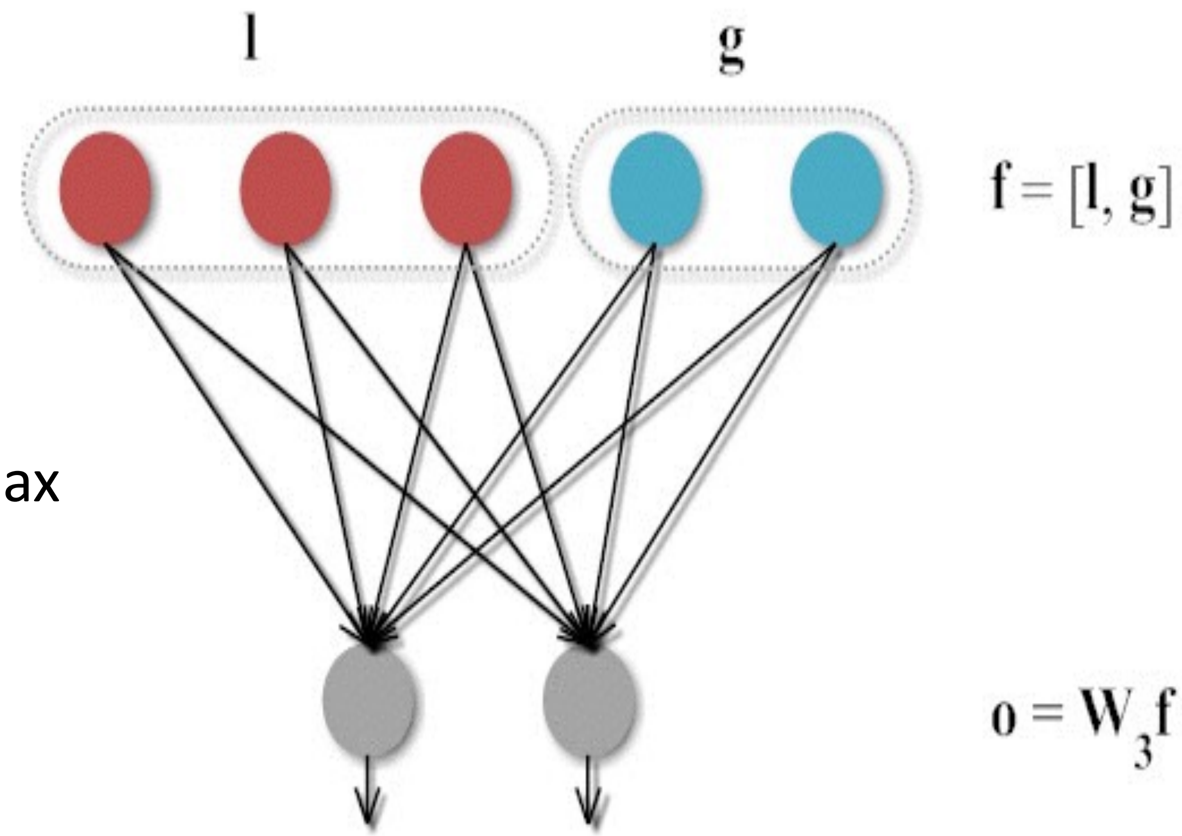


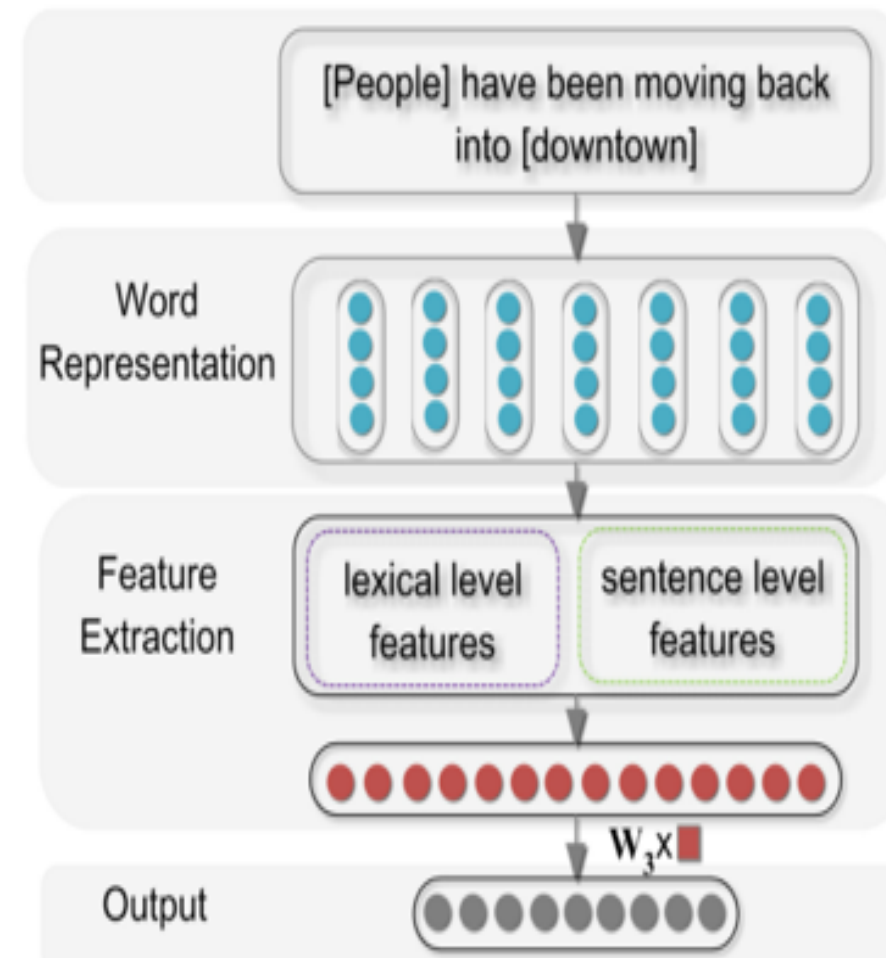
max pooling: $m_i = \max \mathbf{Z}(i, \cdot) \quad 0 \leq i \leq n_1$

