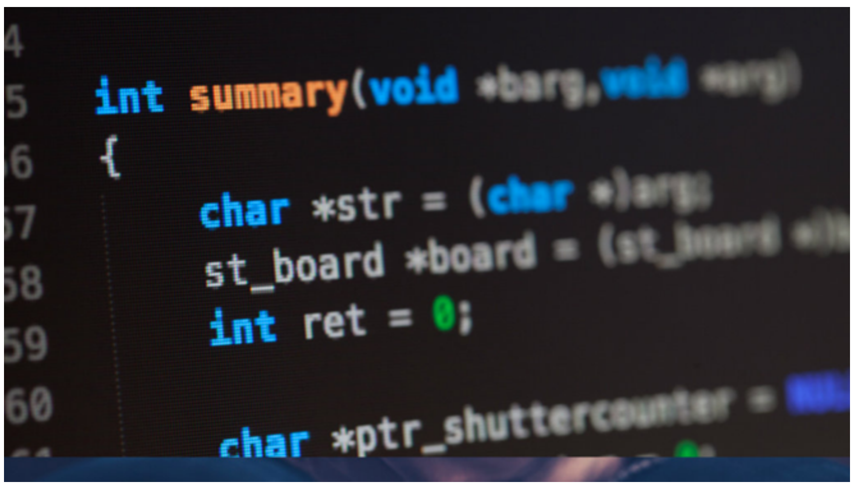
Knowledge-Guided NLP

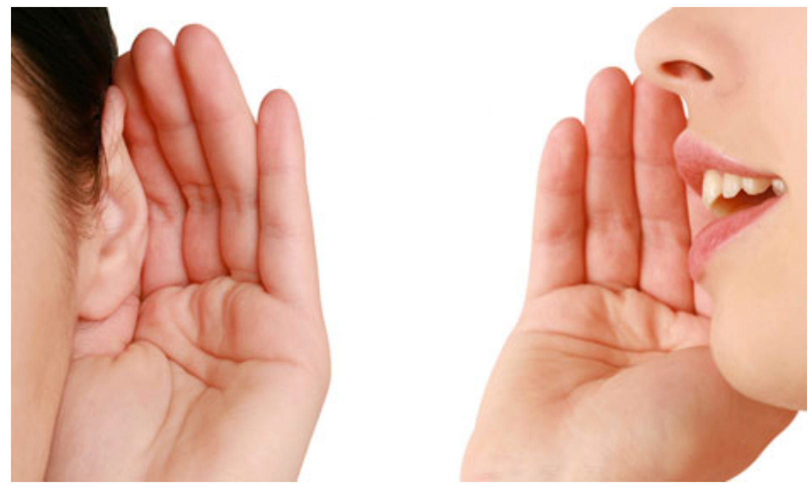
Zhiyuan Liu

Tsinghua NLP

Natural Language

- Natural language is different from programming languages
    - – new words/usages; ambiguity; > CFG



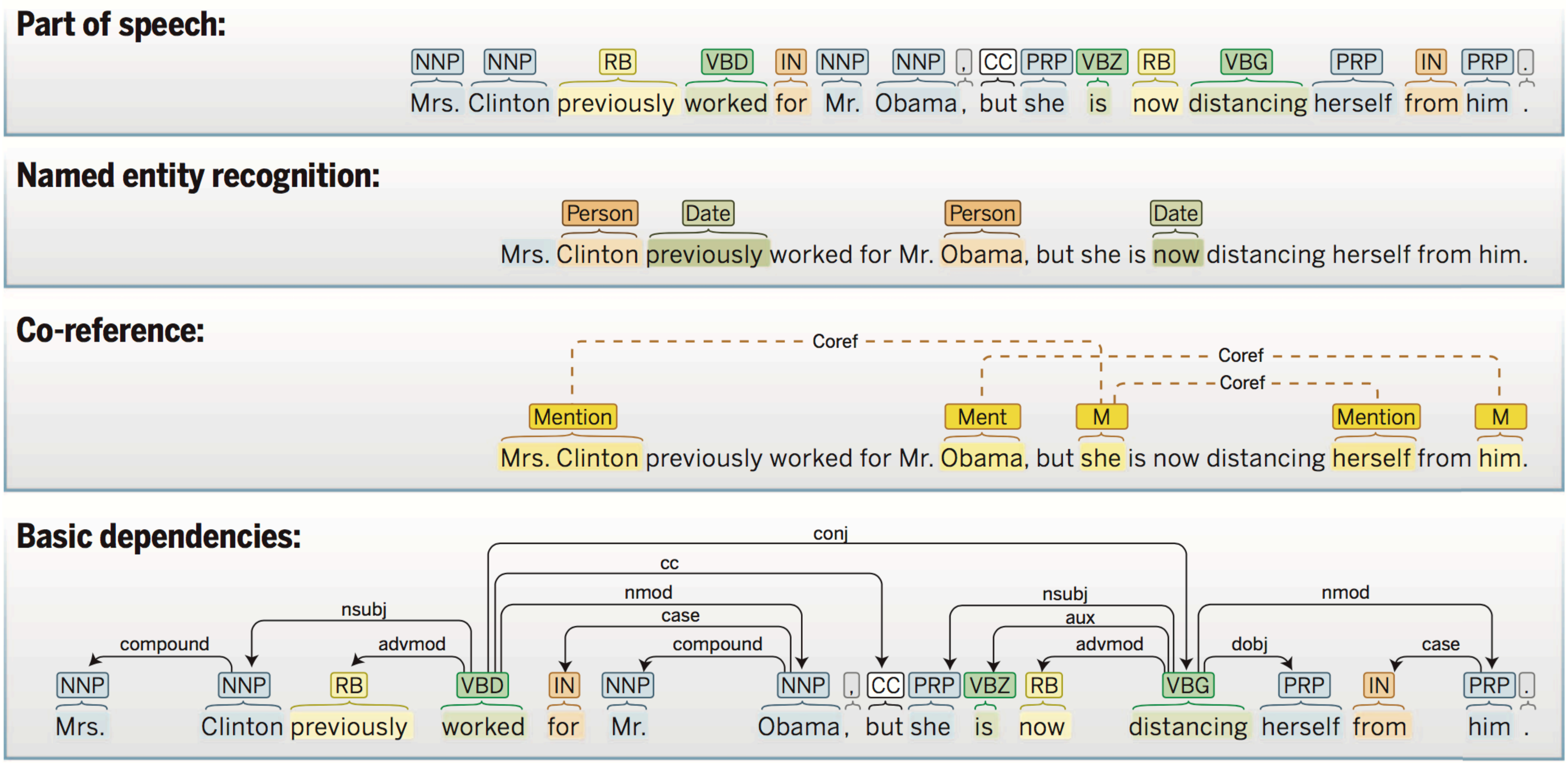


Programming Languages

Natural Languages

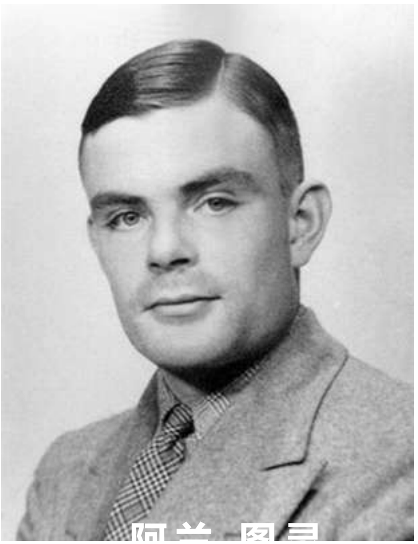
Natural Language Processing

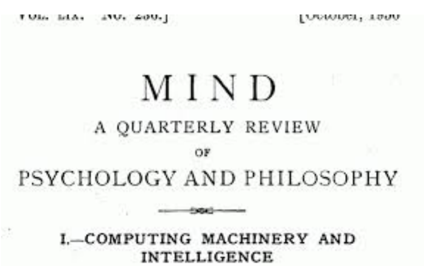
- NLP aims to make computers understand languages
- The nature of NLP is structure prediction



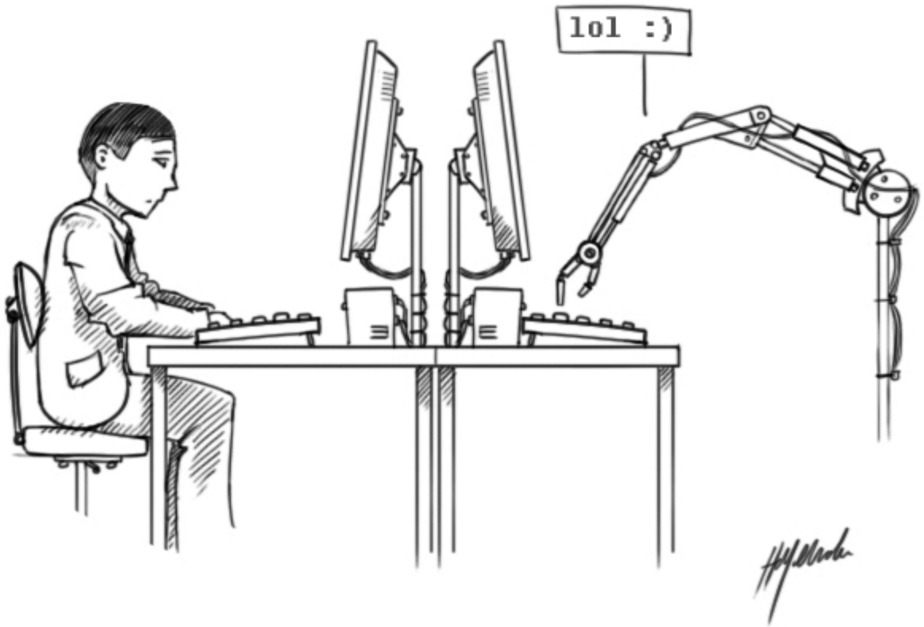
Advances in Natural Language Processing. Science 2015.

