

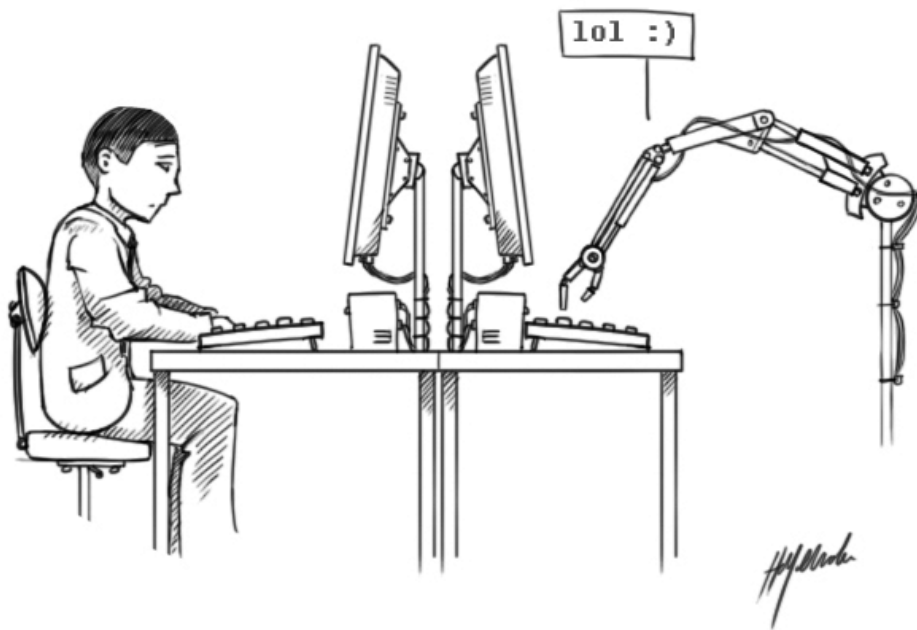
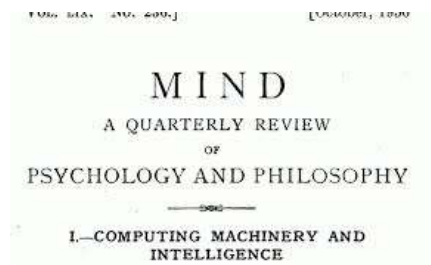
第四届语言与智能高峰论坛