output: $\mathbf{g} = \tanh(\mathbf{W}_2 \mathbf{m})$

Classification Layer

Softmax





$$p(i|x,\theta) = \frac{e^{o_i}}{\sum_{k=1}^{n_4} e^{o_k}}$$

实验结果

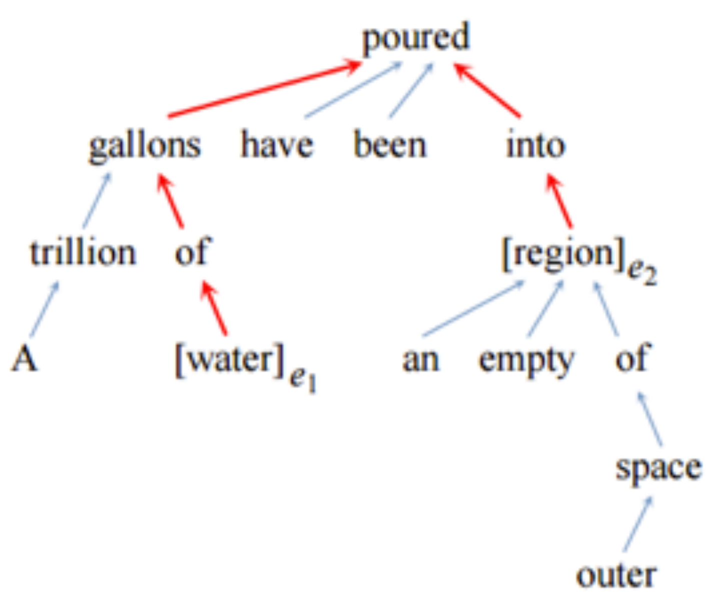
- SemEval-2010 Task 8

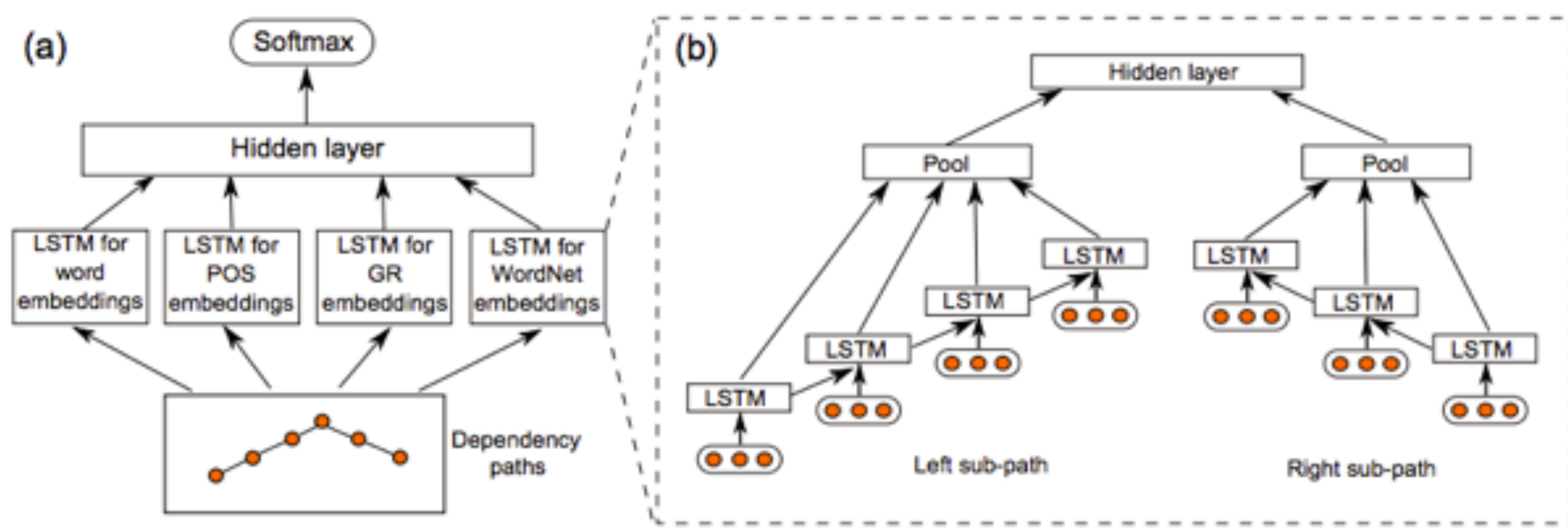
Classifier	Feature Sets	F1
SVM	POS, stemming, syntactic patterns	60.1
SVM	word pair, words in between	72.5
SVM	POS, stemming, syntactic patterns, WordNet	74.8
MaxEnt	POS, morphological, noun compound, thesauri, Google n-grams, WordNet	77.6
SVM	POS, prefixes, morphological, WordNet, dependency parse, Levin classed, ProBank, FrameNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
RNN	-	74.8
	POS, NER, WordNet, syntactic tree	77.6
MVRNN	-	79.1
	POS, NER, WordNet, syntactic tree	82.4
Proposed	word pair, WordNet	**82.7**

实验表明，我们所提出方法在需要NLP预处理和人工设计复杂特征前提下，能够有效提升实体关系分类性能

Long Shortest Memory Network Along Shortest Syntactic Path (Xu et al. EMNLP 2015)

“A trillion gallons of **water** have been poured into an empty **region** of outer space





Results

Classifier	Feature set	F_1
SVM	POS, WordNet, prefixes and other morphological features, depdency parse, Levin classes, PropBank, FanmeNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
RNN	Word embeddings	74.8
	Word embeddings, POS, NER, WordNet	77.6
MVRNN	Word embeddings	79.1
	Word embeddings, POS, NER, WordNet	82.4
CNN	Word embeddings	69.7
	Word embeddings, word position embeddings, WordNet	82.7
Chain CNN	Word embeddings, POS, NER, WordNet	82.7
FCM	Word embeddings	80.6
	Word embeddings, depedency parsing, NER	83.0
CR-CNN	Word embeddings	82.8†
	Word embeddings, position embeddings	82.7
	Word embeddings, position embeddings	**84.1**†
SDP-LSTM	Word embeddings	82.4
	Word embeddings, POS embeddings, WordNet embeddings, grammar relation embeddings	**83.7**

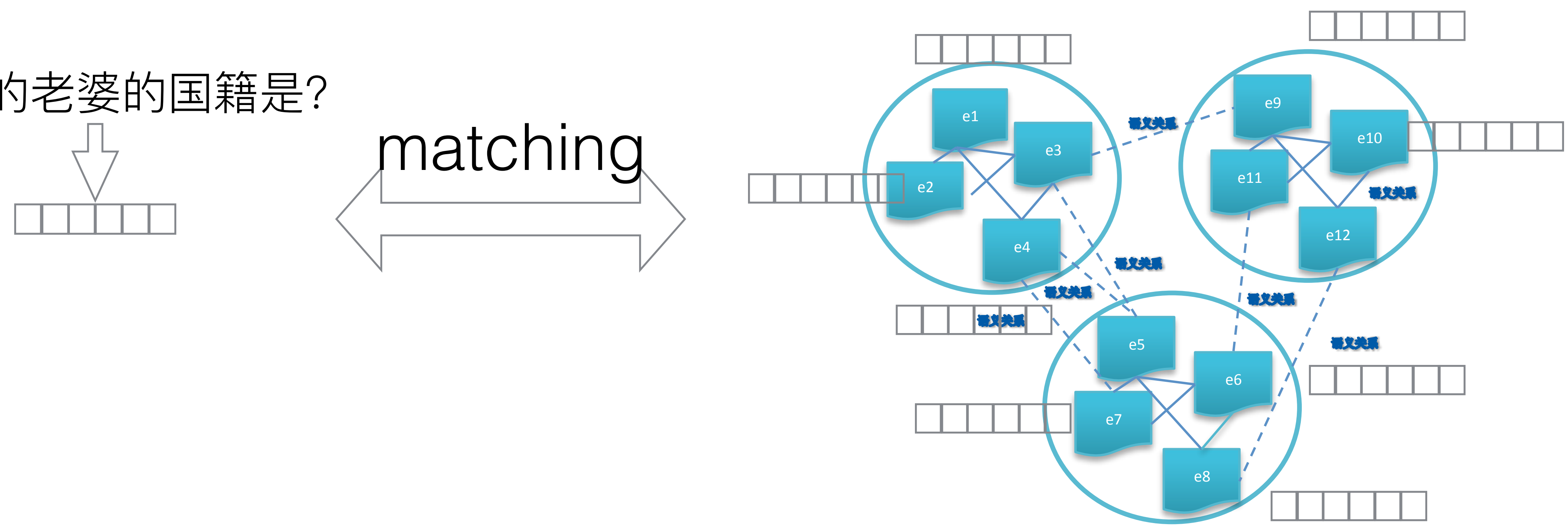
小结

- 已有DL方法对于问答的改进
  - 仍然基于传统知识库问答的基本流程
  - 核心仍然是基于符号表示
    - 语义关系识别：CNN、LSTM-RNN.....