NLP Is The Key of AI



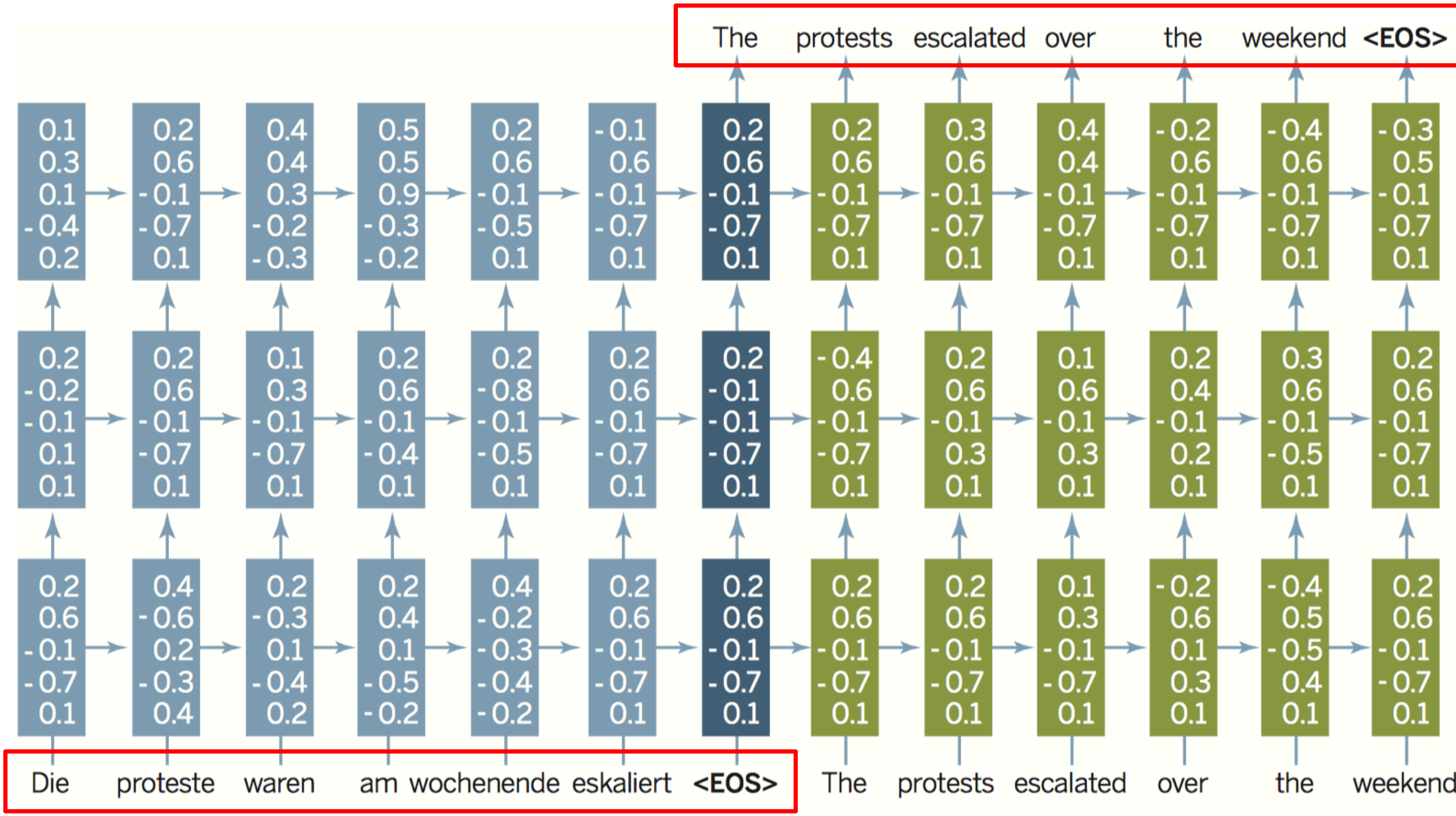

MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE



NLP: The Key to Pass Turing Test and Realize AI

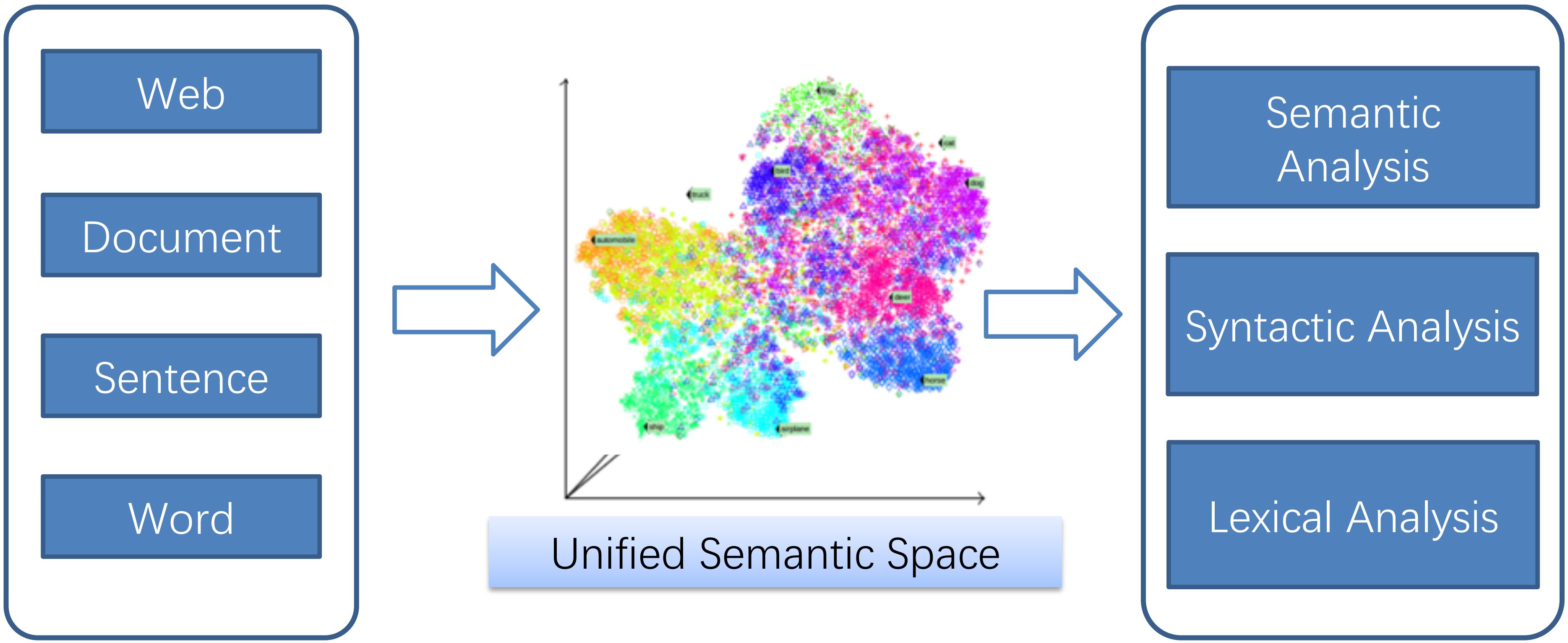
Deep Learning: Data-Driven NLP



Advances in Natural Language Processing. Science 2015.

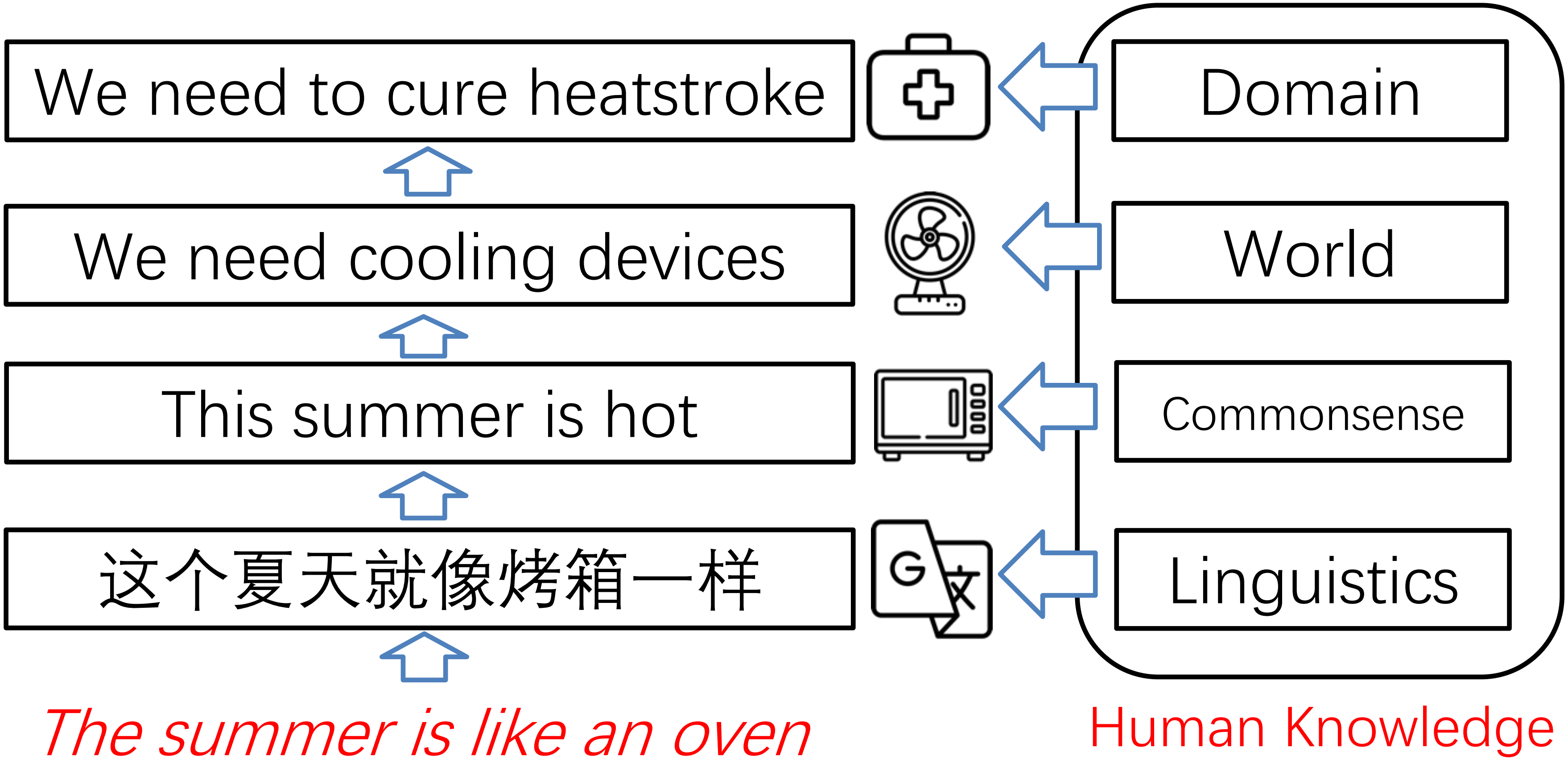
Distributed Representation

- Bridge the multiple-grained units in languages
- Alleviate the issue of data sparsity



What’s Next

- Deep Language Understanding Requires Knowledge



From surface meanings to implicative meanings

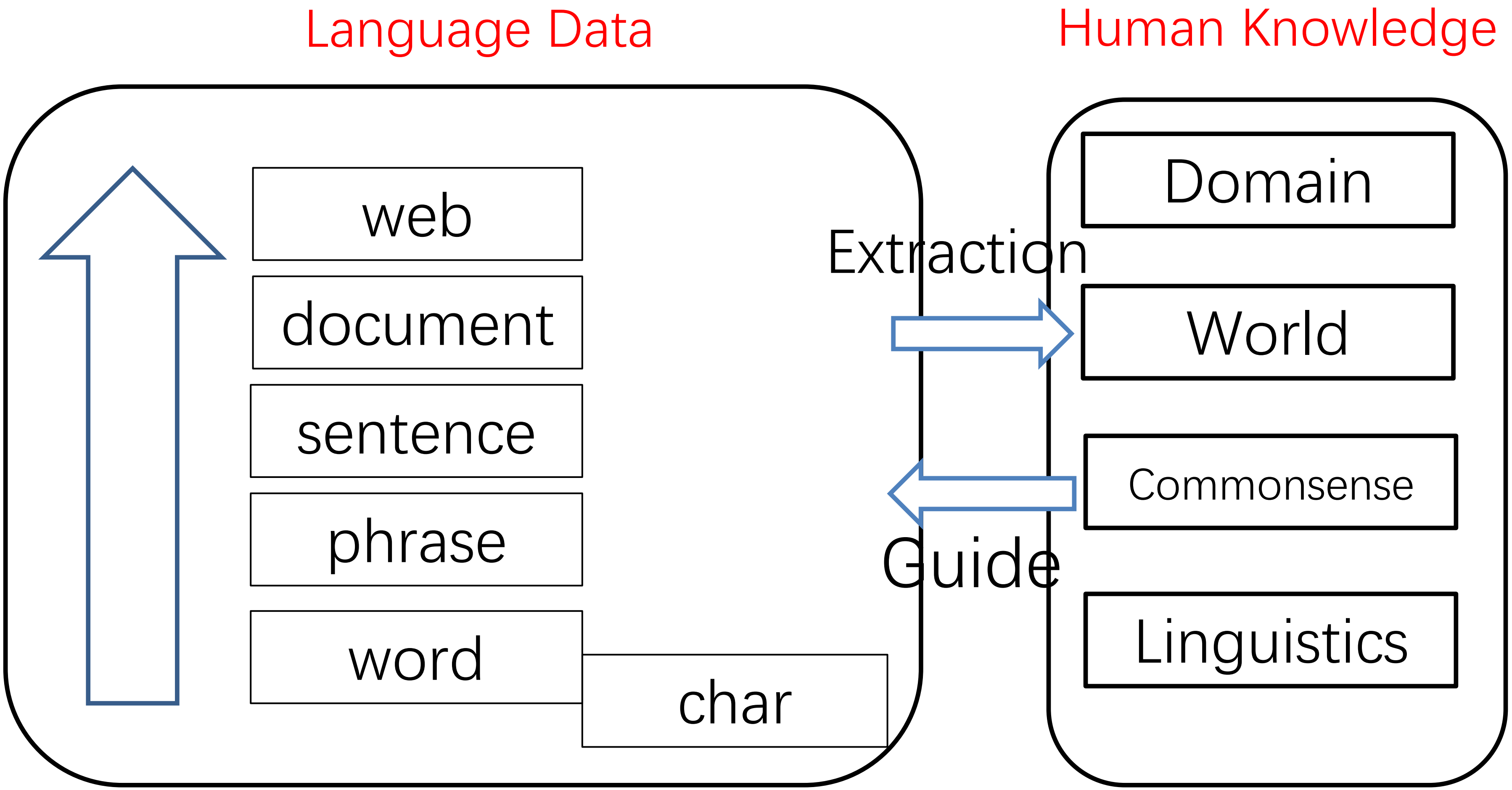
What’s Next

… we feel confident that more data and computation, in addition to recent advances in ML and deep learning, will lead to further substantial progress in NLP. However, the truly <u>difficult problems of semantics, context, and knowledge</u> will probably require new discoveries in linguistics and inference.

Advances in Natural Language Processing. Science 2015.