# 知识计算与语言理解

清华大学自然语言处理实验室

刘知远

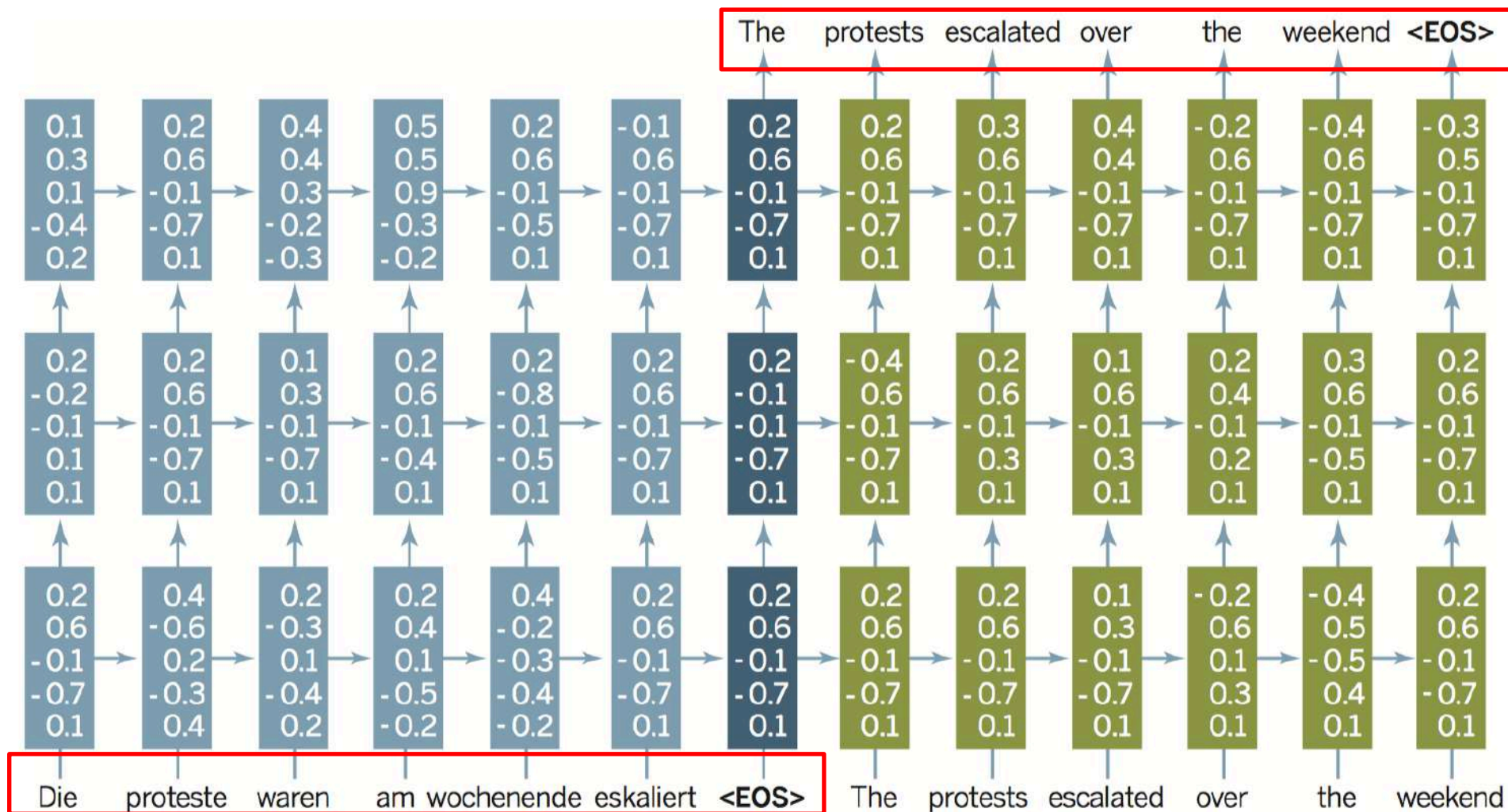
# 自然语言处理是AI关键问题



自然语言处理是实现人工智能、通过图灵测试的关键

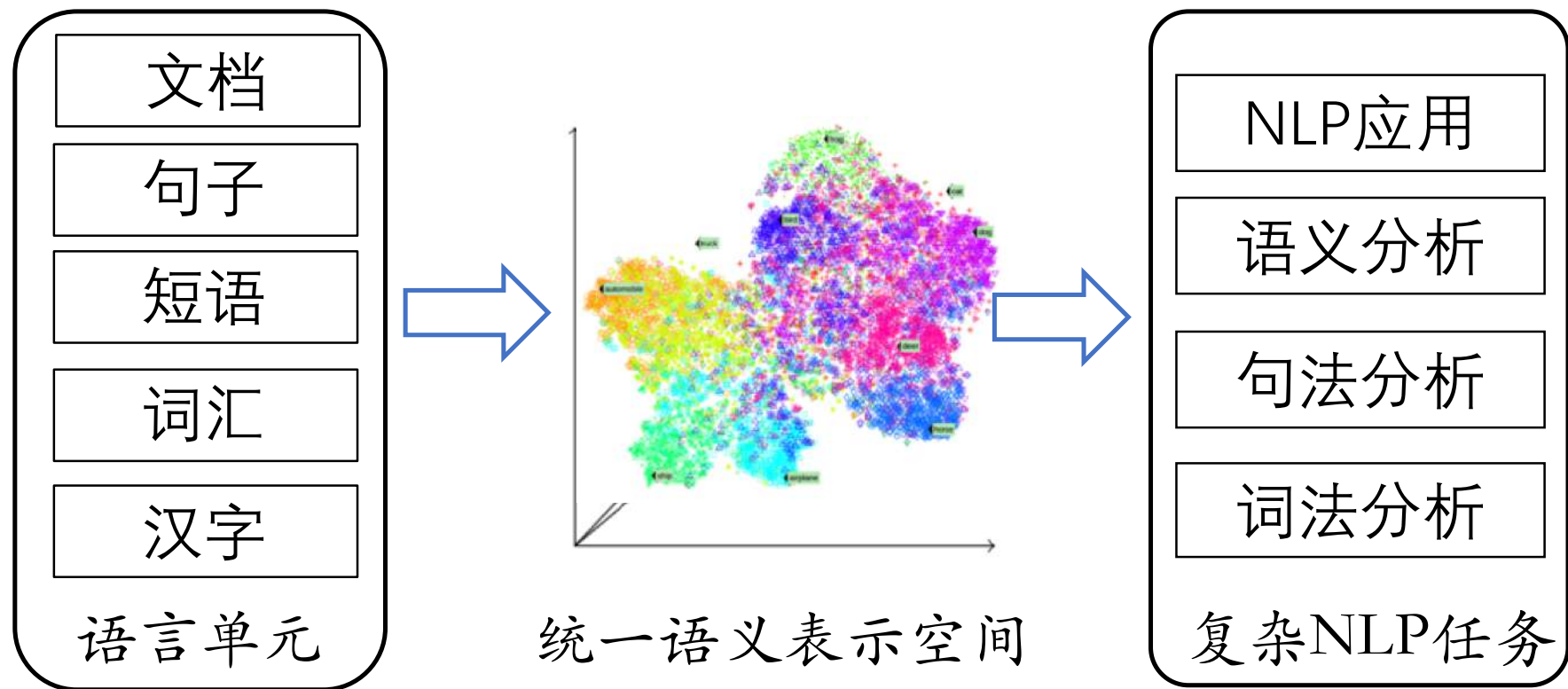
# 数据驱动的自然语言处理：深度学习

- 深度学习技术在自然语言处理取得了巨大突破



# 数据驱动的自然语言处理：深度学习

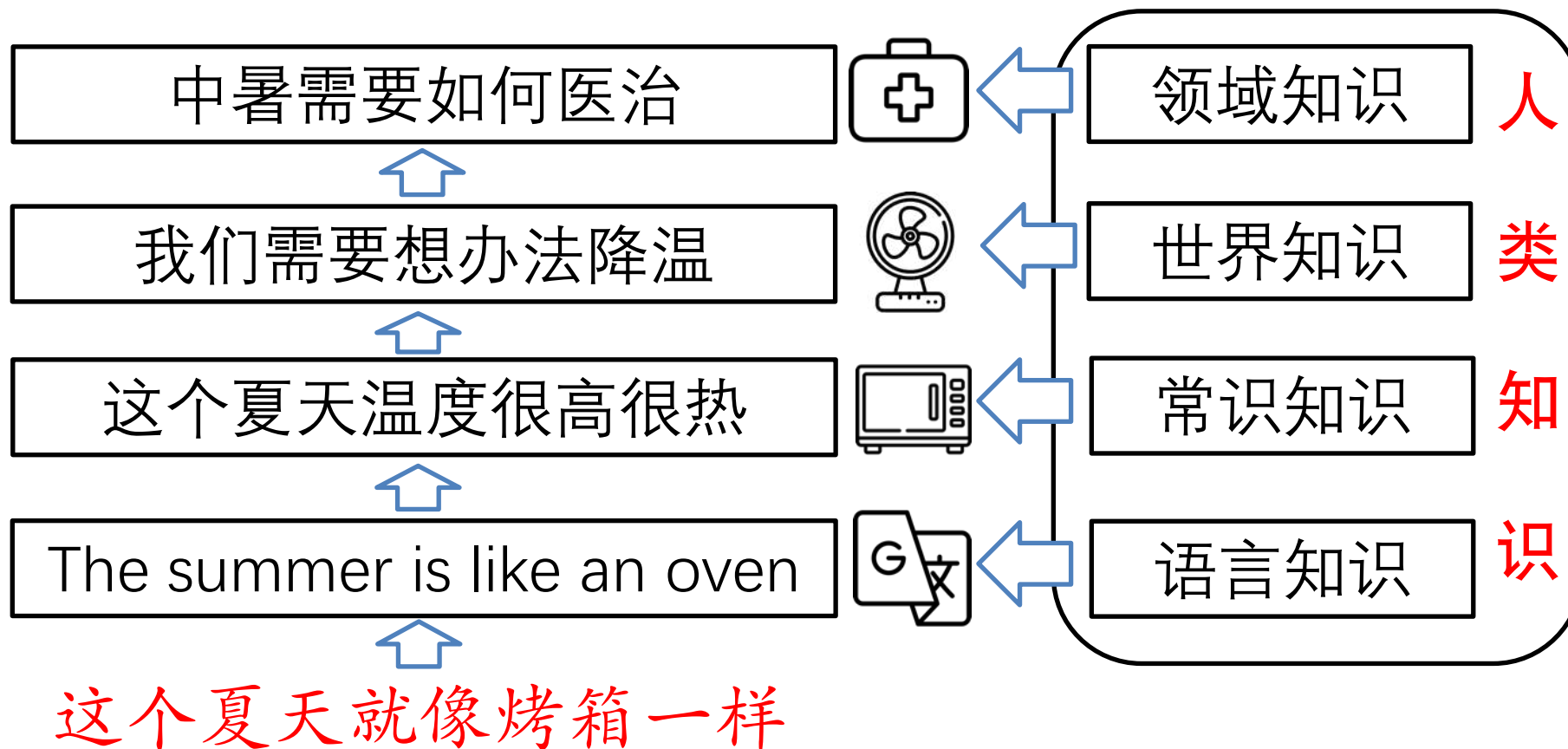
- 深度学习技术在自然语言处理取得了巨大突破



深度学习能够高效学习多粒度语言单元间复杂语义关联

# 面临挑战

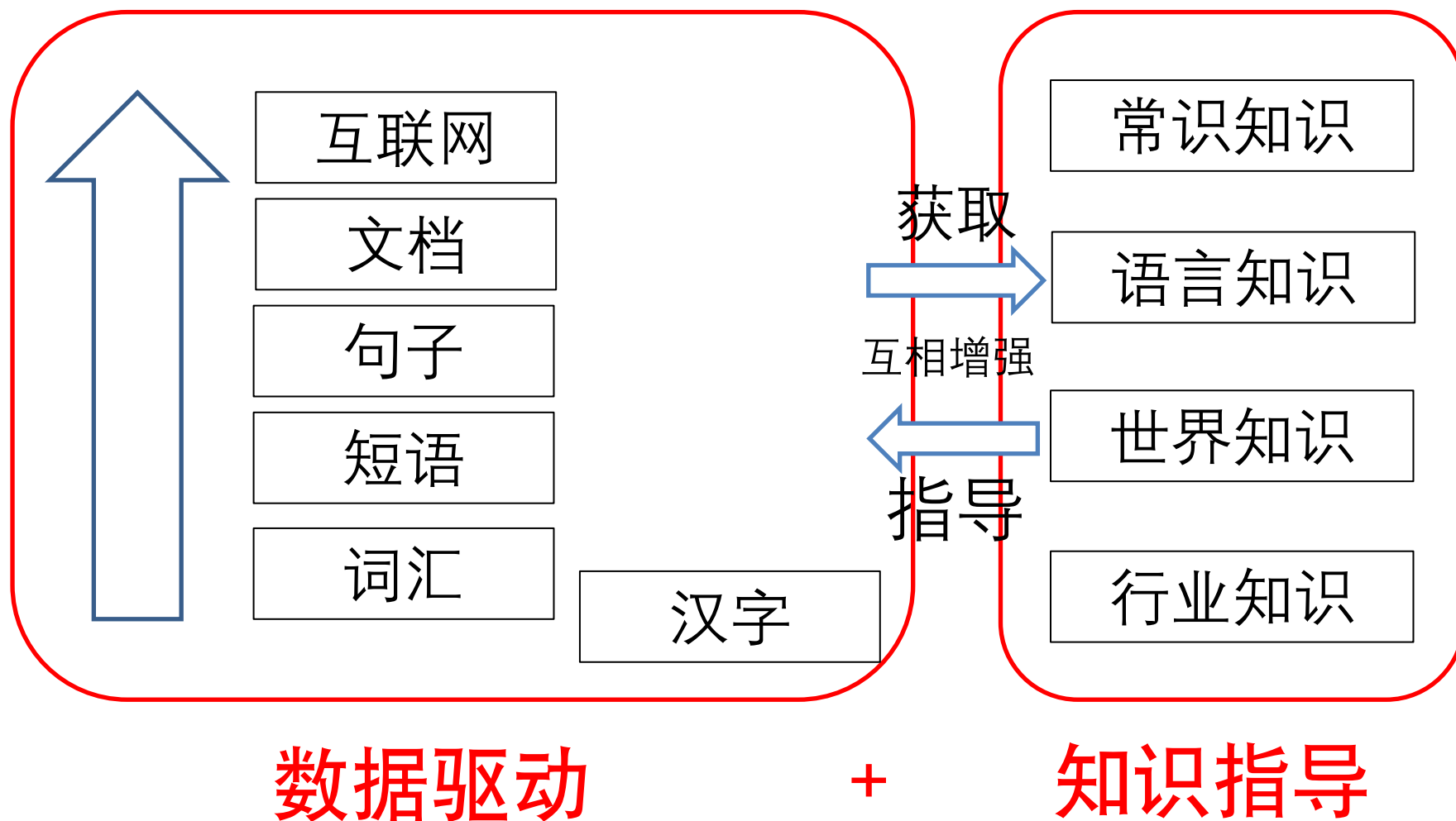
- 对自然语言的深度理解需要复杂知识的支持