End2End based QA

End2End

姚明的老婆的国籍是?

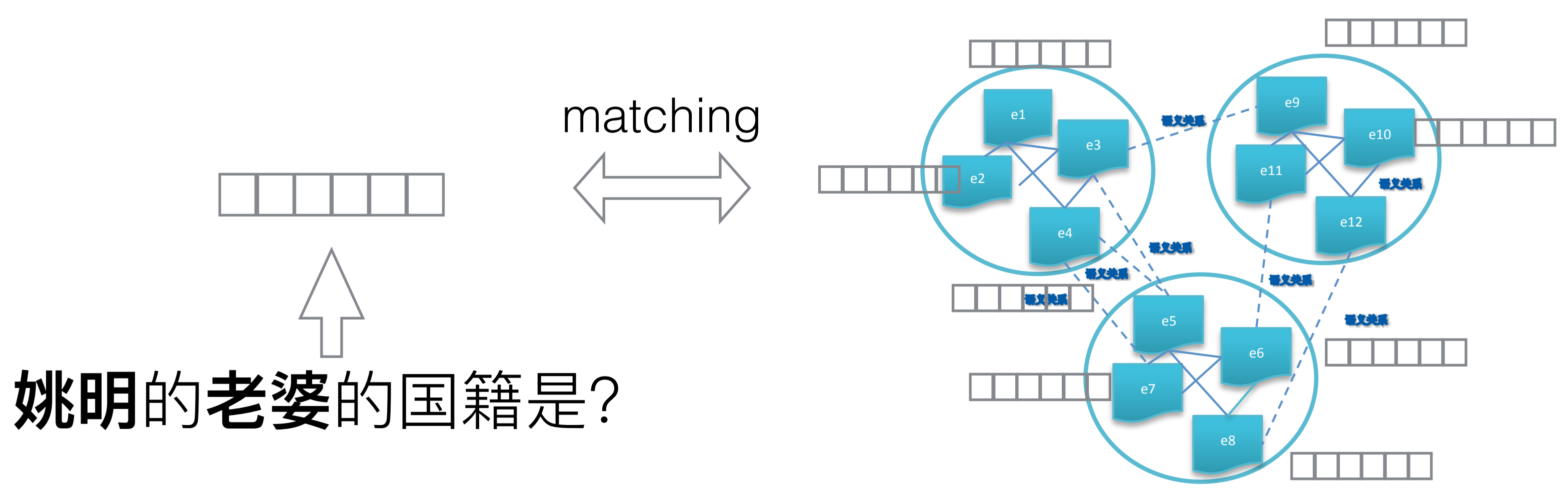


Progress

- Bordes et al. Open Question Answering with Weakly Supervised Embedding Models, In Proceedings of ECML-PKDD 14
  - Basic System
- Bordes et al. Question Answering with Subgraph Embedding, In Proceedings of EMNLP 14
  - Contextual Information of Answers
- Yang et al. Joint Relational Embeddings for Knowledge-Based Question Answering, In Proceedings of EMNLP 2014
  - Entity Type
- Dong et al. Question Answering over Freebase with Multi-Column Convolutional Neural Network, In Proceedings of ACL 2015
  - Topic Entity、Relation Path、Contextual Information
- Bordes et al. Large-scale Simple Question Answering with Memory Network, In Proceedings of ICIR 2015
  - Memory Network

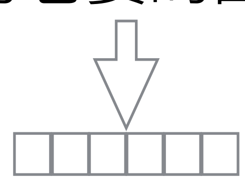
Basic End2End QA System (Bordes et al. 2014)

- 目前只处理单关系（Single Relation）的问句（Simple Question）
- 基本步骤
    - Step1：候选生成
        - 利用Entity Linking找到main entity
        - 在KB中main entity周围的entity均是候选
    - Step2：候选排序



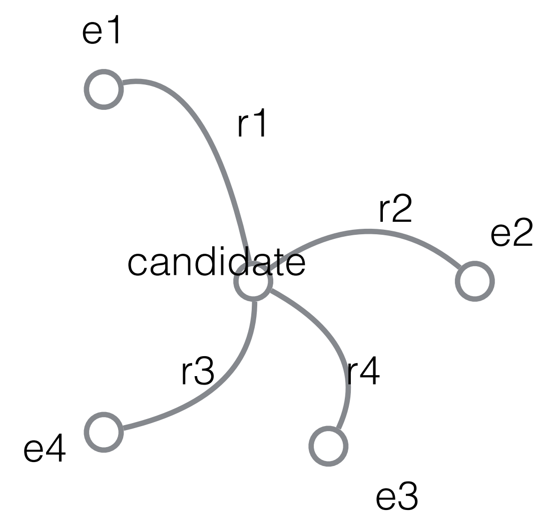
Basic End2End QA System (Bordes et al. 2014)

姚明的老婆的国籍是?