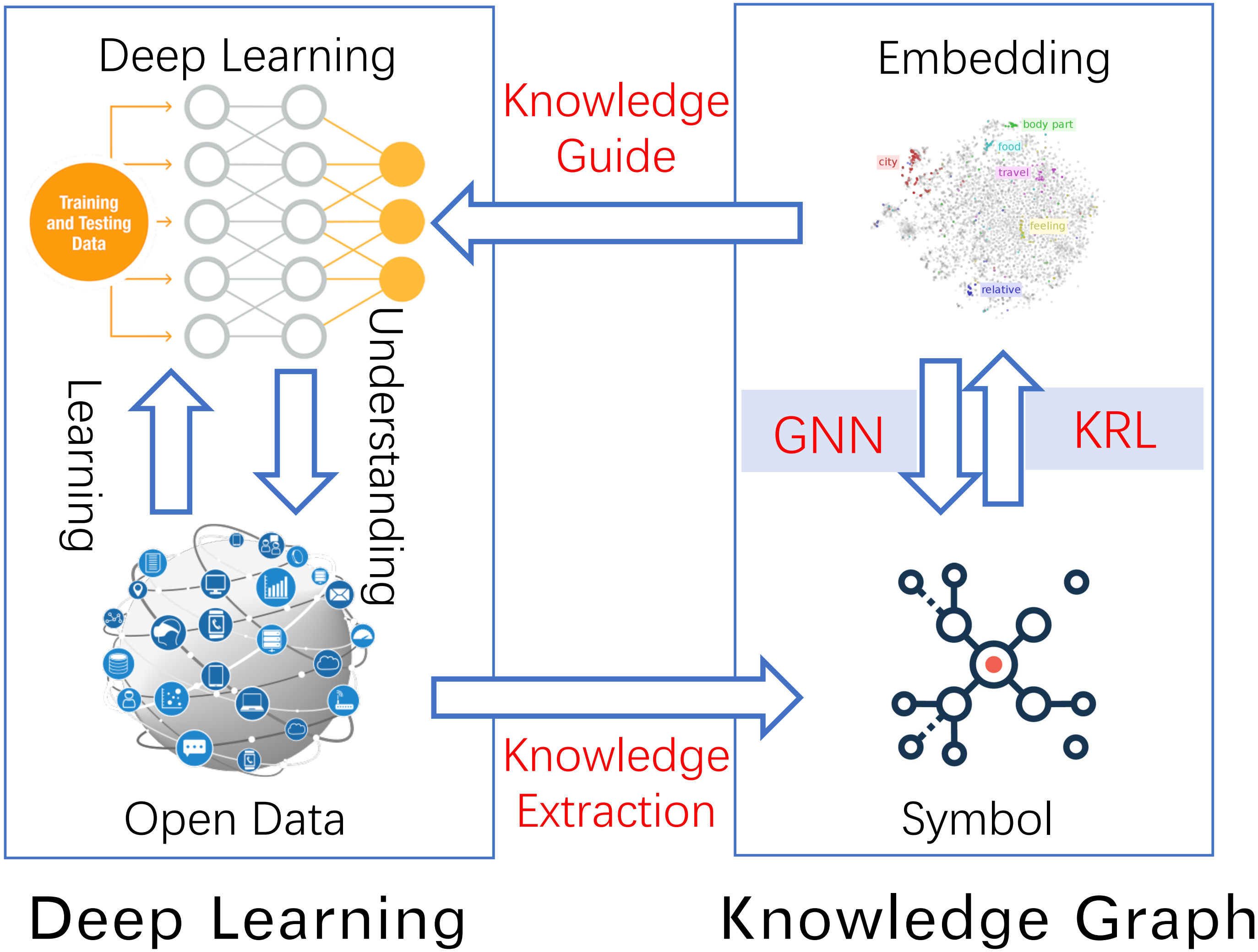
Characteristics of Natural Language

- There are rich knowledge in text



Data Driven + Knowledge Guide

Approach



DL+KG = Knowledge-Guided NLP

PART ONE

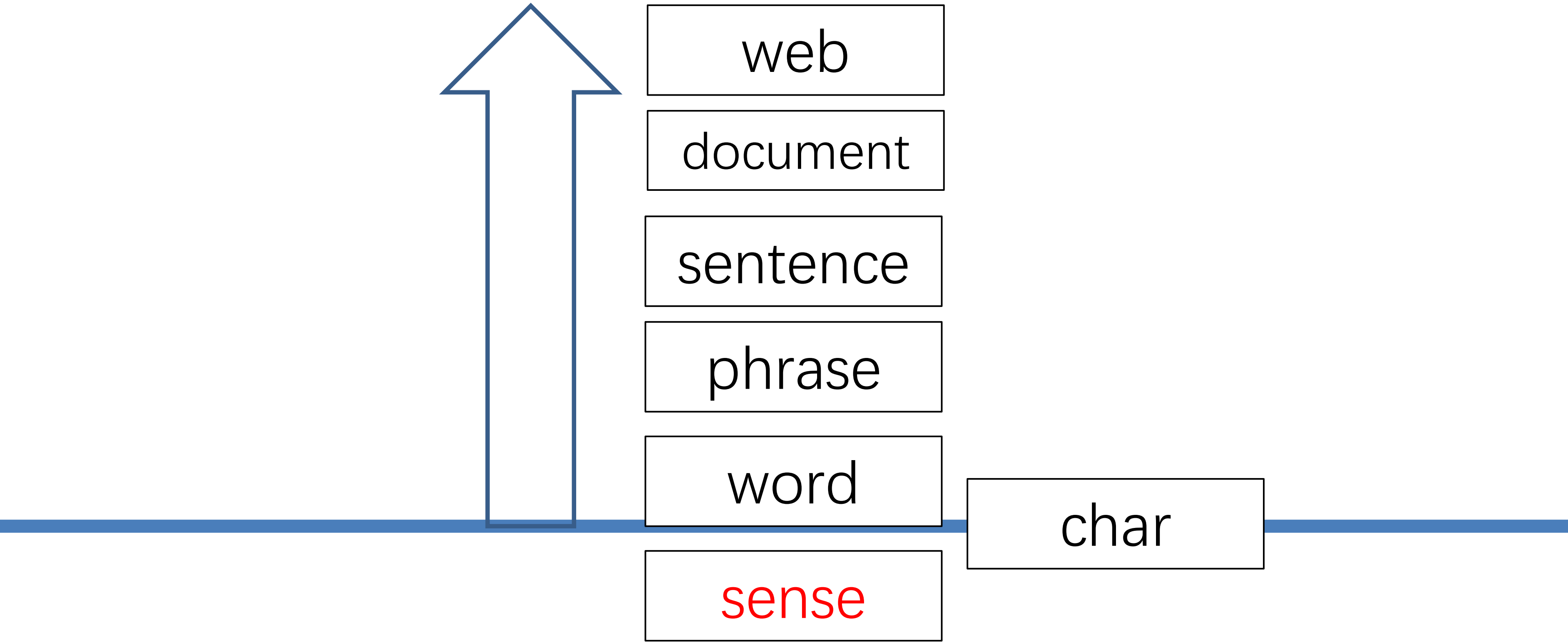
Linguistic / Commonsense Knowledge

Zhiyuan Liu

Tsinghua NLP

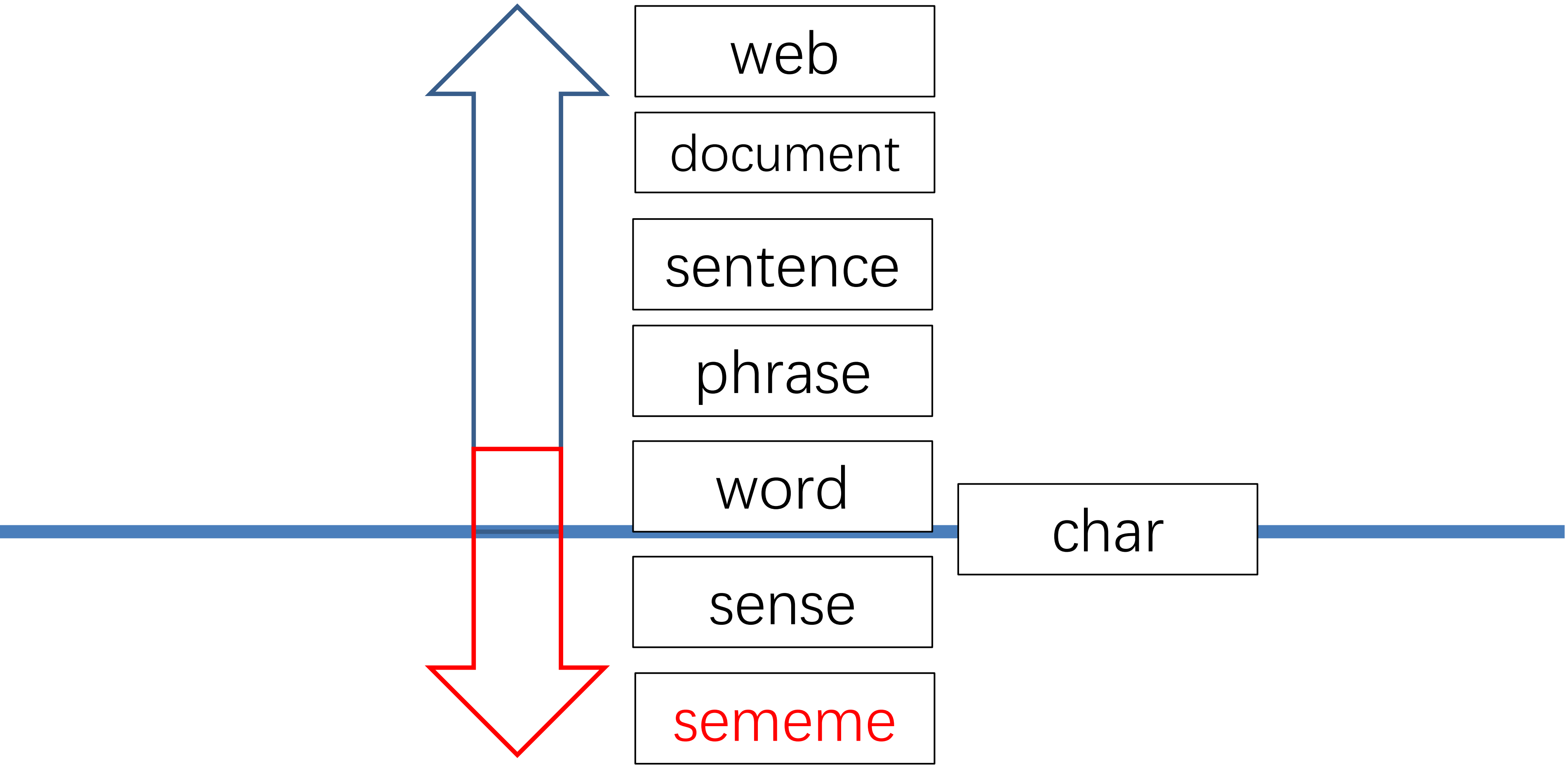
Characteristics of Natural Language

- There are multiple-grained units in languages
- Words/Chinese characters are minimal units of usages, but not minimal units of semantics



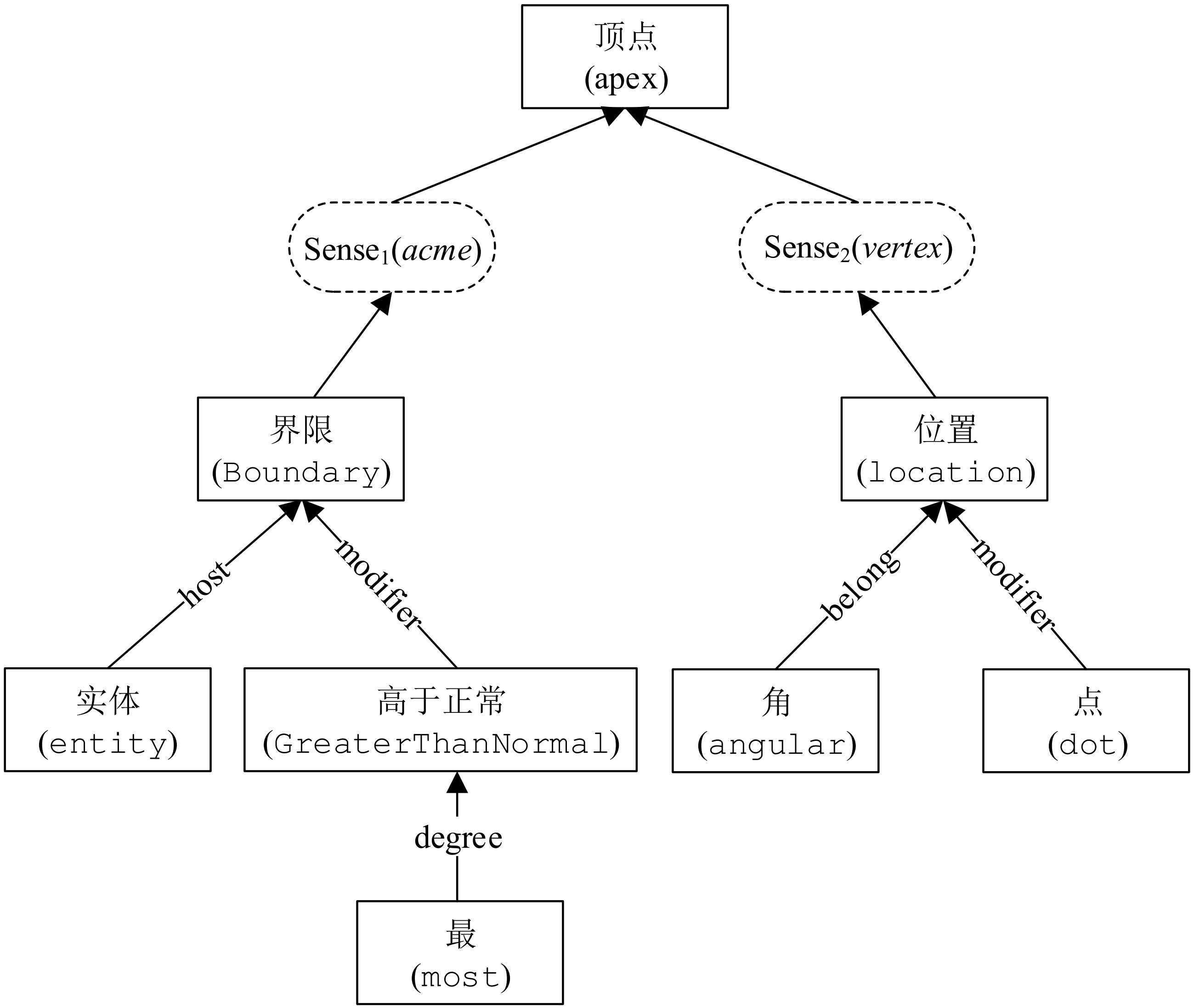
Use Sememes to Break Word Boundary

- Lexical sememes: minimal units of semantics



Linguistic Knowledge with Lexical Sememes

- Lexical sememes: minimal units of semantics



HowNet

- Linguistic knowledge base of lexical sememes, released in 1999
- Manually create ~2,000 sememes
- Manually annotate ~100,000 words with sememes

基于《知网》的词汇语义相似度计算[1]

Word Similarity Computing Based on How-net

刘群[*] 、李素建[+]