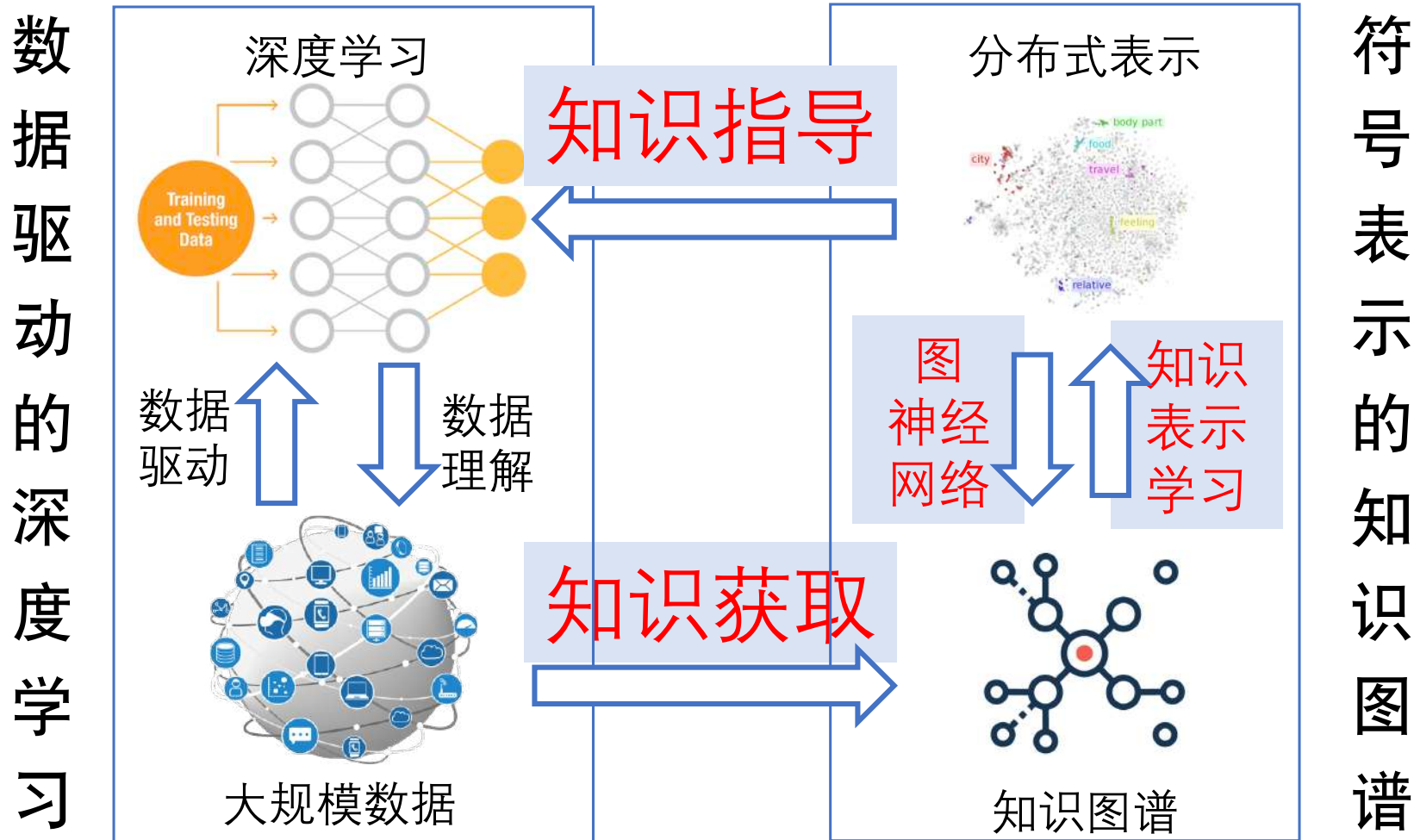
亟需知识支持实现NLP从字面意思到言外之意的跃迁

# 自然语言特点

- 自然语言文本蕴含丰富的语言知识和世界知识



# 研究思路



深度学习+知识图谱 双向驱动的自然语言处理技术体系



# 世界知识库

- 以Google Knowledge Graphs为代表的世界知识库，用三元组形式记录知识



莎士比亚

写作

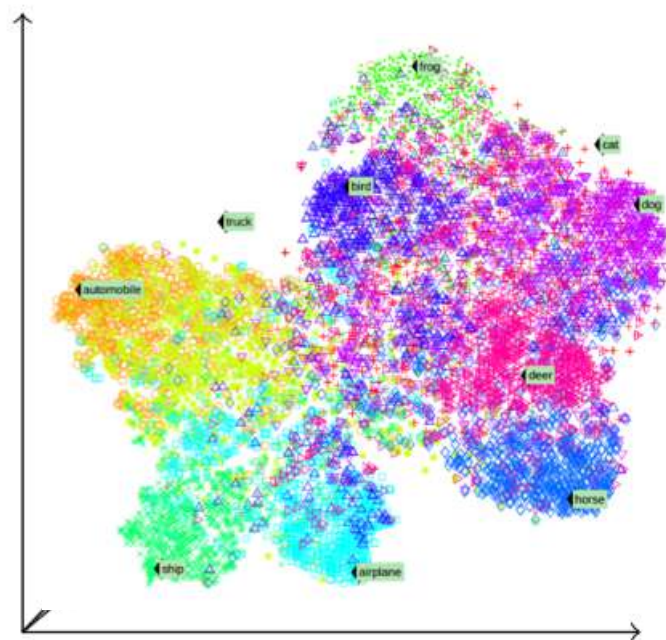
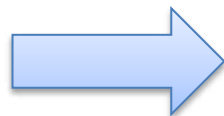
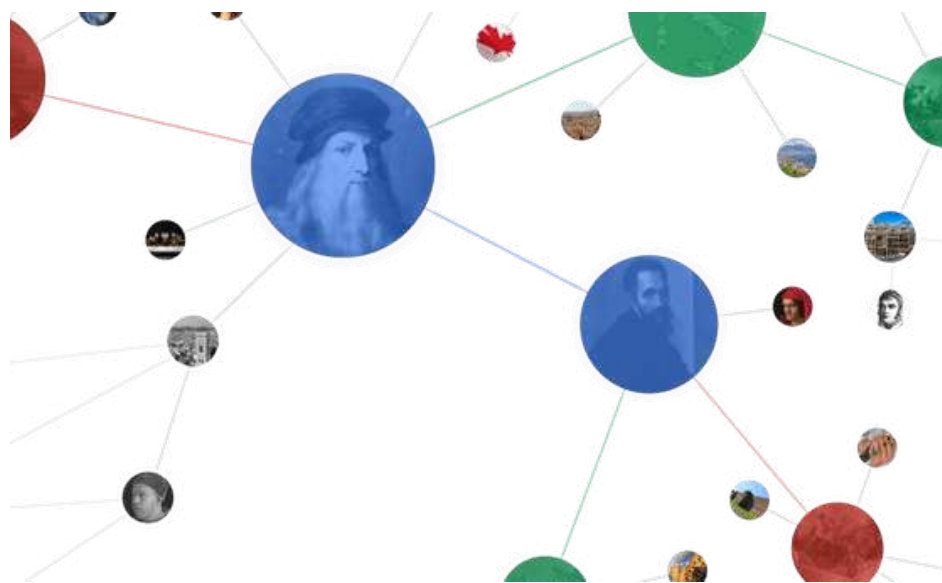


罗密欧与朱丽叶



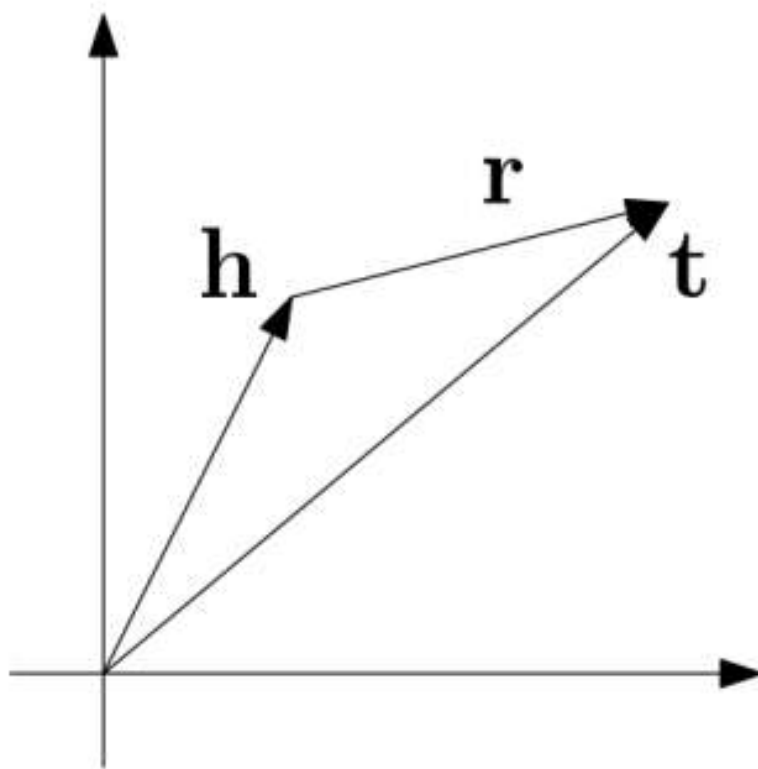
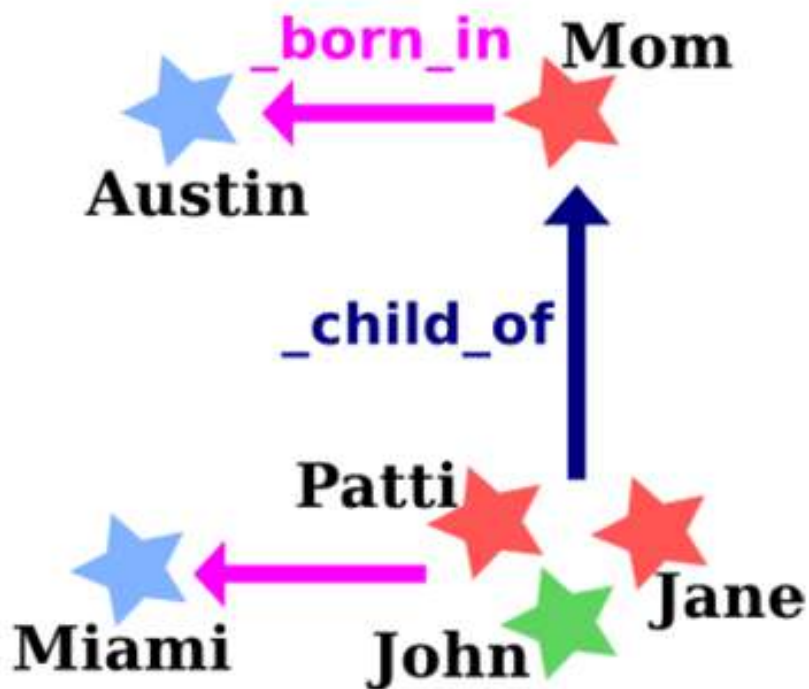