$$f(q_i) = \sum_{w_k \in q_i} E(w_k)$$

matching



$$g(q_i) = \sum_{e_k \in t_i} E(e_k)$$

Object: $L = 0.1 - S(f(q), g(t)) + S(f(q), g(t'))$

$$S(f(q), g(t)) = f(q)^T g(t)$$

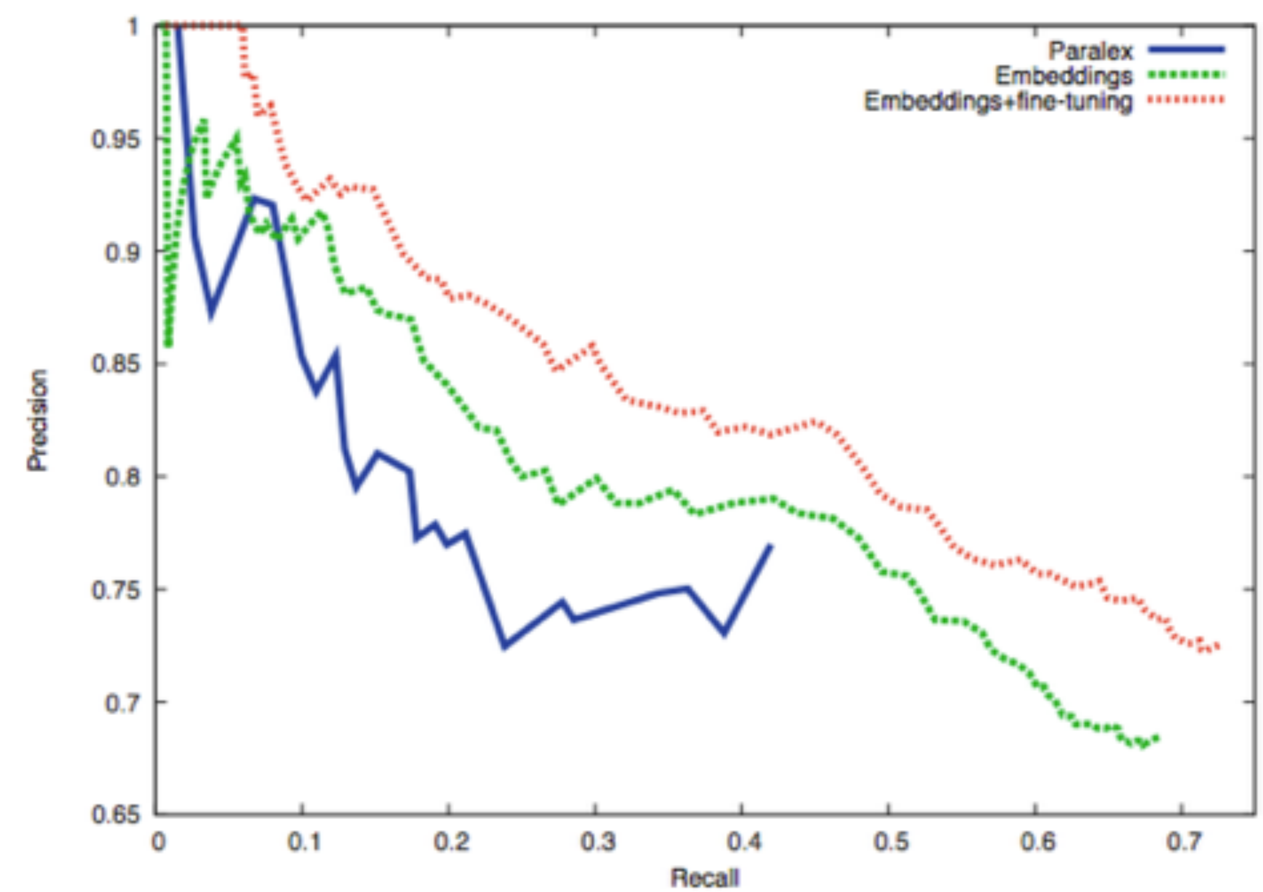
$$S(f(q), g(t)) = f(q)^T M g(t)$$

Multitask learning with paraphrases:

$$S_p(f(q), f(q_p)) = f(q)^T f(q_p)$$

Results

Method	F1	Prec	Recall	MAP
Paralex *(No. 2-arg)*	0.40	**0.86**	0.26	0.12
Paralex	0.54	0. 77	0.42	0.22
Embeddings	0.68	0.68	0.68	0.37
Embeddings *(no paraphrase)*	0.60	0.60	0.60	0.34
Embeddings *(incl. n-grams)*	0.68	0.68	0.68	0.39
Embeddings+fine-tuning	**0.73**	0.73	**0.73**	**0.42**



Considering More Info (Bordes et al. 2014)

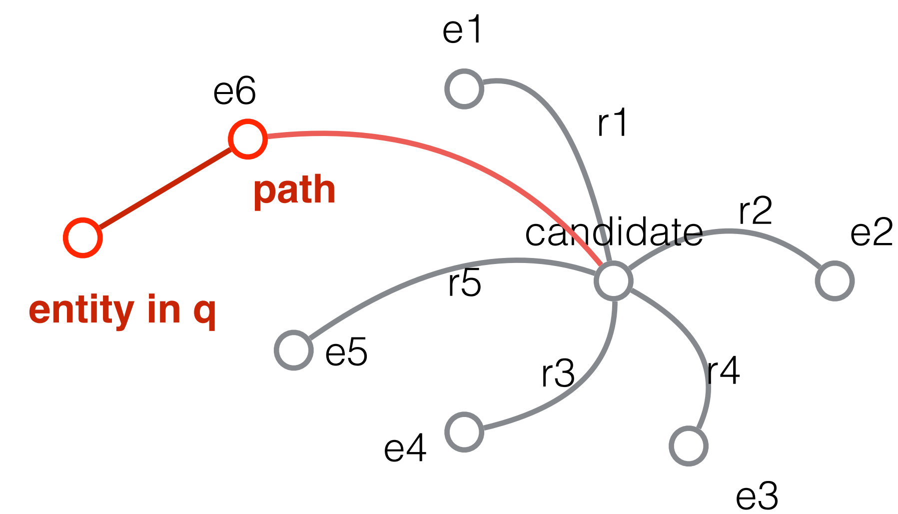
- Entity Information $E(e_k)$

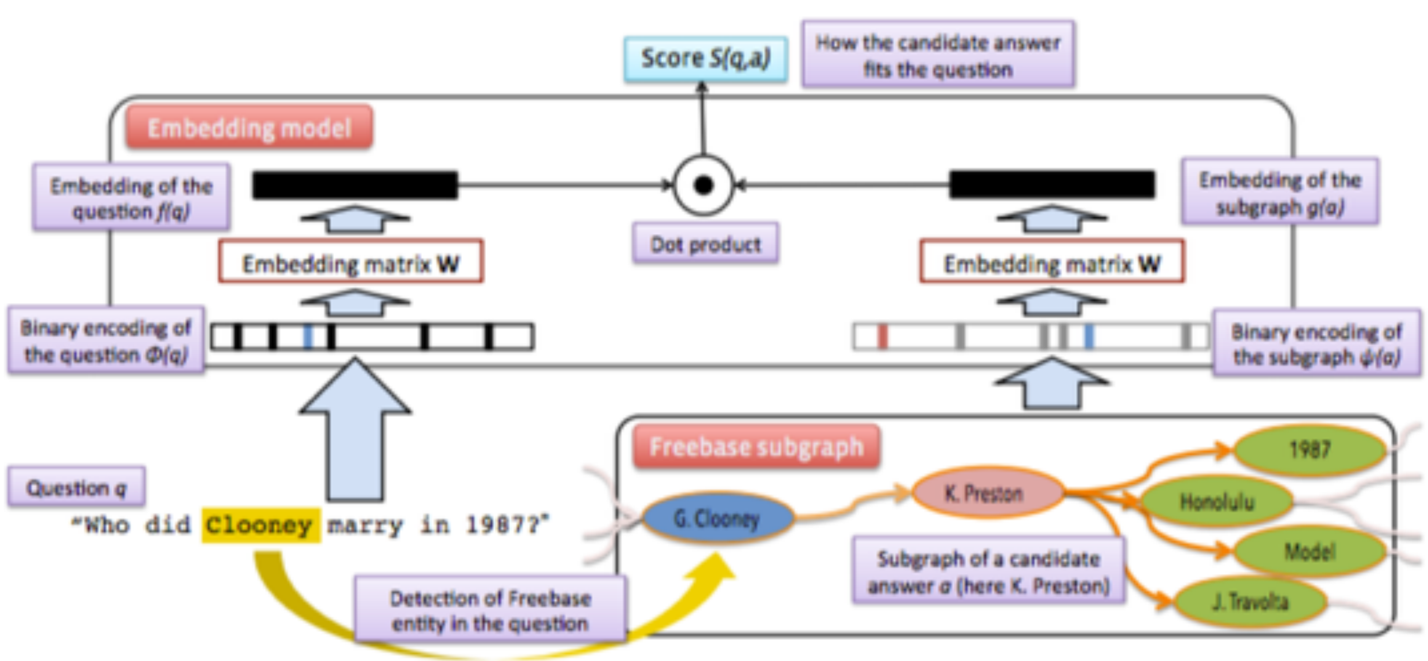
- Path to entities in question

$$g(q_i)=\sum_{e_k \in Path(t_i)} E(e_k)+\sum_{r_k \in Path(t_i)} E(r_k)$$

- Subgraph of the answers (contextual Info)

$$g(q_i)=\sum_{e_k \in Context(t_i)} E(e_k)+\sum_{r_k \in Context(t_i)} E(r_k)$$





$$\sum_{i=1}^{|\mathcal{D}|} \sum_{\bar{a} \in \bar{A}(a_i)} \max\{0, m - S(q_i, a_i) + S(q_i, \bar{a})\}$$