Qun LIU , Sujian LI

摘要

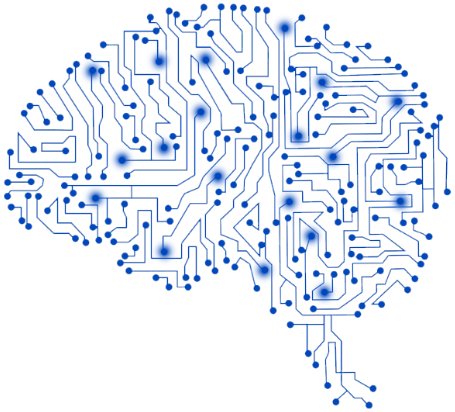
词义相似度计算在很多领域中都有广泛的应用，例如信息检索、信息抽取、文本分类、词义排歧、基于实例的机器翻译等等。词义相似度计算的两种基本方法是基于世界知识（Ontology）或某种分类体系（Taxonomy）的方法和基于统计的上下文向量空间模型方法。这两种方法各有优缺点。

《知网》是一部比较详尽的语义知识词典，受到了人们普遍的重视。不过，由于《知网》中对于一个词的语义采用的是一种多维的知识表示形式，这给词语相似度的计算带来了麻烦。这一点与 WordNet 和《同义词词林》不同。在 WordNet 和《同义词词林》中，所有同类的语义项（WordNet 的 synset 或《同义词词林》的词群）构成一个树状结构，要计算语义项之间的距离，只要计算树状结构中相应结点的距离即可。而在《知网》中词汇语义相似度的计算存在以下问题：

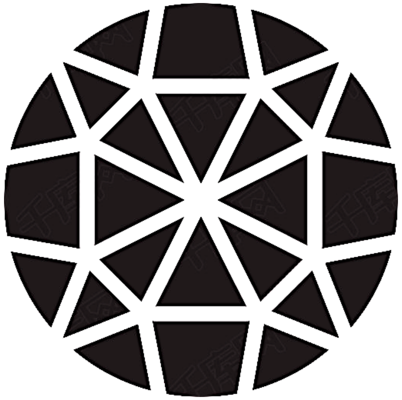
1. 每一个词的语义描述由多个义原组成；
2. 词语的语义描述中各个义原并不是平等的，它们之间有着复杂的关系，通过一种专门的知识描述语言来表示。

我们的工作主要包括：

1. 研究《知网》中知识描述语言的语法，了解其描述一个词义所用的多个义原之间的关系，区分其在词语相似度计算中所起的作用；我们采用一种更



Guide

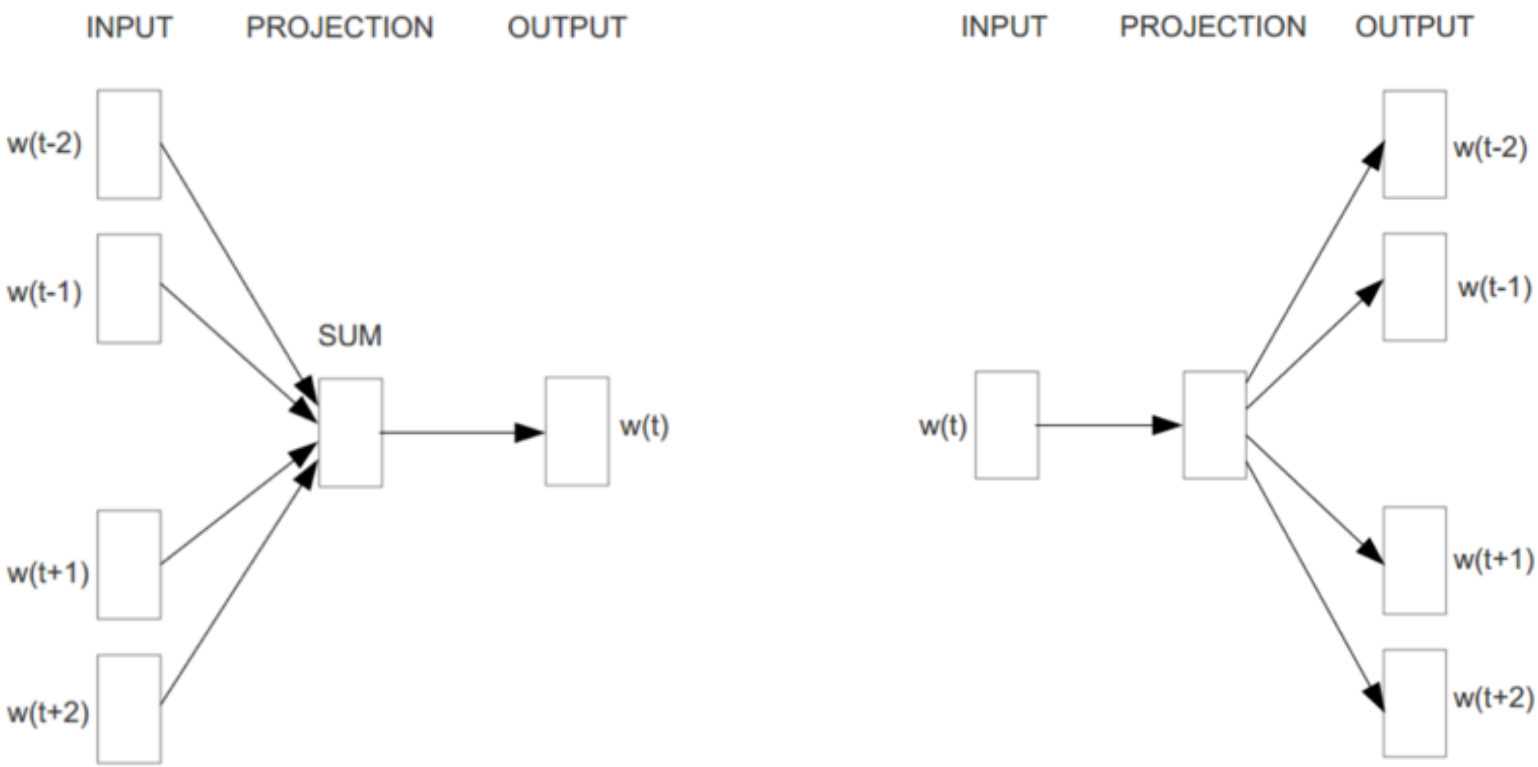


Data-Driven
DL

Symbol-based
Sememe Knowledge

Word Embedding with Sememes

- Learn low-dimensional semantic representations for words

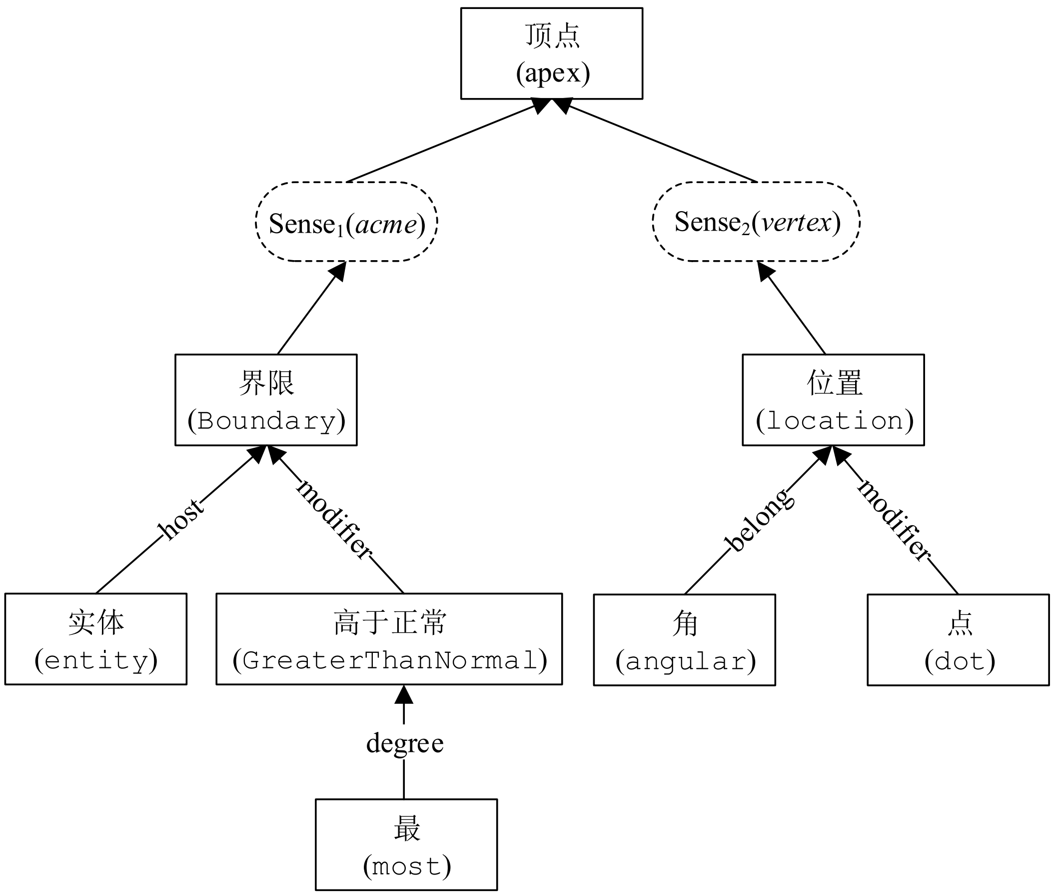


word2vec

Tomas Mikolov et al. Distributed representations of words and phrases and their compositionality. NIPS 2013.

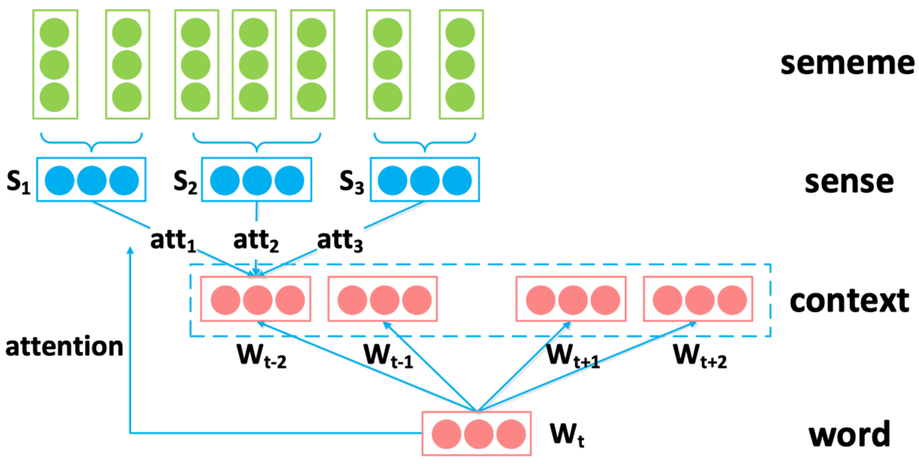
Word Embedding with Sememes

- Incorporate sense-sememe knowledge to learn word embeddings as well as sense and sememe embeddings

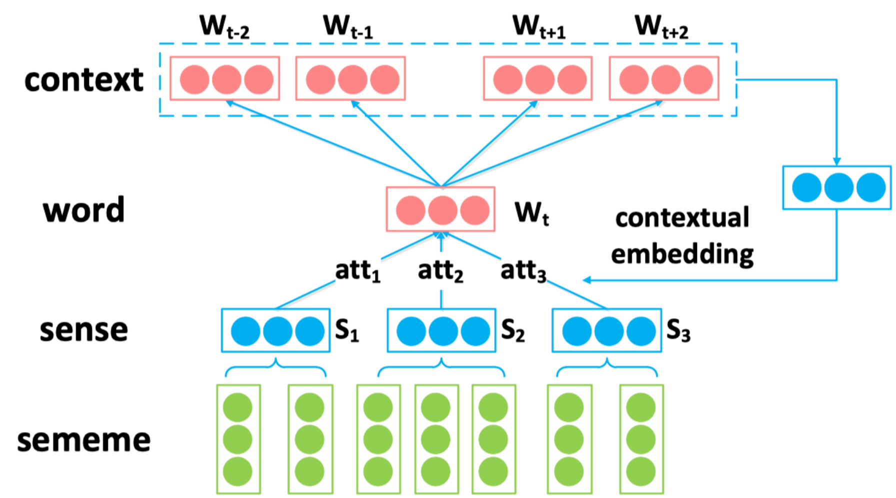


Word Embedding with Sememes

- Incorporate sense-sememe knowledge into word embeddings



Sememe Attention over Context Model



Sememe Attention over Target Model

Yilin Niu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Improved Word Representation Learning with Sememes. ACL 2017.