# 知识表示学习

- 基于知识图谱的知识表示学习



# 世界知识的分布式表示学习

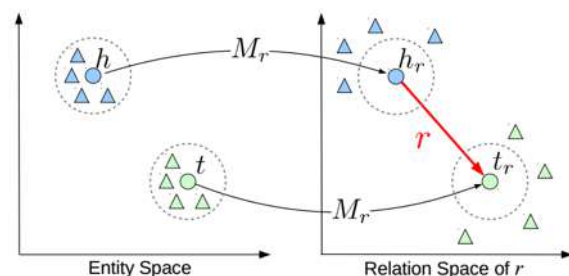
- TransE对每个事实 (head, relation, tail), 将其中的relation作为从head到tail的平移操作



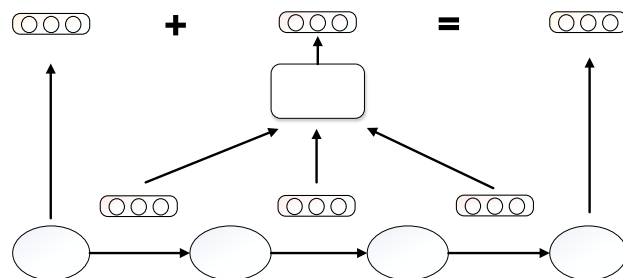
优化目标:  $h + r = t$

# 世界知识的分布式表示学习

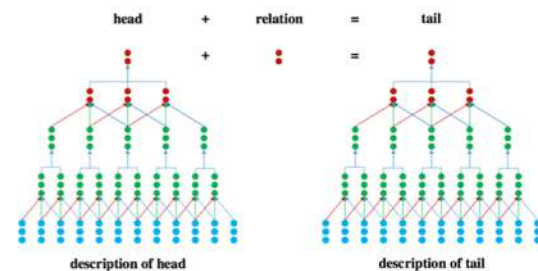
- 利用知识图谱和实体描述、类别和图像等外部信息，实现高效知识表示学习



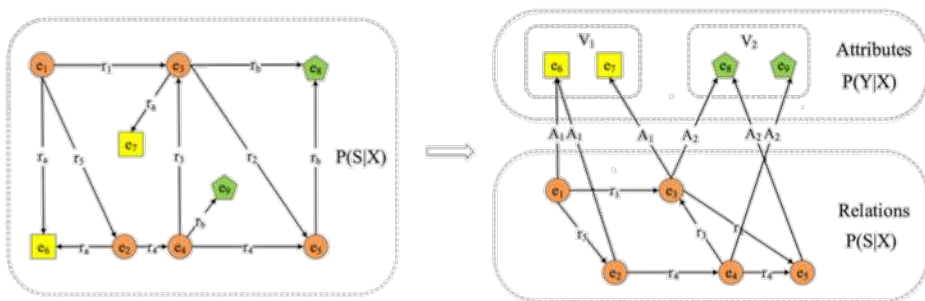
考虑复杂关系类型的知识表示  
TransR (AAAI 2015)