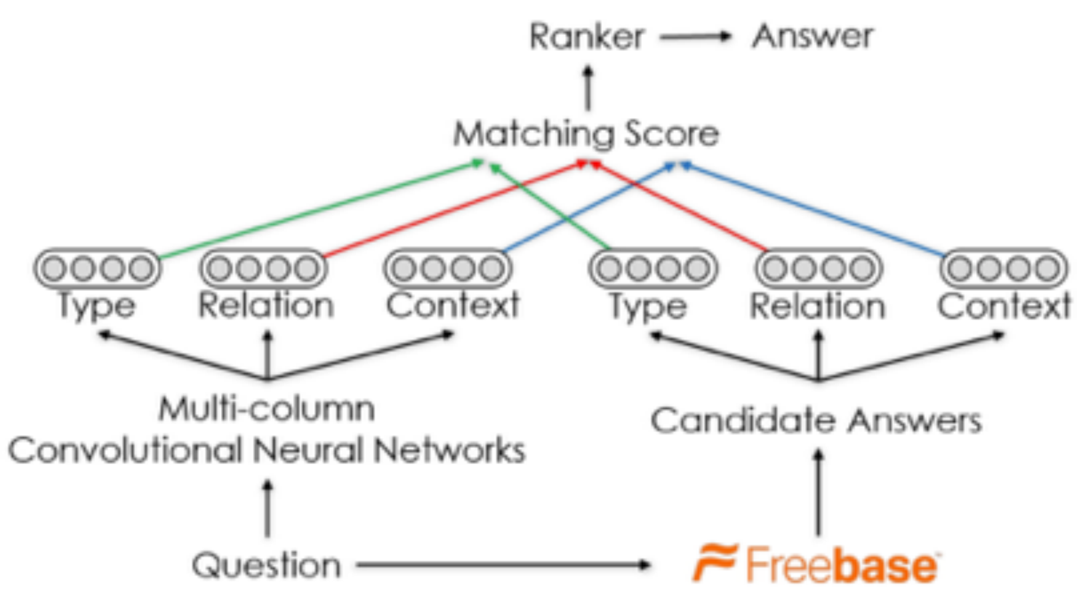
$$S(q,a)=f(q)^T g(a)$$

Results

Method	P@1 (%)	F1 (Berant)	F1 (Yao)
Baselines			
(Berant et al., 2013) [1]	–	31.4	–
(Bordes et al., 2014) [5]	31.3	29.7	31.8
(Yao and Van Durme, 2014) [14]	–	33.0	42.0
(Berant and Liang, 2014) [2]	–	39.9	43.0
Our approach			
Subgraph & $\mathcal{A}(q) = C_2$	**40.4**	39.2	43.2
Ensemble with (Berant & Liang, 14)	–	**41.8**	**45.7**
Variants			
Without multiple predictions	**40.4**	31.3	34.2
Subgraph & $\mathcal{A}(q) =$ All 2-hops	38.0	37.1	41.4
Subgraph & $\mathcal{A}(q) = C_1$	34.0	32.6	35.1
Path & $\mathcal{A}(q) = C_2$	36.2	35.3	38.5
Single Entity & $\mathcal{A}(q) = C_1$	25.8	16.0	17.8

Multi-Column CNN (Dong et al. 2015)

- 依据问答特点，考虑答案不同维度的信息

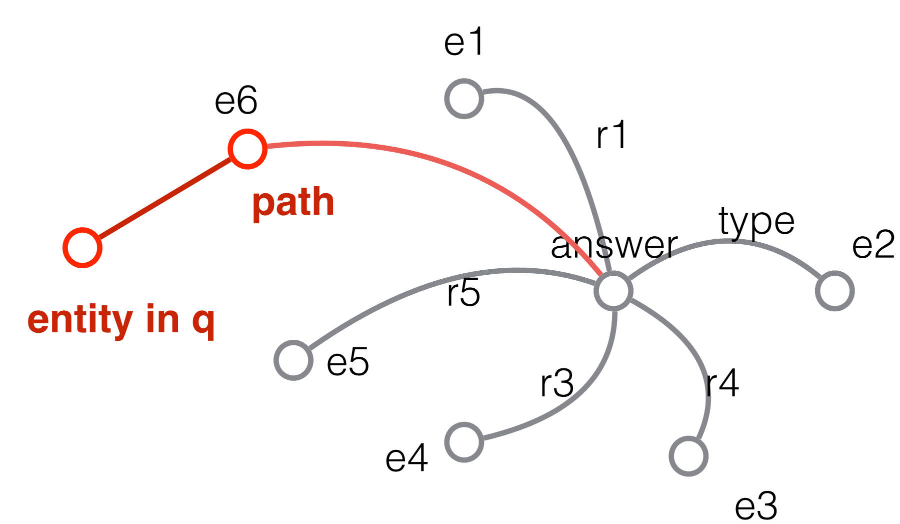


$$S(q,a) = \underbrace{\mathbf{f}_1(q)^{\mathrm{T}}\mathbf{g}_1(a)}_{answer\ path} + \underbrace{\mathbf{f}_2(q)^{\mathrm{T}}\mathbf{g}_2(a)}_{answer\ context} + \underbrace{\mathbf{f}_3(q)^{\mathrm{T}}\mathbf{g}_3(a)}_{answer\ type}$$