Experiment Results

- The enhanced word embeddings perform better on the tasks of analogy reasoning and word similarity

Model	Accuracy				Mean Rank			
	Capital	City	Relationship	All	Capital	City	Relationship	All
CBOW	49.8	85.7	**86.0**	64.2	36.98	1.23	62.64	37.62
GloVe	57.3	74.3	81.6	65.8	19.09	1.71	3.58	12.63
Skip-gram	66.8	93.7	76.8	73.4	137.19	1.07	2.95	83.51
SSA	62.3	93.7	81.6	71.9	45.74	1.06	3.33	28.52
MST	65.7	95.4	82.7	74.5	50.29	1.05	2.48	31.05
SAC	79.2	97.7	75.0	81.0	28.88	1.02	2.23	18.09
SAT	**82.6**	**98.9**	80.1	**84.5**	**14.78**	**1.01**	**1.72**	**9.48**

Experiment Examples

- The model can conduct sense disambiguation based on sememes and contexts

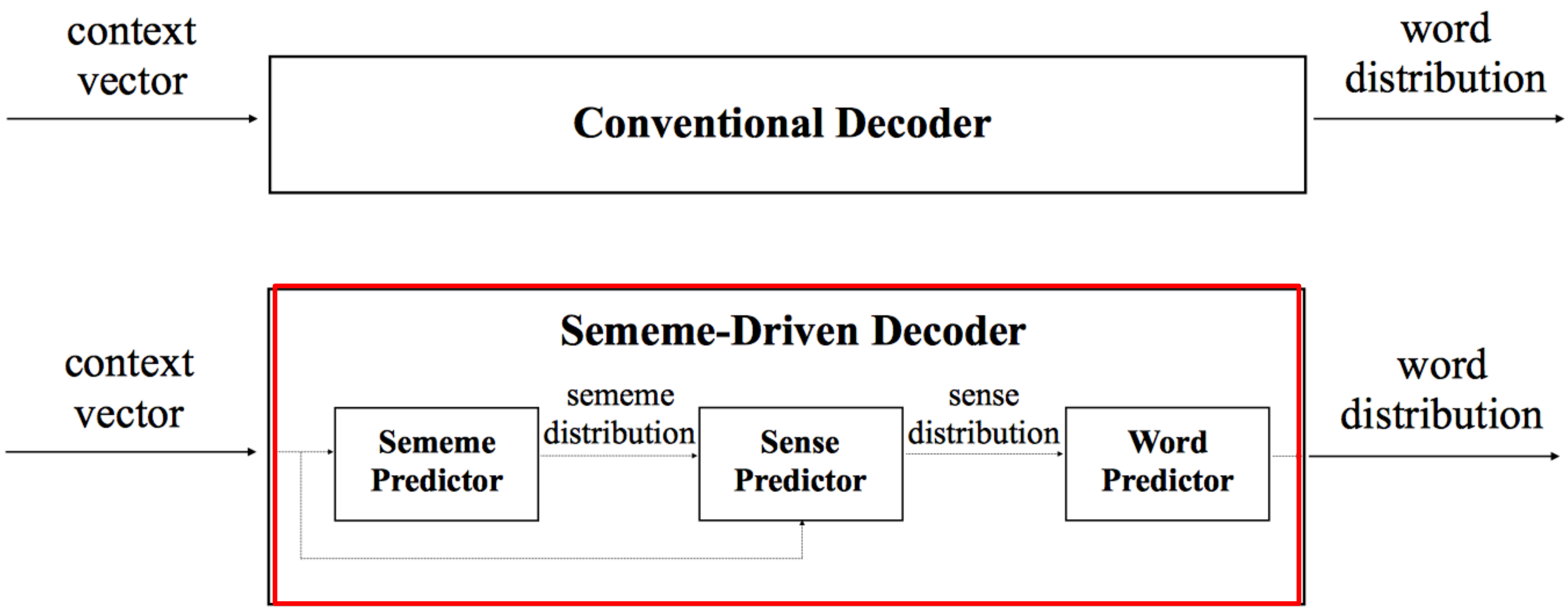
Word: 苹果("Apple brand/apple") sense1: *Apple brand* (computer, PatternValue, able, bring, SpeBrand) sense2: *duct* (fruit)		
苹果 素有果中王美称（**Apple** is always famous as the king of fruits） **苹果** 电脑无法正常启动（The **Apple brand** computer can not startup normally）	*Apple brand*: 0.28 *Apple brand*: 0.87	*apple*: 0.72 *apple*: 0.13
Word: 扩散("proliferate/metastasize") sense1: *proliferate* (disperse) sense2: *metastasize* (disperse, disease)		
防止疫情**扩散**（Prevent epidemic from **metastasizing**） 不**扩散**核武器条约（Treaty on the Non-**Proliferation** of Nuclear Weapons）	*proliferate*: 0.06 *proliferate*: 0.68	*metastasize*: 0.94 *metastasize*: 0.32
Word: 队伍("contingent/troops") sense1: *contingent* (community) sense2: *troops* (army)		
八支**队伍**进入第二阶段团体赛（Eight **contingents** enter the second stage of team competition） 公安基层**队伍**组织建设（Construct the organization of public security's **troops** in grass-roots unit）	*contingent*: 0.90 *contingent*: 0.15	*troops*: 0.10 *troops*: 0.85

Language Modeling

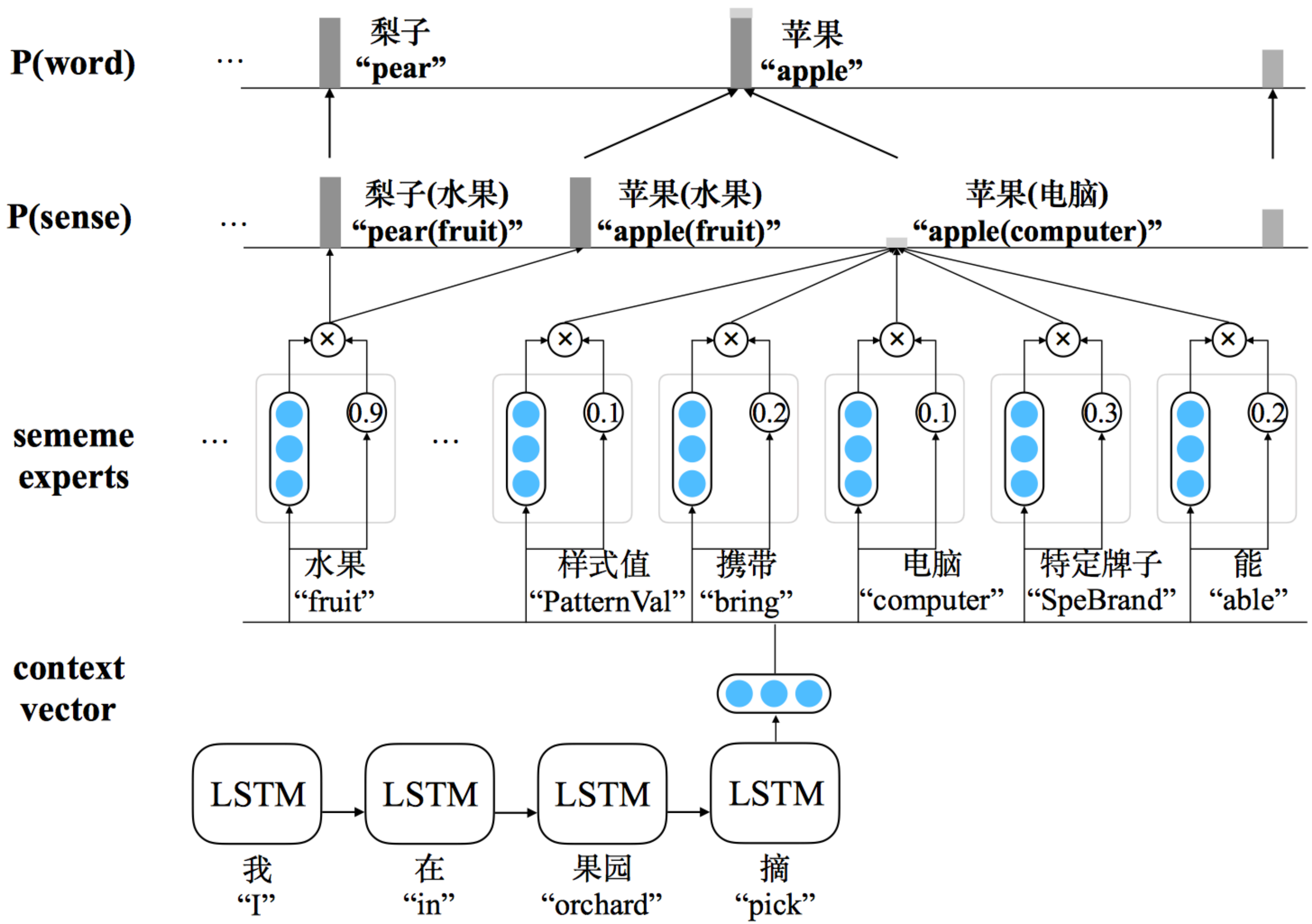
- Modeling word sequence with Markov property

The U.S. trade deficit last year is initially estimated to be 40 billion ______ .

- Sememe-Driven Language Modeling



Sememe-Driven Neural Language Modeling



Yihong Gu, Jun Yan, Hao Zhu, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin and Leyu Lin. Language Modeling with Sparse Product of Sememe Experts. EMNLP 2018.

Experiment Results

- Sememe knowledge can significantly reduce the perplexity of language models

Model	#Paras	Validation	Test
LSTM (medium)	24M	116.46	115.51
+ cHSM	24M	129.12	128.12
+ tHSM	24M	151.00	150.87
Tied LSTM (medium)	15M	105.35	104.67
+ cHSM	15M	116.78	115.66
+ MoS	17M	98.47	98.12
+ SDLM	17M	**97.75**	**97.32**
LSTM (large)	76M	112.39	111.66
+ cHSM	76M	120.07	119.45
+ tHSM	76M	140.41	139.61
Tied LSTM (large)	56M	101.46	100.71
+ cHSM	56M	108.28	107.52
+ MoS	67M	94.91	94.40
+ SDLM	67M	**94.24**	**93.60**
AWD-LSTM[4]	26M	89.35	88.86
+ MoS	26M	92.98	92.76
+ SDLM	27M	**88.16**	**87.66**

Experiment Examples

Example (1)		
去年 美国 贸易逆差 初步 估计 为 <N> _______ 。		
The U.S. trade deficit last year is initially estimated to be <N> _______ .		
Top 5 word prediction		
美元 “dollar”	， “,”	。 “.”
日元 “yen”	和 “and”	
Top 5 sememe prediction		
商业 “commerce”	**金融 “finance”**	**单位 “unit”**
多少 “amount”	专 “proper name”	
Example (2)		
阿 总理 _______ 已 签署 了 一 项 命令 。		
Albanian Prime Minister _______ has signed an order.		
Top 5 word prediction		
内 “inside”	**<unk>**	在 “at”
塔 “tower”	和 “and”	
Top 5 sememe prediction		
政 “politics”	**人 “person”**	花草 “flowers”
担任 “undertake”	水域 “waters”	

Semantic Composition

农民 (peasant)

起义 (uprising)

农民起义 (peasant uprising)

画 (draw)

句号 (a period)

画句号 (draw a period)

Semantic Composition (SC)

- Semantic Composition: The phenomenon that the meaning of a complex linguistic unit can be composed of the meanings of its constituents
- SC is regarded as fundamental truth of semantics
- A general modeling framework (Taking two-constituent unit for example)

composition function → f; combination rule → R

$$\boldsymbol{p} = f(w_1, w_2, R, \boldsymbol{K})$$

embedding of the complex linguistic unit → $\boldsymbol{p}$; embeddings of the constituents → w_1, w_2; external knowledge → $\boldsymbol{K}$

Modeling SC with Sememes

- A preliminary experiment of semantic composition degree of Multi-word Expressions (MWEs)

SCD	Our Computation Formulae	Examples	
		MWEs and Constituents	Sememes
3	$S_p = S_{w_1} \cup S_{w_2}$	农民起义(peasant uprising) 农民 (peasant) 起义(uprising)	事情|fact,职位|occupation,政|politics,暴动|uprise,人|human,农|agricultural 职位|**occupation**,人|**human**,农|**agricultural** 暴动|**uprise**,事情|**fact**,政|**politics**
2	$S_p \subsetneq (S_{w_1} \cup S_{w_2})$	几何图形(geometric figure) 几何 (geometry; how much) 图形(figure)	数学|math,图像|image 数学|**math**,知识|knowledge,疑问|question,功能词|funcword 图像|**image**
1	$S_p \cap (S_{w_1} \cup S_{w_2}) \neq \emptyset$ $\wedge S_p \not\subset (S_{w_1} \cup S_{w_2})$	应考(engage a test) 应 (deal with; echo; agree) 考(quiz; check)	考试|exam,从事|engage 处理|handle,回应|respond,同意|agree,遵循|obey,功能词|funcword,姓|surname 考试|**exam**,查|check
0	$S_p \cap (S_{w_1} \cup S_{w_2}) = \emptyset$	画句号(end) 画 (draw) 句号(period)	完毕|finish 画|draw,部件|part,图像|image, 文字|character,表示|express 符号|symbol,语文|text

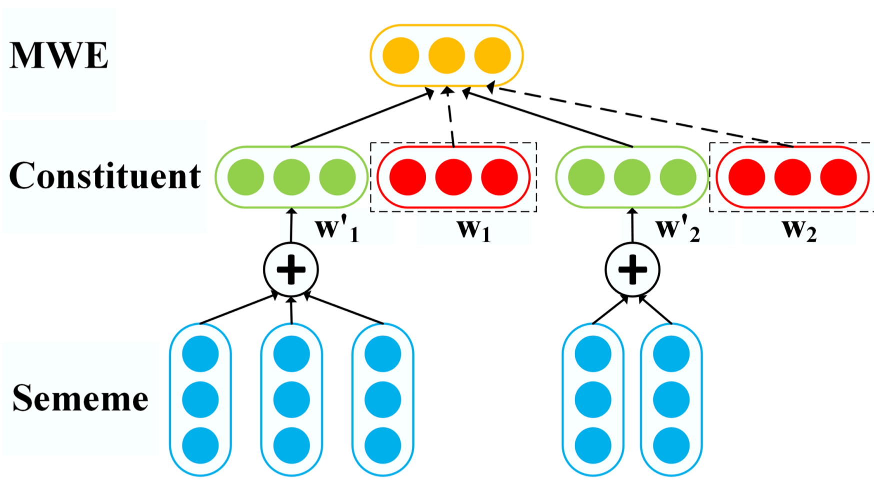
S_p, S_{w_1} and S_{w_2} : sememe sets of an MWE, its first constituent and second constituent.