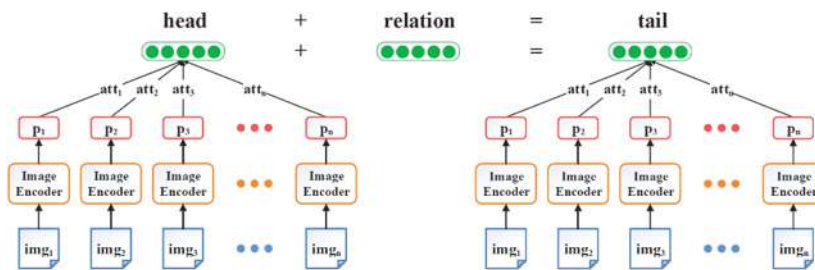
考虑关系路径的知识表示  
PTransE (EMNLP 2015)



考虑实体描述信息的知识表示  
DKRL (AAAI 2016)



综合考虑实体、属性与关系的知识表示  
KR-EAR (IJCAI 2016)



考虑实体图像信息的知识表示  
IKRL (IJCAI 2017)

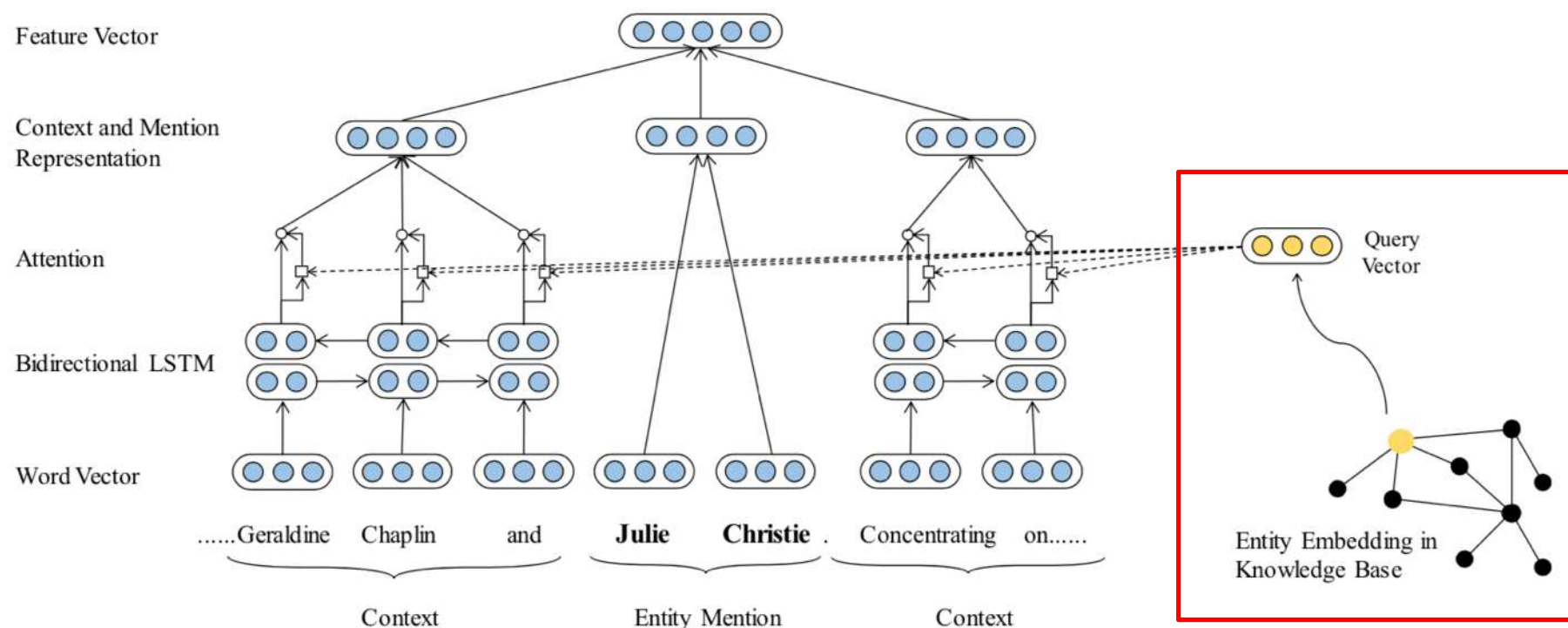
# 知识表示学习论文列表

<https://github.com/thunlp/KRLPapers>

- Xin Lv, Lei Hou, Juanzi Li, Zhiyuan Liu. **Differentiating Concepts and Instances for Knowledge Graph Embedding**. EMNLP 2018.
- Ruobing Xie, Zhiyuan Liu, Fen Lin, Leyu Lin. **Does William Shakespeare REALLY Write Hamlet? Knowledge Representation Learning with Confidence**. AAAI 2018.
- Ruobing Xie, Zhiyuan Liu, Huanbo Luan, Maosong Sun. **Image-embodied Knowledge Representation Learning**. IJCAI 2017.
- Yankai Lin, Zhiyuan Liu, Maosong Sun. **Knowledge Representation Learning with Entities, Attributes and Relations**. IJCAI 2016.
- Ruobing Xie, Zhiyuan Liu, Maosong Sun. **Representation Learning of Knowledge Graphs with Hierarchical Types**. IJCAI 2016.
- Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, Maosong Sun. **Representation Learning of Knowledge Graphs with Entity Descriptions**. AAAI 2016.
- Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, Song Liu. **Modeling Relation Paths for Representation Learning of Knowledge Bases**. EMNLP 2015.
- Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, Xuan Zhu. **Learning Entity and Relation Embeddings for Knowledge Graph Completion**. AAAI 2015.