Answer Type

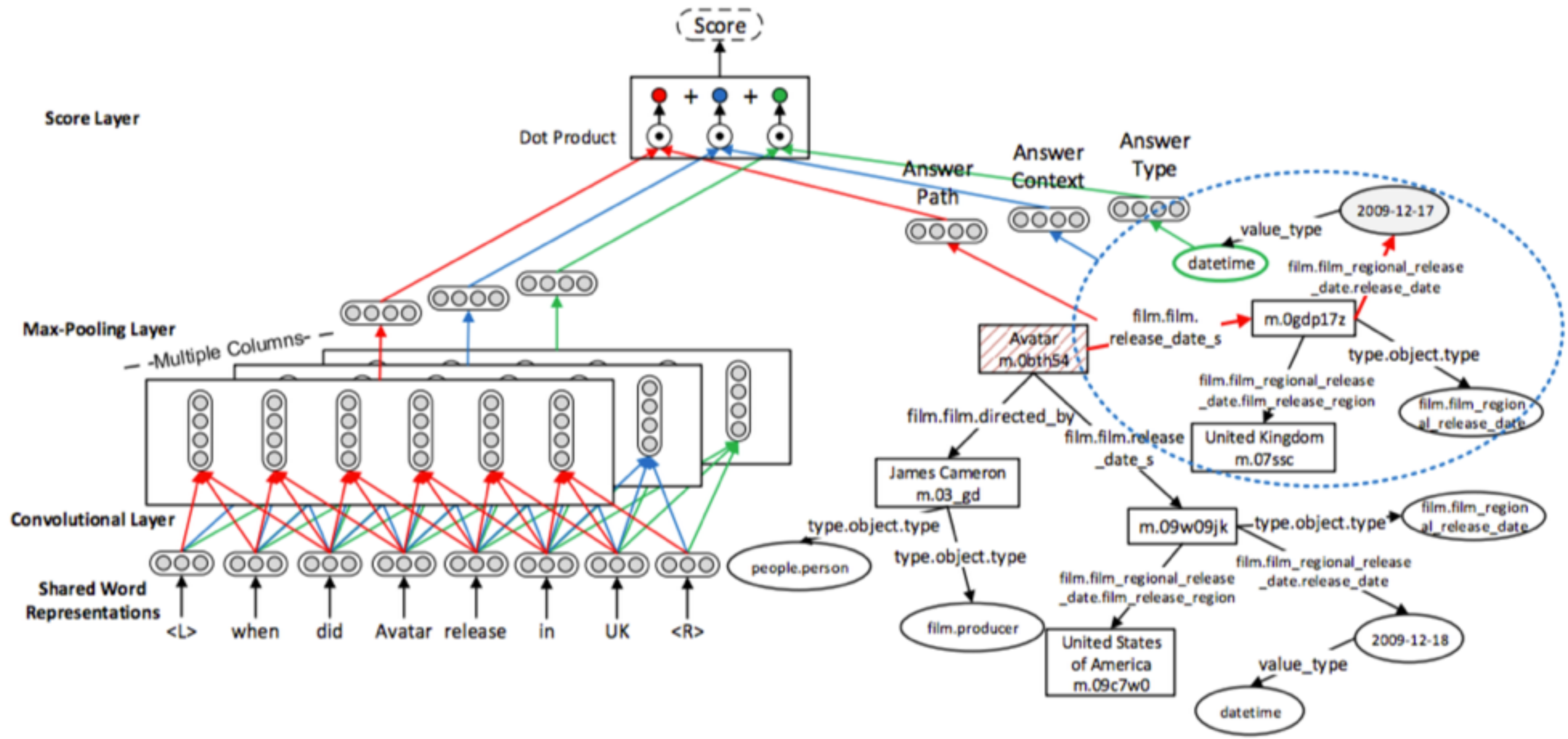
Answer Context

Answer Path



Framework

when did Avatar release in UK



Results

Method	F1	P@1
(Berant et al., 2013)	31.4	-
(Berant and Liang, 2014)	39.9	-
(Bao et al., 2014)	37.5	-
(Yao and Van Durme, 2014)	33.0	-
(Bordes et al., 2014a)	39.2	40.4
(Bordes et al., 2014b)	29.7	31.3
MCCNN (our)	**40.8**	**45.1**

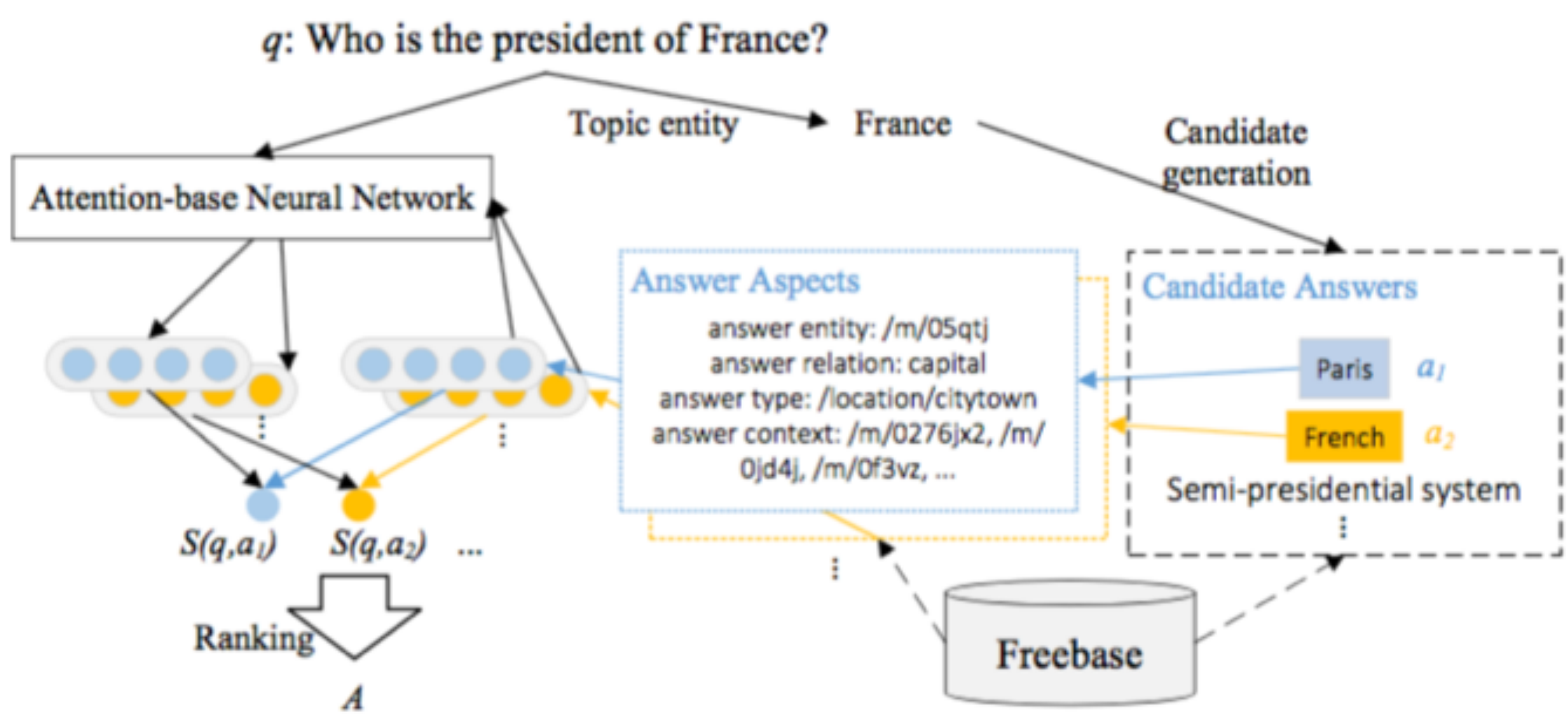
Neural Attention-based Neural Model for QA with Combining Global Knowledge Information

(Zhang et al. 2016)

- 问题
  - 问句表示
    - 已有方法多采用Word Embedding的平均对于问句进行语义表示，问句表示过于简单
    - 关注答案不同的部分，问句的表示应该是不一样的
  - 知识表示
    - 受限于训练语料
    - 未考虑全局信息