Pearson's correlation with human evaluation: **0.75**

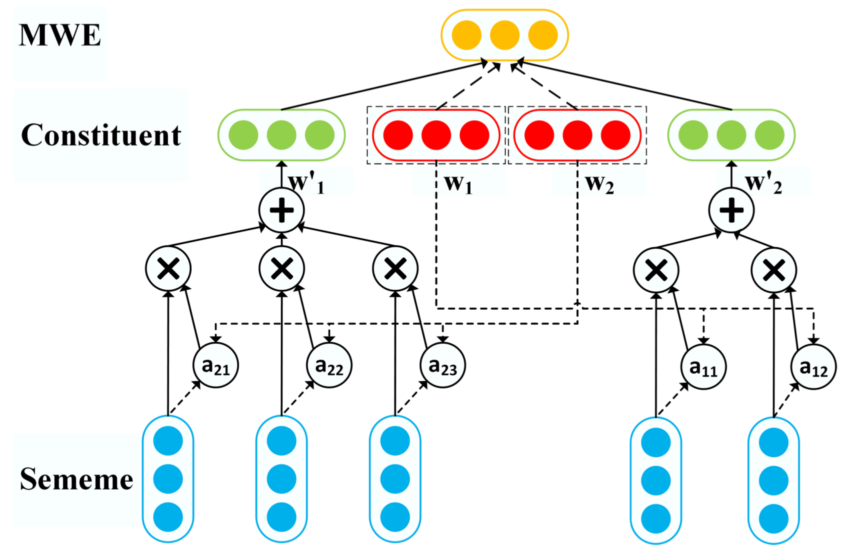
Fanchao Qi, Junjie Huang, Chenghao Yang, Zhiyuan Liu, Xiao Chen, Qun Liu, Maosong Sun. Modeling Semantic Compositionality with Sememe Knowledge. ACL 2019.

Modeling SC with Sememes

- Sememe-incorporated SC models



SC with Aggregated Sememe Model (**SC-AS**)



SC with Mutual Sememe Attention Model (**SC-MSA**)

Fanchao Qi, Junjie Huang, Chenghao Yang, Zhiyuan Liu, Xiao Chen, Qun Liu, Maosong Sun. Modeling Semantic Compositionality with Sememe Knowledge. ACL 2019.

Modeling SC with Sememes

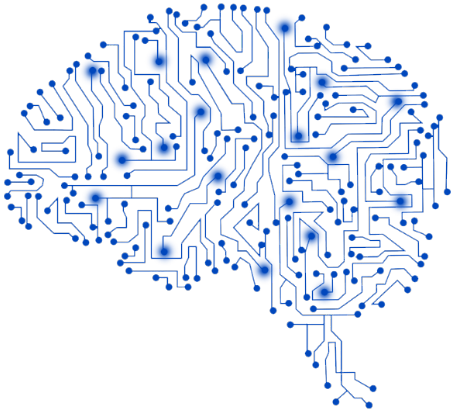
- Experimental Results

Framework	Method	WS240	WS297	COS960
$f(\mathbf{w}_1, \mathbf{w}_2)$	ADD	50.8	53.1	49.1
	MUL	19.6	21.6	−3.9
	TIM	47.4	54.2	50.5
	RNTN	42.5	53.6	55.8
	RAE	61.3	59.9	59.6
	SCAS-S	61.4	57.0	60.1
$f(\mathbf{w}_1, \mathbf{w}_2, K)$	SCAS	60.2	60.5	61.4
	SCMSA	**61.9**	58.7	60.5
$f(\mathbf{w}_1, \mathbf{w}_2, K, R)$	SCAS+R	59.0	60.8	**61.8**
	SCMSA+R	61.4	**61.2**	60.4

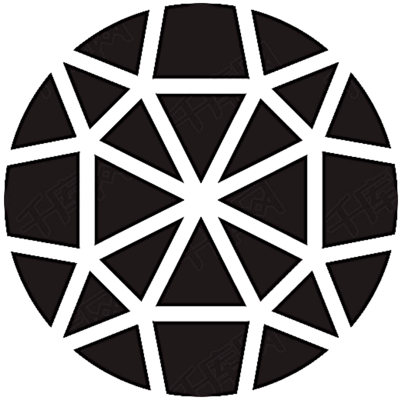
Intrinsic Evaluation
(MWE Similarity)

Framework	Method	Sememe Prediction	
		MAP	F1 Score
$f(\mathbf{w}_1, \mathbf{w}_2)$	ADD	40.7	23.2
	MUL	11.2	0.3
	TIM	46.8	35.3
	RNTN	47.7	35.3
	RAE	44.0	30.8
	SCAS-S	39.0	27.9
$f(\mathbf{w}_1, \mathbf{w}_2, K)$	SCAS	52.2	41.3
	SCMSA	55.1	43.4
$f(\mathbf{w}_1, \mathbf{w}_2, K, R)$	SCAS+R	56.8	**46.1**
	SCMSA+R	**58.3**	46.0

Extrinsic Evaluation
(MWE Sememe Prediction)



Prediction



Data-Driven DL

Symbol-based Sememe Knowledge

HowNet Is Not Large Enough

- HowNet: 127,266 Chinese words, 50,879 English words (only 32.8% of WordNet vocabulary)
- New words are consistently emerging
- Meanings of existing words keeps changing
- No words in other languages are annotated with sememes

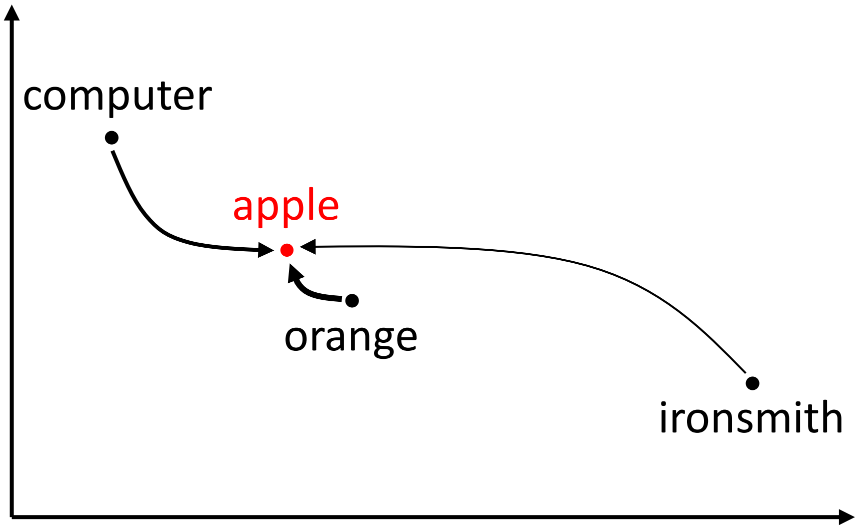
Application of Sememe KB is limited!

Expanding Sememe KBs by Sememe Prediction

- Manual Annotation: time-consuming, labor-intensive and hard to keep agreement
- Data-Driven ML methods can help annotate sememes for unannotated words automatically
- Simplification: ignore hierarchical structures of sememes and modeling it as a multi-label classification problem

Sememe Prediction by Collaborative Filtering and Matrix Factorization

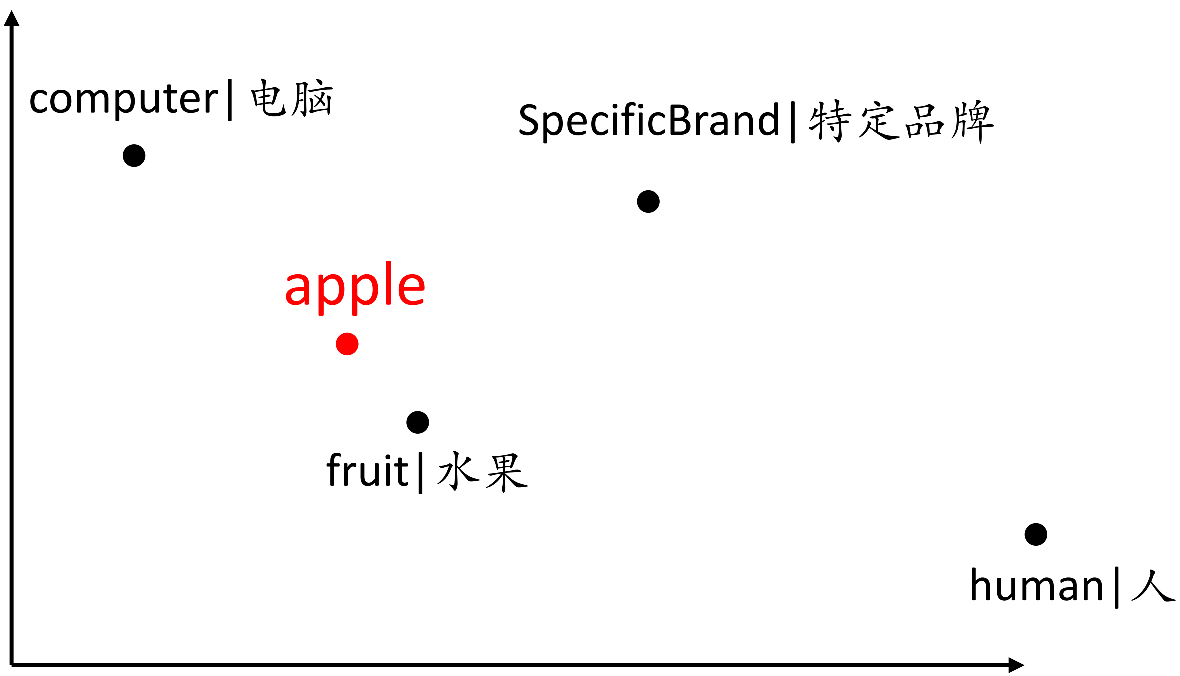
- Word - User, Sememe - Item
  - Sememe Prediction: recommending items for users
- Collaborative Filtering: Semantically similar words should have similar sememes



Ruobing Xie, Xingchi Yuan, Zhiyuan Liu, Maosong Sun. Lexical Sememe Prediction via Word Embeddings and Matrix Factorization. IJCAI 2017.

Sememe Prediction by Collaborative Filtering and Matrix Factorization

- Matrix Factorization
  - The embedding of a word should be close to the embeddings of its sememes



Sememe Prediction by Collaborative Filtering and Matrix Factorization

- Experimental Results

Method	MAP
SPSE	0.554
SPASE	0.506
GloVe+LR	0.662
SPWE	0.676
SPWE+SPASE	0.683
SPWE+SPSE	**0.713**

Overall Sememe Prediction Performance

POS	number of words	MAP
adverb	136	0.568
adjective	808	0.544
verb	1,867	0.583
noun	3,556	0.747

Effect of POS Tags

word frequency	number of words	MAP
<800	1,659	0.817
800 - 3,000	1,494	0.736
3,001 - 15,000	1,672	0.690
>15,000	1,311	0.596

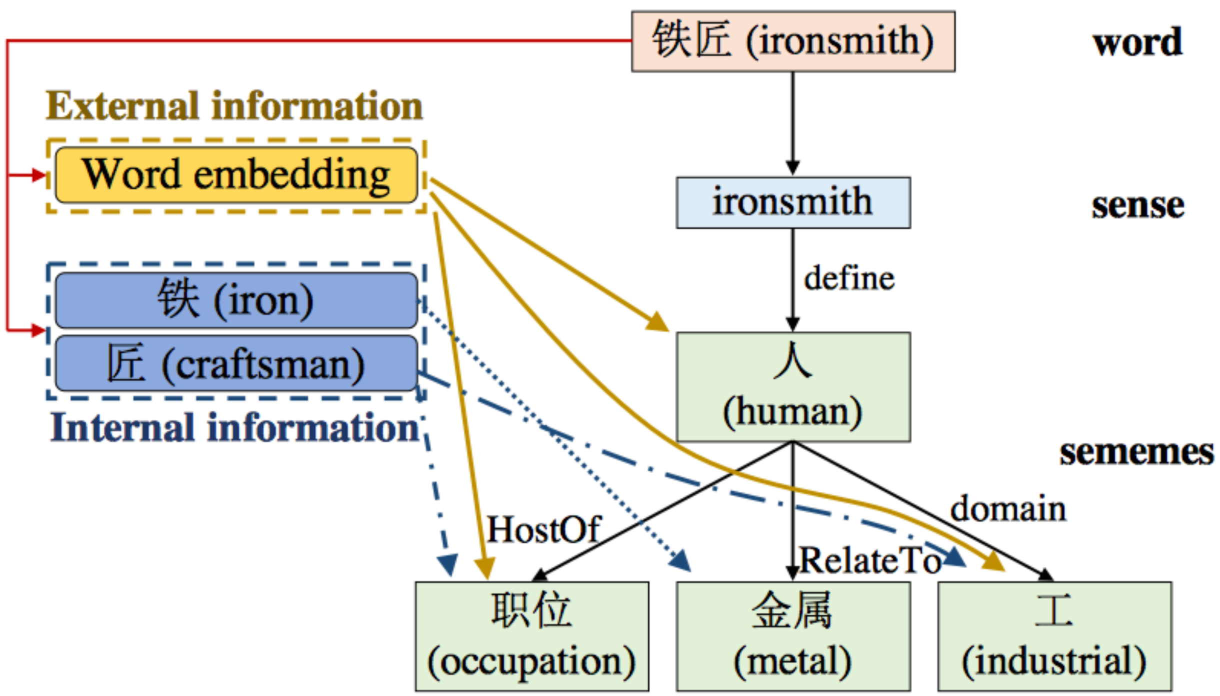
Effect of Word Frequency

words	Top 5 sememes prediction
网迷(webaholic)	**人(human)**, **因特网(internet)**, **经常(frequency)**, **利用(use)**, 喜欢(fond of)
专递(express mail)	**邮寄(post)**, **信件(letter)**, **快(fast)**, 事情(fact), 车(landvehicle)
电影业(film industry)	**事务(affairs)**, **艺(entertainment)**, **表演物(shows)**, 拍摄(take picture), **制造(produce)**
漂流(rafting)	船(ship), 旅游(tour), 游(swim), 水域(waters), 消闲(whileaway)
公羊(ram)	**牲畜(livestock)**, **男(male)**, 女(female), 走兽(beast), 饲养(foster)