# 知识指导的实体细粒度分类

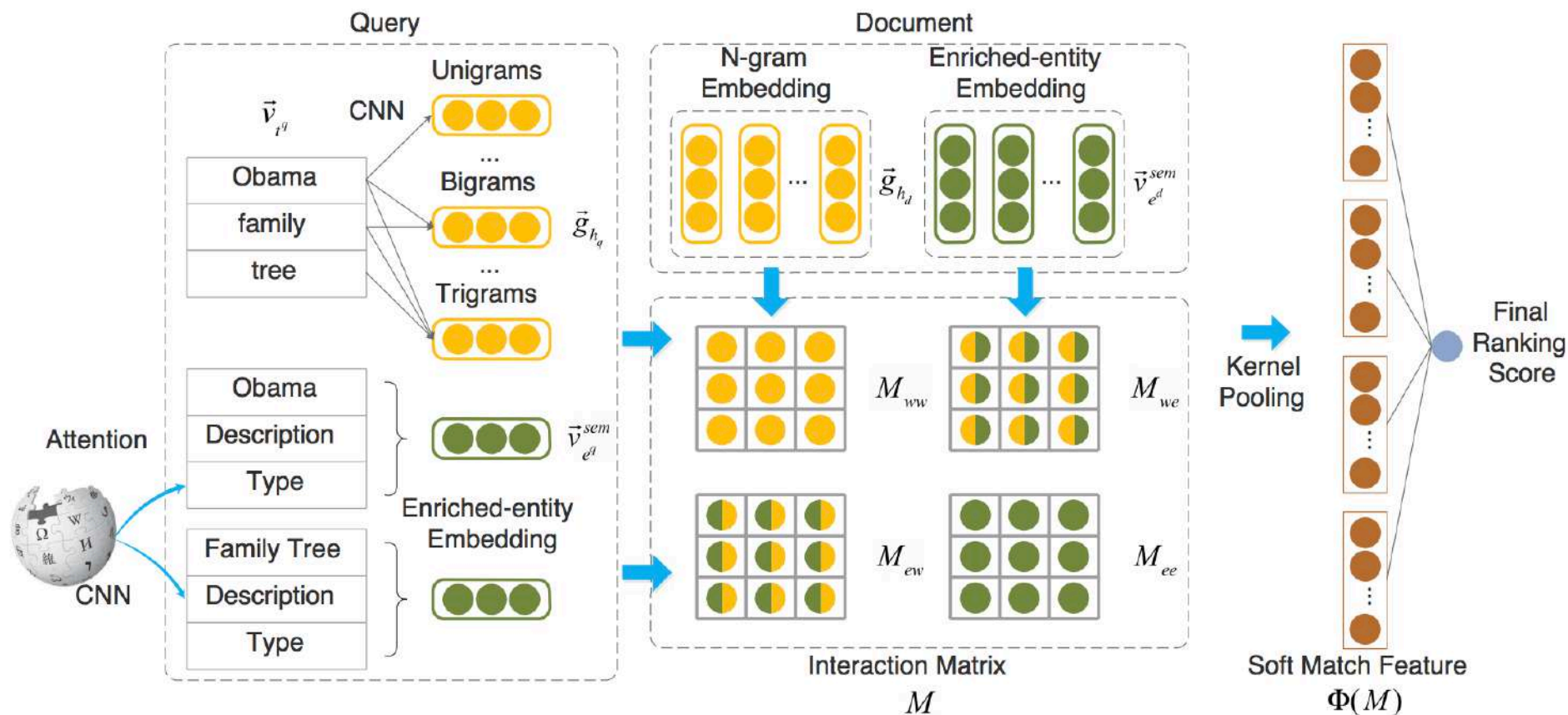
- 对文本实体进行细粒度分类，助力深度分析
- 充分利用KG实体表示，提出知识注意力机制，建立对上下文的高效建模





# 知识指导的神经网络文档排序

- 在利用神经网络学习查询-文档匹配关系模型 (KNRM) 中, 引入 KG 世界知识











# 知识指导的预训练语言模型

- 深度学习对大规模无监督数据建模的最新进展
- ELMo、GPT、BERT、XLNet、...

2001	•	Neural language models
2008	•	Multi-task learning
2013	•	Word embeddings
2013	•	Neural networks for NLP
2014	•	Sequence-to-sequence models
2015	•	Attention
2015	•	Memory-based networks
2018	•	Pretrained language models

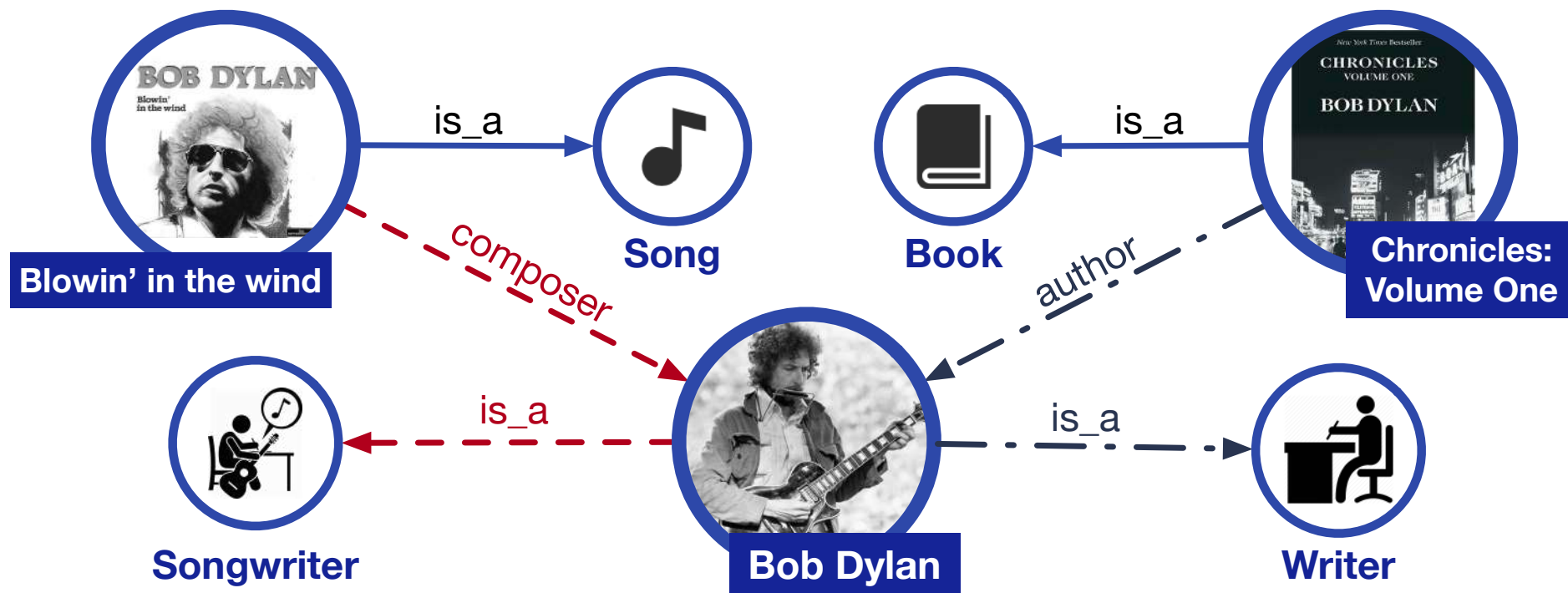
Rank	Name	Model	URL	Score
1	bigbird he	Microsoft D365 AI & MSR AI		81.9
— 2	Jacob Devlin	BERT: 24-layers, 1024-hidden, 16		80.4
		BERT: 12-layers, 768-hidden, 12-l		78.3
3	Jason Phang	GPT on STILTs		76.9
4	Alec Radford	Singletask Pretrain Transformer		72.8
+ 5	Samuel Bowman	BiLSTM+ELMo+Attn		70.5
6	GLUE Baselines	BiLSTM+ELMo+Attn		68.9

Sebastian Ruder <http://ruder.io/a-review-of-the-recent-history-of-nlp/>

Leaderboard of GLUE benchmark (2019.1)