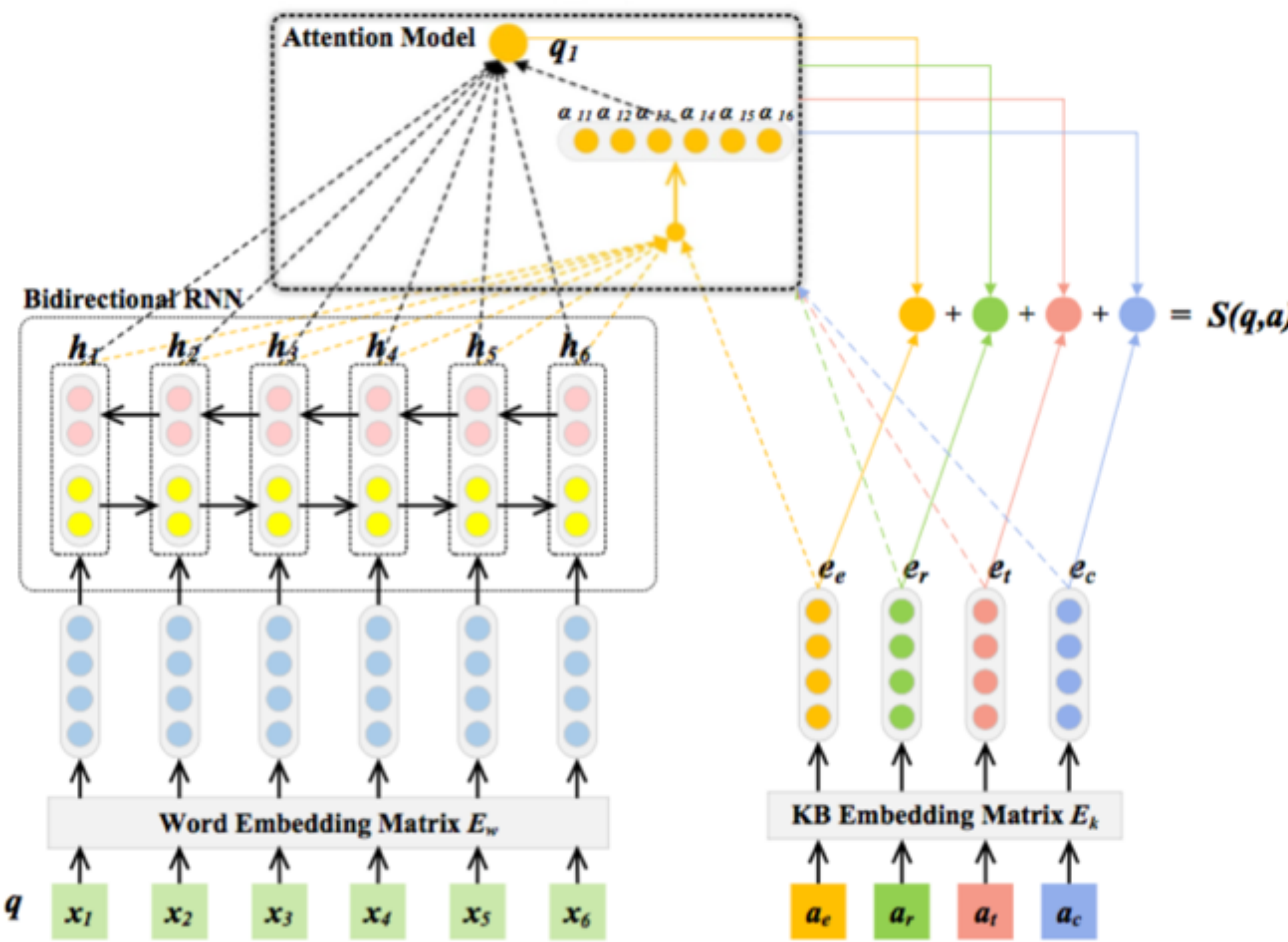
Neural Attention-based Neural Model for QA with Combining Global Knowledge Information

(Zhang et al. 2016)



Neural Attention-based Neural Model for QA with Combining Global Knowledge Information

(Zhang et al. 2016)



$$\alpha_{ij} = \frac{\exp(w_{ij})}{\sum_{k=1}^{L_q} \exp(w_{ik})}$$

$$w_{ij} = W^T(tanh[h_j; e_i]) + b$$

$$q_i = \sum_{j=1}^{L_q} \alpha_{ij} h_j$$

$$S(q,a) = \sum_{e_i \in \{e_e, e_r, e_t, e_c\}} q_i \cdot e_i$$

$$L_{q,a,a'} = [\gamma + S(q,a') - S(q,a)]_+$$

Neural Attention-based Neural Model for QA with Combining Global Knowledge Information

(Zhang et al. 2016)

- 融入全局信息
  - 利用TransE得到knowledge embedding
- Multi-task Learning

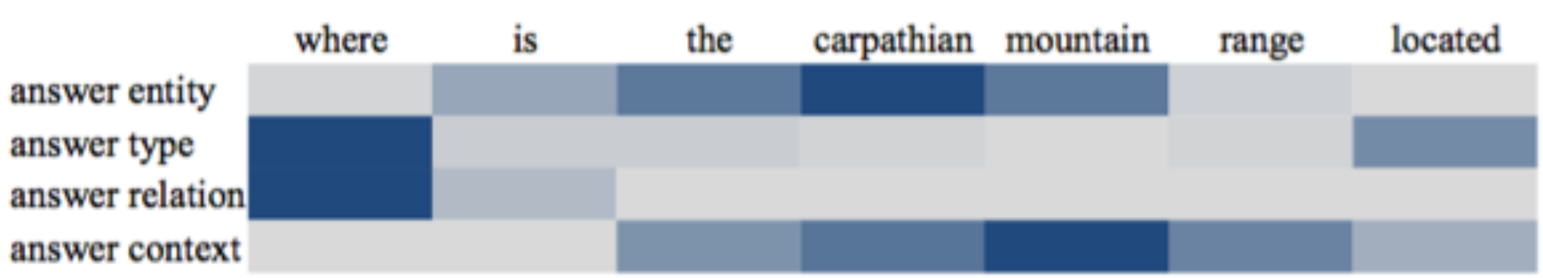
$$\begin{cases} L_{q,a,a'} = [\gamma + S(q,a') - S(q,a)]_+ \\ L_k = \sum_{(s,p,o)\in S} \sum_{(s',p,o')\in S'} [\gamma_k + d(s+p,o) - d(s'+p,o')]_+ \end{cases}$$

TransE

Experimental Results

Method	F_1
Bordes et al., 2014b	29.7
Bordes et al., 2014a	39.2
Yang et al., 2014	41.3
Dong et al., 2015	40.8
Bordes et al., 2015	42.2
ours	**42.6**