Examples

Incorporating Chinese Characters

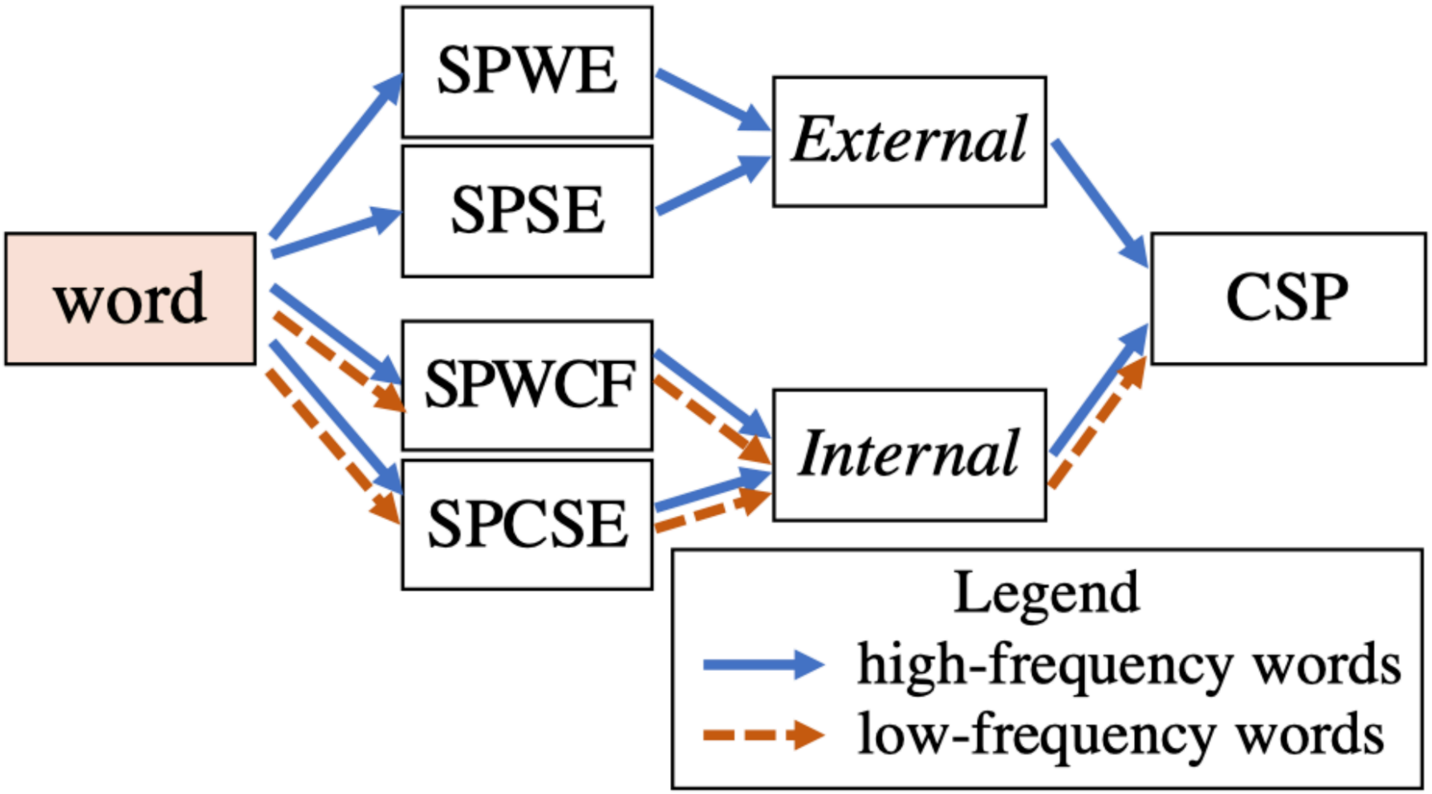
- Incorporating Chinese characters can alleviate the problem of bad sememe prediction performance of low-frequency words caused by poor embeddings



Huiming Jin, Hao Zhu, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin, Leyu Lin. Incorporating Chinese Characters of Words for Lexical Sememe Prediction. ACL 2018.

Incorporating Chinese Characters

- Combining word embeddings (external information) and characters (internal information)



Experiment Results

word frequency occurrences	⩽ 50 8537	51– 100 4868	101 – 1,000 3236	1,001 – 5,000 2036	5,001 – 10,000 663	10,001 – 30,000 753	>30,000 686
SPWE	0.312	0.437	0.481	0.558	0.549	0.556	0.509
SPSE	0.187	0.273	0.339	0.409	0.407	0.424	0.386
SPWE + SPSE	0.284	0.414	0.478	0.556	0.548	0.554	0.511
SPWCF	0.456	0.414	0.400	0.443	0.462	0.463	0.479
SPCSE	0.309	0.291	0.286	0.312	0.339	0.353	0.342
SPWCF + SPCSE	0.467	0.437	0.418	0.456	0.477	0.477	0.494
SPWE + fastText	0.495	0.472	0.462	0.520	0.508	0.499	0.490
CSP	**0.527**	**0.555**	**0.555**	**0.626**	**0.632**	**0.641**	**0.624**

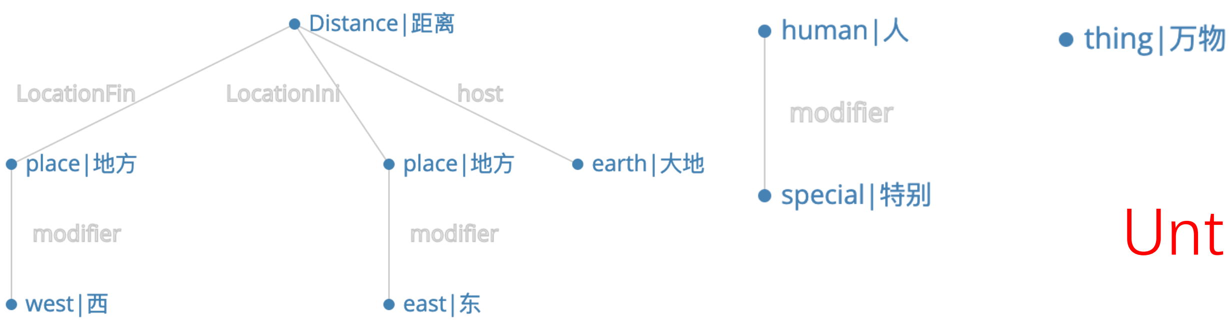
Sememe Prediction Performance of Different Models

words	models	Top 5 sememes
钟表匠 (clockmaker)	internal	**人(human)**, **职位(occupation)**, 部件(part), **时间(time)**, **告诉(tell)**
	external	**人(human)**, 专(ProperName), 地方(place), 欧洲(Europe), 政(politics)
	ensemble	**人(human)**, **职位(occupation)**, **告诉(tell)**, **时间(time)**, **用具(tool)**
奥斯卡 (Oscar)	internal	**专(ProperName)**, 地方(place), 市(city), 人(human), 国都(capital)
	external	**奖励(reward)**, **艺(entertainment)**, **专(ProperName)**, 用具(tool), **事情(fact)**
	ensemble	**专(ProperName)**, **奖励(reward)**, **艺(entertainment)**, 著名(famous), 地方(place)

Examples

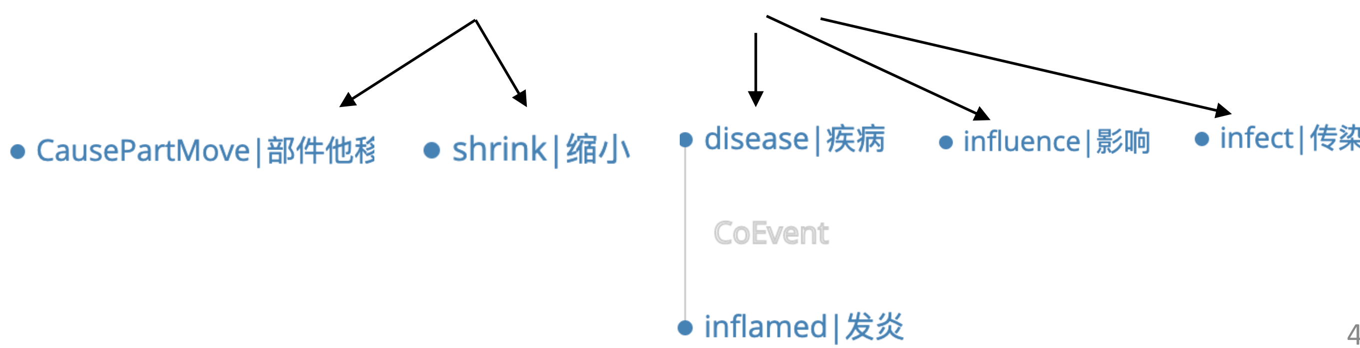
Cross-lingual Lexical Sememe Prediction

- Goal: Predicting sememes for words in other languages
- Translation doesn't work
  - E.g., 东西 → thing; stuff; east and west:



Untranslatability

  - E.g., contract → 合同；收缩； 感染



Cross-lingual Lexical Sememe Prediction

- Learn bilingual word embeddings which are in a unified semantic space and then adopt mono-lingual sememe prediction method
- Source Language: the language whose words have sememe annotation
- Target Language: the language to which we target at adding sememe annotation

Fanchao Qi, Yankai Lin, Maosong Sun, Hao Zhu, Ruobing Xie and Zhiyuan Liu. Cross-lingual Lexical Sememe Prediction. EMNLP 2018.

Cross-lingual Lexical Sememe Prediction

- Framework
  - Bilingual Word Representation Learning
    - Mono-lingual word representation learning
    - Cross-lingual word embeddings alignment
    - Incorporating sememe information of source language
  - Sememe Prediction for words in the target language

Experimental Results

- Cross-lingual Sememe Prediction Results

Method	Seed Lexicon	Sememe Prediction	
		MAP	F1
BiLex	1000	27.57	16.08
	2000	33.79	22.33
	2000	35.78	25.74
	6000	38.29	28.71
CLSP-WR	1000	28.12	18.55
	2000	33.78	23.64
	2000	38.30	27.74
	6000	41.23	28.71
CLSP-SE	1000	31.78	18.22
	2000	37.70	24.31
	2000	40.77	29.33
	6000	**43.16**	**32.49**

Method	Word Frequency	Sememe Prediction	
		MAP	F1
BiLex	< 200	30.35	21.83
	200 - 500	34.83	25.95
	501 - 2500	40.21	28.62
	> 2500	47.56	35.80
CLSP-WR	< 200	34.73	24.41
	200 - 500	39.50	29.49
	501 - 2500	43.92	33.87
	> 2500	47.33	34.99
CLSP-SE	< 200	36.54	27.49
	200 - 500	41.46	30.09
	501 - 2500	45.35	35.01
	> 2500	**49.34**	**37.16**

Experimental Results

- Byproduct: Better monolingual and bilingual word embeddings

Bilingual Lexicon Induction

Method	Lexicon Induction	
	P@1	P@5
BiLex	25.89	29.59
CLSP-WR	25.83	31.03
CLSP-SE	**26.91**	**32.17**

Word Similarity Computation

Method	Chinese (source)		English (target)	
	WS-240	WS-297	WS-353	SL-999
BiLex	60.36	62.17	60.46	27.22
CLSP-WR	**61.27**	65.25	60.46	27.22
CLSP-SE	60.84	**65.62**	**62.47**	**28.79**