# 知识指导的预训练语言模型

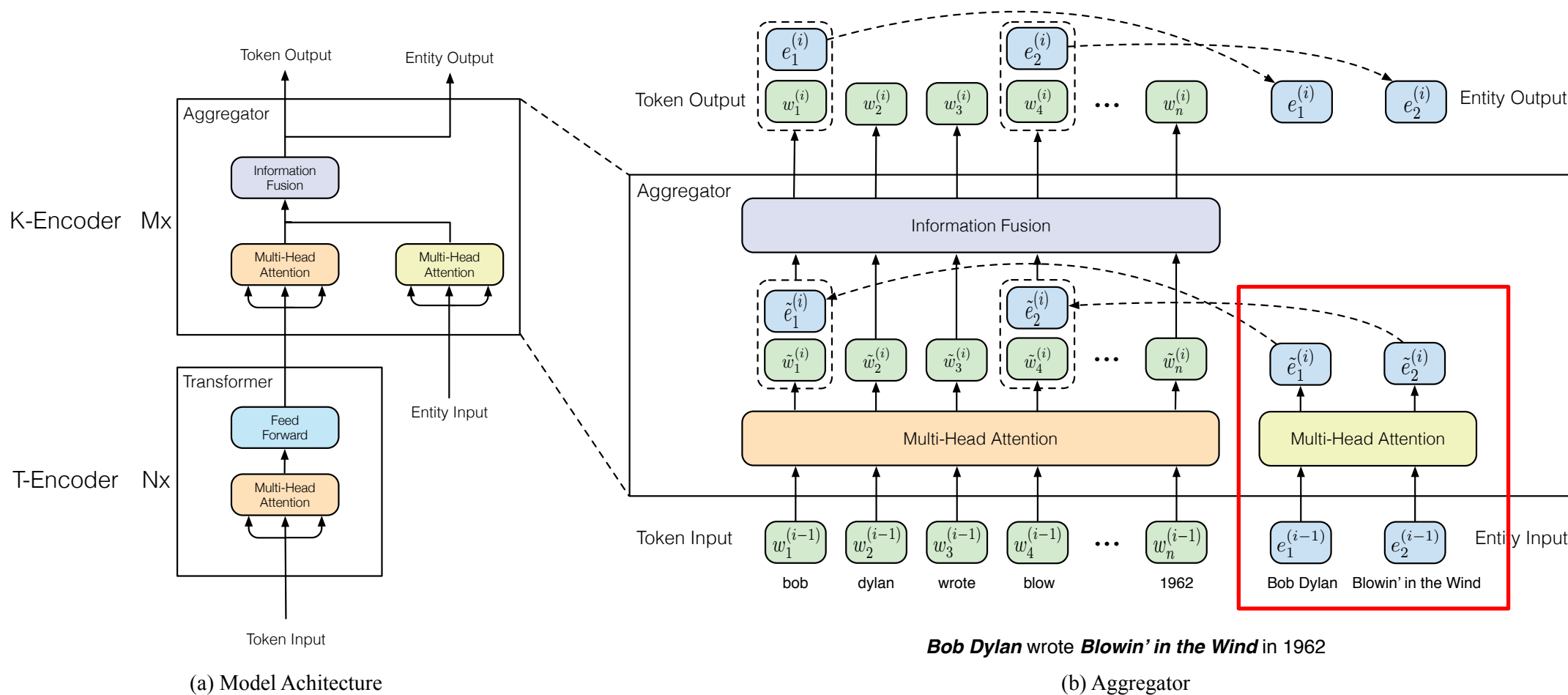
- 预训练模型未考虑知识图谱中的结构化知识
- 结构化知识可以有效提升模型对于文本中的低频实体的理解能力



**Bob Dylan** wrote **Blowin' in the Wind** in 1962, and wrote **Chronicles: Volume One** in 2004.

# 知识指导的预训练语言模型

- 在BERT模型中引入知识图谱表示学习向量，提出实体预测任务



# 世界知识指导NLP论文列表

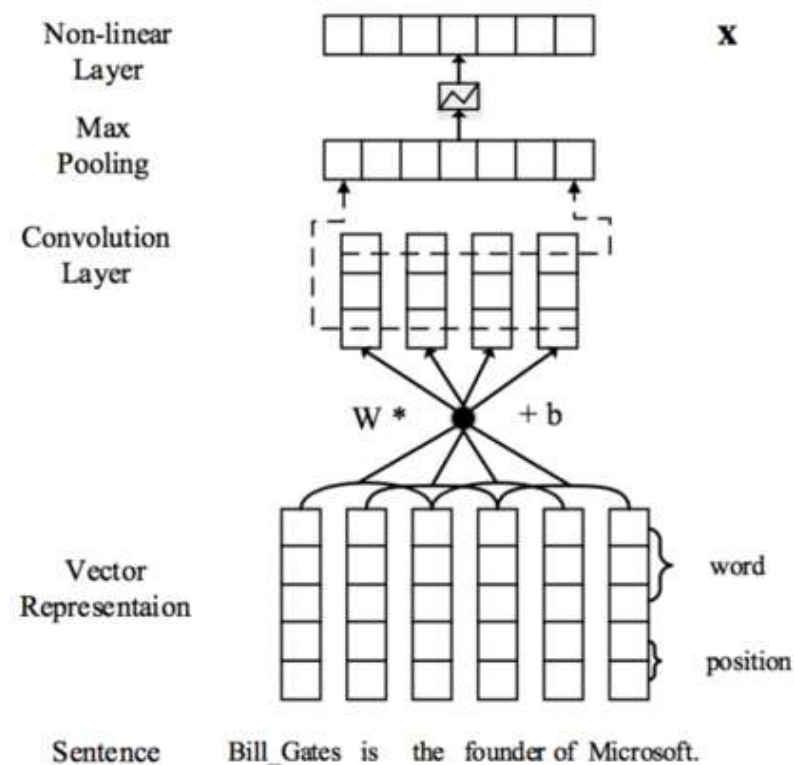
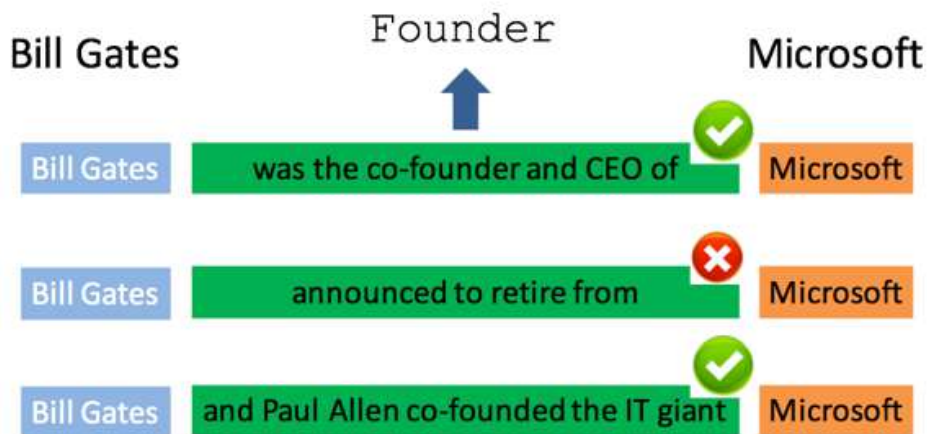
- Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, Qun Liu. **ERNIE: Enhanced Language Representation with Informative Entities**. ACL 2019.
- Zhenghao Liu, Chenyan Xiong, Maosong Sun, Zhiyuan Liu. **Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval**. ACL 2018.
- Ji Xin, Yankai Lin, Zhiyuan Liu, Maosong Sun. **Improving Neural Fine-Grained Entity Typing with Knowledge Attention**. AAAI 2018.
- Hao Zhu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. **Iterative Entity Alignment via Joint Knowledge Embeddings**. IJCAI 2017.
- Yankai Lin, Zhiyuan Liu, Maosong Sun. **Knowledge Representation Learning with Entities, Attributes and Relations**. IJCAI 2016.

# 神经网络知识获取技术

- 采用神经网络对句子进行语义理解
- 使用大规模自动标注训练数据学习

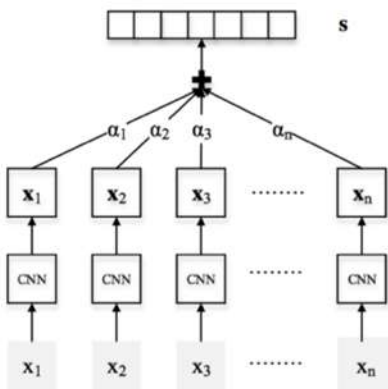
Freebase

(Bill Gates,  
Founder,  
Microsoft)