Method	F_1
LSTM	38.2
Bi_LSTM	38.9
Bi_LSTM + ATT	41.6
Bi_LSTM + GKI	40.4
Bi_LSTM + ATT + GKI	42.6

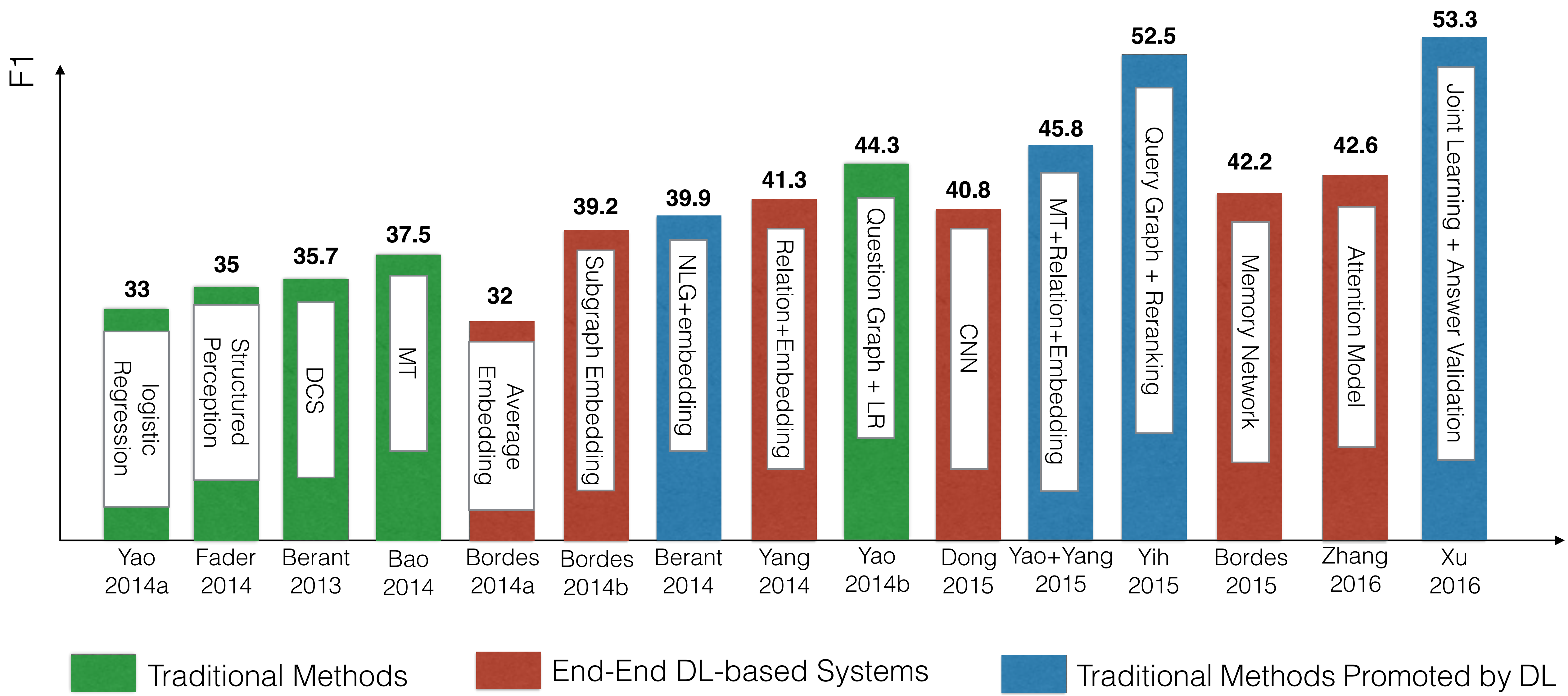


entity: Slovakia
type: /location/country
relation: partially/containedby
context: /m/04dq9kf, /m/01mp…

小结

- 基于DL的End2End问答基于检索框架，其核心是计算知识库中实体、关系与当前问句在语义向量空间中的语义相似度
  - 需要依据知识库问答的任务特点设计神经网络（实体类型、语义关系、上下文结构）
  - 目前，End2End问答仍只能解决Simple（单关系）类型的问题
  - 对于复杂问题、推理性问题还有待解决

Comparison on WebQuestion



参考

公开的评测数据集

数据集	# 训练集	# 测试集	知识库	形式	发布时间
ATIS	8297	3211	ATIS	答案	1994
Geo880	880		GeoBase	逻辑形式	2001
QALD-1	50	50	DBpedia	逻辑形式 & 答案	2011
QALD-2	100	99	DBpedia & YAGO	逻辑形式 & 答案	2012
QALD-3	100	99	DBpedia & YAGO	逻辑形式 & 答案	2013
Free917	641	276	Freebase	答案	2013
WebQuestion	3782	2037	Freebase	答案	2013
WikiAnswers	2.4M	698	Reverb	答案	2013
QALD-4	100	50	DBpedia & YAGO	逻辑形式 & 答案	2014
QALD-5	170	59	DBpedia & YAGO	逻辑形式 & 答案	2015
SimpleQuestions	86755	21687	Freebase & Reverb	答案	2015

Reference

[Zelle, 1995] J. M. Zelle and R. J. Mooney, "Learning to parse database queries using inductive logic programming," in Proceedings of the National Conference on Artificial Intelligence, 1996, pp. 1050–1055.

[Wong, 2007] Y. W. Wong and R. J. Mooney, "Learning synchronous gram-mars for semantic parsing with lambda calculus," in Proceedings of the 45th Annual Meeting-Association for computational Linguistics, 2007.

[Lu, 2008] W. Lu, H. T. Ng, W. S. Lee, and L. S. Zettlemoyer, "A generative model for parsing natural language to meaning representations," in Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, 2008, pp. 783–792.

[Zettlemoyer, 2005] L. S. Zettlemoyer and M. Collins, "Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars," in Proceedings of the 21st Uncertainty in Artificial Intelligence, 2005, pp. 658–666.

[Clarke, 2010] J. Clarke, D. Goldwasser, M.-W. Chang, and D. Roth, "Driving semantic parsing from the world's response," inProceedings of the 14th Conference on Computational Natural Language Learning, 2010, pp. 18–27

[Liang, 2011] P. Liang, M. I. Jordan, and D. Klein, "Learning dependency-based compositional semantics," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011, pp. 590–599.

[Cai, 2013] Qingqing Cai and Alexander Yates. 2013. Large-scale semantic parsing via schema matching and lexicon extension. In ACL.

[Berant, 2013]Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In EMNLP.

[Yao, 2014] Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In ACL.

[Unger, 2012] Christina Unger, Lorenz B¨ uhmann, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, Daniel Gerber, and Philipp Cimiano. 2012. Template-based question answering over rdf data. In WWW.

[Yahya, 2012]Mohamed Yahya, Klaus Berberich, Shady Elbas-suoni, Maya Ramanath, Volker Tresp, and Gerhard Weikum. 2012. Natural language questions for the web of data. In EMNLP.

[He, 2014] S. He, K. Liu, Y. Zhang, L. Xu, and J. Zhao. 2014. Question answering over linked data using first-order logic. In EMNLP.

[Artzi, 2011]Y. Artzi and L. Zettlemoyer. 2011. Bootstrap-ping semantic parsers from conversations. In Empirical Methods in Natural Language Processing (EMNLP), pages 421–432. Dan Goldwasser, Roi Reichart, James Clarke and Dan Roth. Confidence Driven Unsupervised Semantic Parsing. Proceedings of ACL (2011)

[Krishnamurthy, 2012] Jayant Krishnamurthy and Tom M. Mitchell. Weakly Supervised Training of Semantic Parsers. In Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2012

[Reddy, 2014] S. Reddy, M. Lapata, and M. Steedman. 2014. Large-scale semantic parsing without question-answer pairs. Transactions of the Association for Compu-tational Linguistics (TACL), 2(10):377–392.

[Fader, 2013] A. Fader, L. S. Zettlemoyer, and O. Etzioni, "Paraphrase-driven learning for open question answering." in Proceedings of the 51th Annual Meeting-Association for computational Linguistics, 2013, pp. 1608–1618.

Reference

[Bordes, 2014a] Antoine Bordes, Jason Weston, and Nicolas Usunier, "Open Question Answering with Weakly Supervised Embedding Models", In Proceedings of ECML-PKDD, 2014.
[Bordes, 2014b] Antoine Bordes, Sumit Chopra, and Jason Weston, "Question Answering with Subgraph Embedding", In Proceedings of EMNLP 2014.
[Yang, 2014] Min-Chul Yang, Nan Duan, Ming Zhou, and Hae-Chang Rim, "Joint Relational Embeddings for Knowledge-Based Question Answering", In Proceedings of EMNLP 2014.
[Dong, 2015] Li Dong, Furu Wei, Ming Zhou, and Ke Xu, "Question Answering over Freebase with Multi-Column Convolutional Neural Network", In Proceedings of ACL 2015.
[Bordes, 2015] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston, "Large-scale Simple Question Answering with Memory Network", In Proceedings of ICIR 2015.
[Yih, 2015] Wen-tau Yih, Ming
Wei Chang, Xiaodong He, and Jianfeng Gao, "Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base", In Proceedings of ACL 2015. (Outstanding Paper)
[Zeng, 2014] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou and Jun Zhao, "Relation Classification via Convolutional Deep Neural Network", in Proceedings of COLING 2014. (Best Paper)
[Zeng, 2015] Daojian Zeng, Kang Liu, Yubo Chen and Jun Zhao, "Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks", in Proceedings of EMNLP 2015.
[Xu, 2015] Xu Yan, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin, "Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths", In Proceedings of EMNLP 2015.
[Berant, 2014] Jonathan Berant, and Percy Liang, "Semantic Parsing via Paraphrasing", In Proceeding of ACL 2014. (Best long paper honorable mention)
[Zhang, 2016] Yuanzhe Zhang, Shizhu He, Kang Liu and Jun Zhao. A Joint Model for Question Answering over Multiple Knowledge Bases, To appear in Proceedings of AAAI 2016, Phoenix, USA, Phoenix.

谢谢! Q&A!