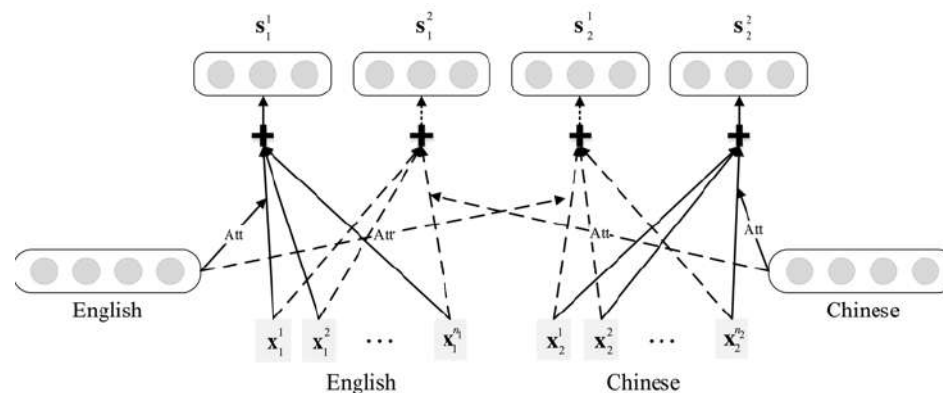


# 高效鲁棒的知识获取技术

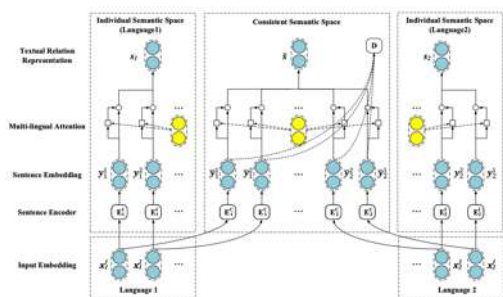
- 提出**选择注意力**机制自动降噪并整合多源信息



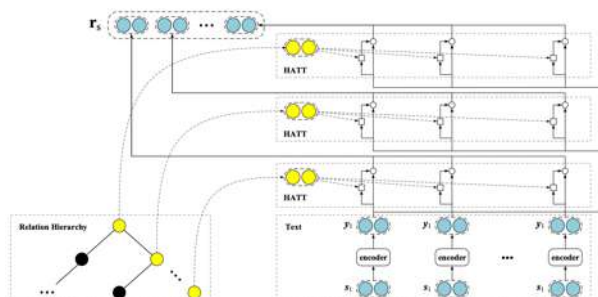
基于**句级注意力**的远程监督  
神经网络关系抽取(ACL 2016)



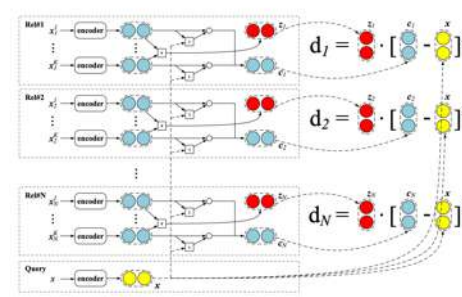
基于**跨语言注意力**的  
神经网络关系抽取(ACL 2017)



基于**对抗注意力**的神经网络  
关系抽取(COLING 2018)



基于**层次注意力**的神经网络  
关系抽取(EMNLP 2018)



基于**混合注意力**的  
少次关系抽取(AAAI 2019)



# 知识获取论文列表

<https://github.com/thunlp/NREPapers>

- Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun, Peng Li. **Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention**. EMNLP 2018.
- Xiaozhi Wang, Xu Han, Yankai Lin, Zhiyuan Liu, Maosong Sun. **Adversarial Multi-lingual Neural Relation Extraction**. COLING 2018.
- Xu Han, Zhiyuan Liu, Maosong Sun. **Neural Knowledge Acquisition via Mutual Attention between Knowledge Graph and Text**. AAAI 2018.
- Wenyuan Zeng, Yankai Lin, Zhiyuan Liu, Maosong Sun. **Incorporating Relation Paths in Neural Relation Extraction**. EMNLP 2017.
- Yankai Lin, Zhiyuan Liu, Maosong Sun. **Neural Relation Extraction with Multi-lingual Attention**. ACL 2017.
- Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, Maosong Sun. **Neural Relation Extraction with Selective Attention over Instances**. ACL 2016.

# 开源工具

- 义原计算、知识表示、知识获取等相关算法工具均在全球最大开源社区GitHub发布，获得超过20, 000+星标关注

<https://github.com/thunlp>

- THULAC : 中文词法分析
- THUCTC : 中文文本分类
- THUTAG : 关键词抽取与社会标签推荐
- OpenKE : 知识表示学习
- OpenNRE : 神经网络关系抽取
- OpenNE : 网络表示学习
- OpenQA : 开放域自动问答

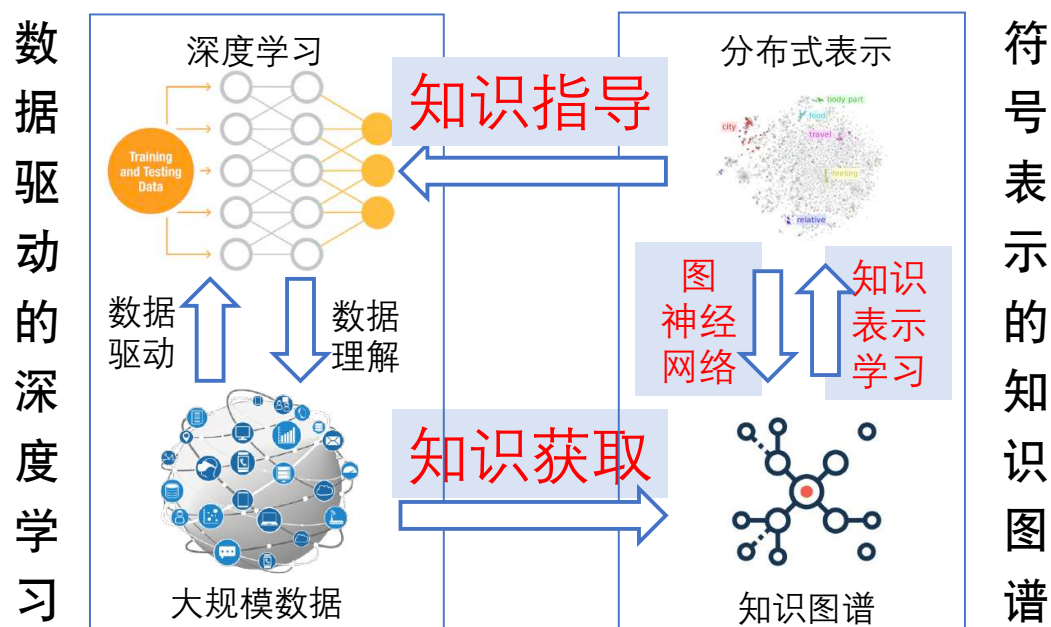


The screenshot displays the GitHub profile for 'thunlp'. It includes the repository logo (Tsinghua University NLP), the profile name 'thunlp', location 'Beijing', and a 'Refresh your profile' button. To the right, there are two ranking tables: 'c++ ranking' and 'python ranking'. The 'c++ ranking' table shows 12 / 2 413 stars for Beijing, 30 / 9 212 stars for China, and 519 / 251 037 stars Worldwide, with 11 repositories and 822 stars. The 'python ranking' table shows 33 / 3 336 stars for Beijing, 91 / 12 113 stars for China, and 2 045 / 419 419 stars Worldwide, with 6 repositories and 529 stars.

Ranking	Location	Stars
c++ ranking	Beijing	12 / 2 413
	China	30 / 9 212
	Worldwide	519 / 251 037
python ranking	Beijing	33 / 3 336
	China	91 / 12 113
	Worldwide	2 045 / 419 419

# 总结展望

- 知识对于富知识文本深度理解具有重要意义，知识表示学习是目前较好的解决方案
- 深度学习自然语言处理技术反过来可以帮助从大规模文本中获取知识



## 五个更加

1. 更加全面的知识类型
2. 更加复杂的知识结构
3. 更加有效的知识获取
4. 更加强大的知识指导
5. 更加精深的知识推理

# 感谢各位!

<http://nlp.csai.tsinghua.edu.cn/~lzy/>  
[liuzy@Tsinghua.edu.cn](mailto:liuzy@Tsinghua.edu.cn)