Experimental Results

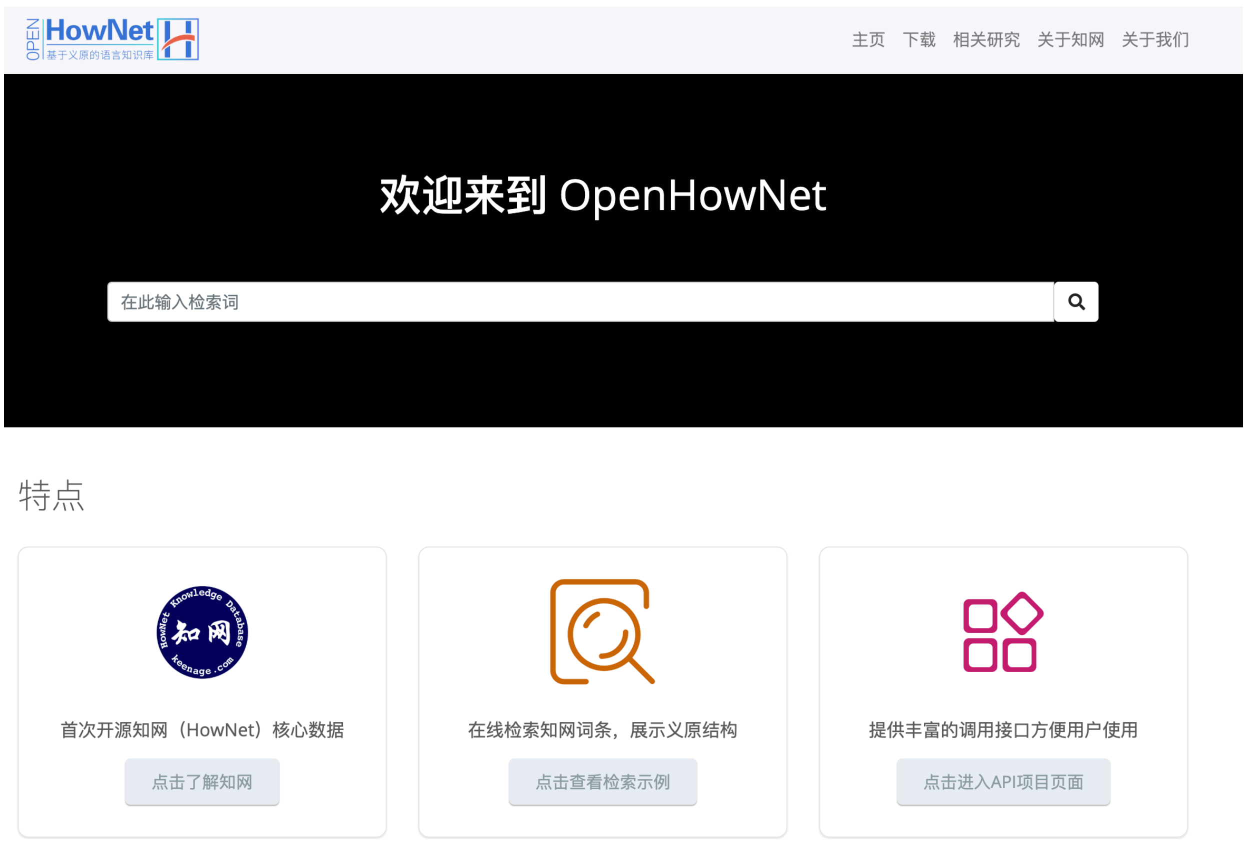
- Two examples

Type	Words	Sememes
Target Word	handcuffs	**用具"tool",警"police",扣住"detain",人"human",有罪"guilty"**
5 Nearest Source Words	手铐"handcuffs"	**有罪"guilty",警"police",人"human",扣住"detain",用具"tool"**
	镣铐"shackles"	**有罪"guilty",警"police",人"human",扣住"detain",用具"tool"**
	绑"tie"	包扎"wrap"
	螺丝刀"screwdriver"	**用具"tool"**,放松"loosen",勒紧"tighten"
	绳"rope"	线"linear",材料"material",拴连"fasten"
Target Word	canoeist	**锻炼"exercise", 人"human",体育"sport"**,事情"fact", 船"ship"
5 Nearest Source Words	短跑"sprint"	事情"fact", **锻炼"exercise"**, **体育"sport"**
	独木舟"canoe"	船"ship"
	皮艇"kayak"	船"ship"
	名将"sportsstar"	著名"famous", **人"human"**, 官"official", 军"military"
	皮划艇"kayak"	事情"fact", **锻炼"exercise"**, **体育"sport"**

Sememes in **red boldface** are annotated ones for the target word in HowNet.

OpenHowNet

https://openhownet.thunlp.org/



OpenHowNet

OPEN HowNet 基于义原的语言知识库

主页 下载 相关研究 关于知网 关于我们

铁匠(ironsmith)

ID:160550

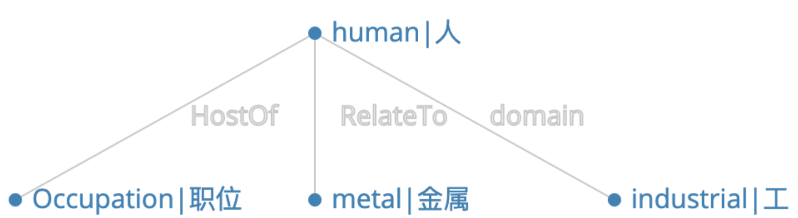
词性:

noun

HowNet定义:

{human|人:HostOf={Occupation|职位},RelateTo={metal|金属},domain={industrial|工}}

义原树演示



搜索

在此输入检索词

近义词

板金工(015234)
锻工(050692)
锻工(050690)
锻工(050693)
锻工(050691)
钢铁工人(062869)

OpenHowNet

https://github.com/thunlp/OpenHowNet-API

OpenHowNet API

接口说明

接口	功能说明	参数说明
get(self, word, language=None)	检索HowNet中词语标注的完整信息	word表示待查词，language为en(英文)/ch(中文)，默认双语同时查找
get_sememes_by_word(self, word, structured=False, lang='ch', merge=False, expanded_layer=-1)	检索输入词的义原，可以选择是否合并多义，也可以选择是否以结构化的方式返回，还可以指定展开层数。	word表示待查词，language为en(英文)/ch(中文), structured表示是否以结构化的方式返回，merge控制是否合并多义项，expanded_layer控制展开层数，默认全展开。
initialize_sememe_similarity_calculation(self)	初始化基于义原的词语相似度计算（需要读取相关文件并有短暂延迟）	
calculate_word_similarity(self, word0, word1)	计算基于义原的词语相似度，调用前必须先调用上一个函数进行初始化	word0和word1表示待查的词语相似度对
get_nearest_words_via_sememes(self, word, K=10)	在使用基于义原的词语相似度度量下，计算和检索词最接近的K个词	Word表示检索词，K表示K近邻算法取的Top-K

请阅读文档 `./HowNet/Standards.html` 或查看Demo `DemoForHowNetPackage.ipynb` 了解更多。

Fanchao Qi, Chenghao Yang, Zhiyuan Liu, Qiang Dong, Maosong Sun, Zhendong Dong. OpenHowNet: An Open Sememe-based Lexical Knowledge Base. arXiv 2019.

Sememe Computation Paper List

https://github.com/thunlp/SCPapers

- Fanchao Qi, Junjie Huang, Chenghao Yang, Zhiyuan Liu, Xiao Chen, Qun Liu, Maosong Sun. **Modeling Semantic Compositionality with Sememe Knowledge**. ACL 2019.
- Yihong Gu, Jun Yan, Hao Zhu, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin and Leyu Lin. **Language Modeling with Sparse Product of Sememe Experts**. EMNLP 2018.
- Fanchao Qi, Yankai Lin, Maosong Sun, Hao Zhu, Ruobing Xie, Zhiyuan Liu. **Cross-lingual Lexical Sememe Prediction**. EMNLP 2018.
- Huiming Jin, Hao Zhu, Zhiyuan Liu, Ruobing Xie, Maosong Sun, Fen Lin, Leyu Lin. **Incorporating Chinese Characters of Words for Lexical Sememe Prediction**. ACL 2018.
- Xiangkai Zeng, Cheng Yang, Cunchao Tu, Zhiyuan Liu, Maosong Sun. **Chinese LIWC Lexicon Expansion via Hierarchical Classification of Word Embeddings with Sememe Attention**. AAAI 2018.
- Ruobing Xie, Xingchi Yuan, Zhiyuan Liu, Maosong Sun. **Lexical Sememe Prediction via Word Embeddings and Matrix Factorization**. IJCAI 2017.
- Yilin Niu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. **Improved Word Representation Learning with Sememes**. ACL 2017.

PART TWO

World Knowledge

Zhiyuan Liu

Tsinghua NLP

World Knowledge Graph

- Example: Google Knowledge Graphs
- Represent world knowledge with triples

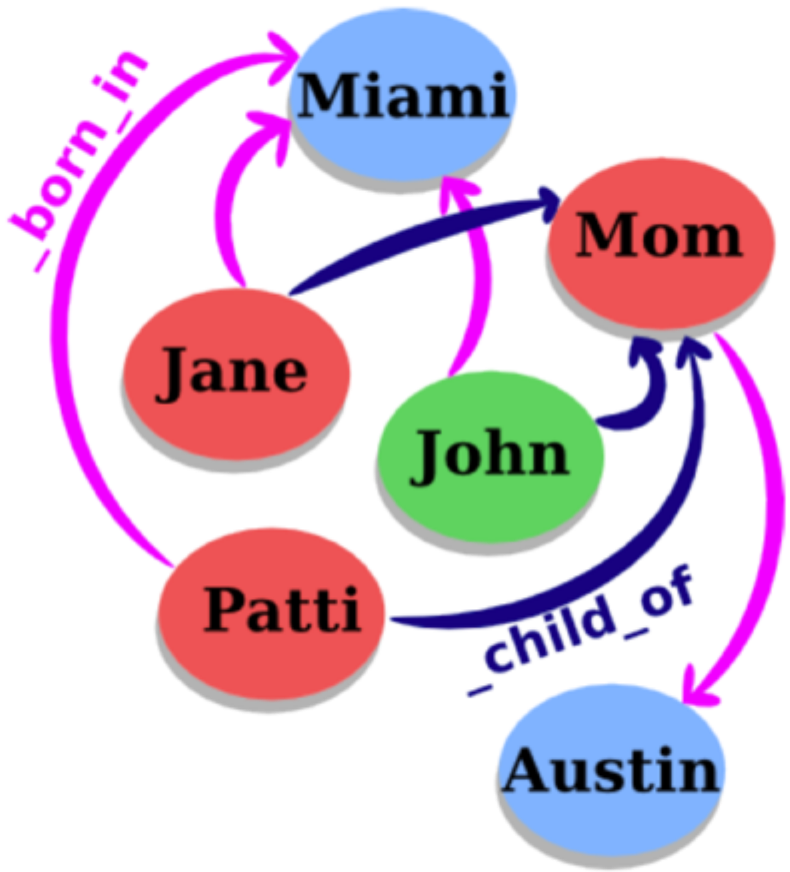
write

William Shakespeare

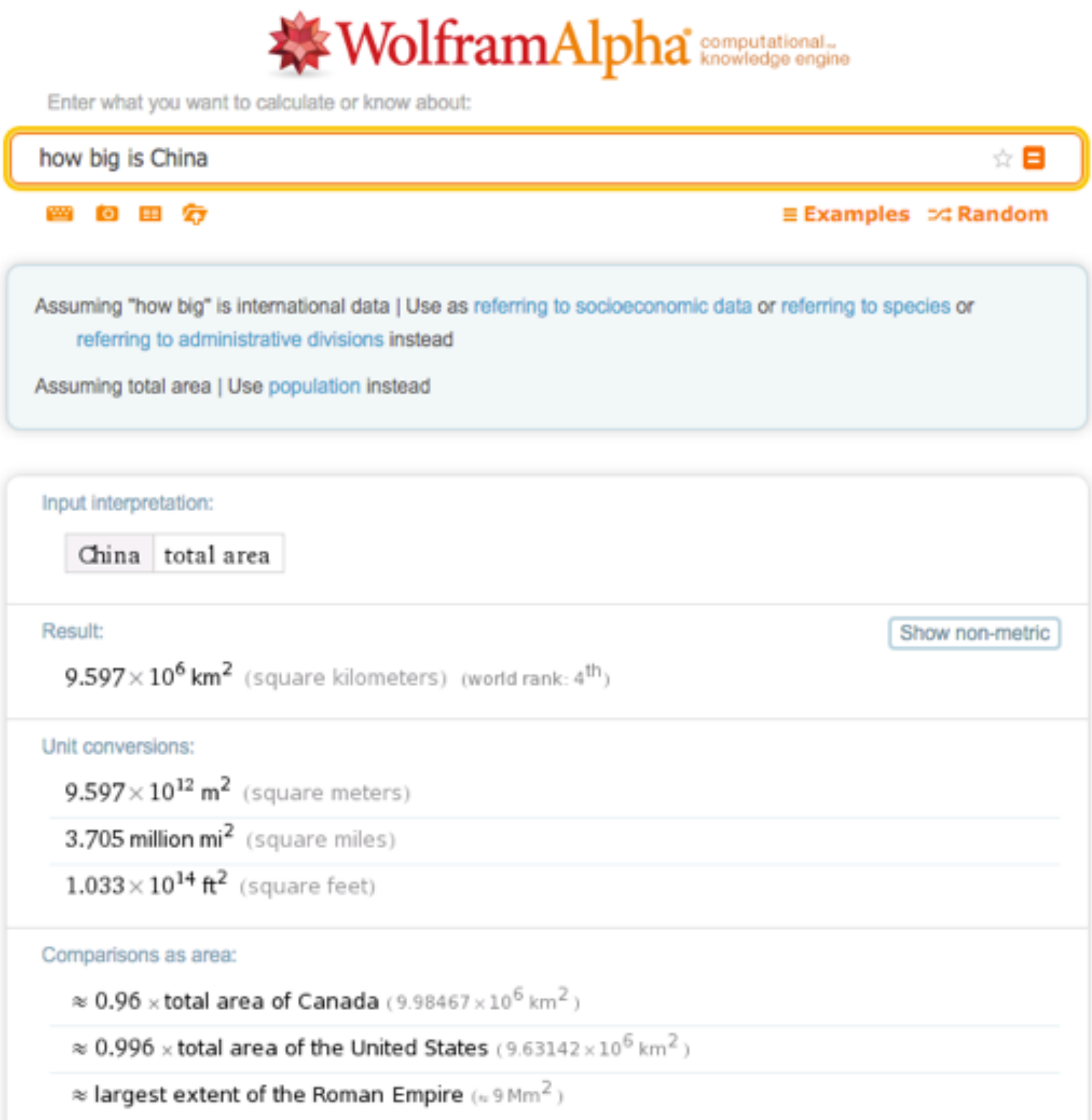
Romeo and Juliet

Entities and Relations

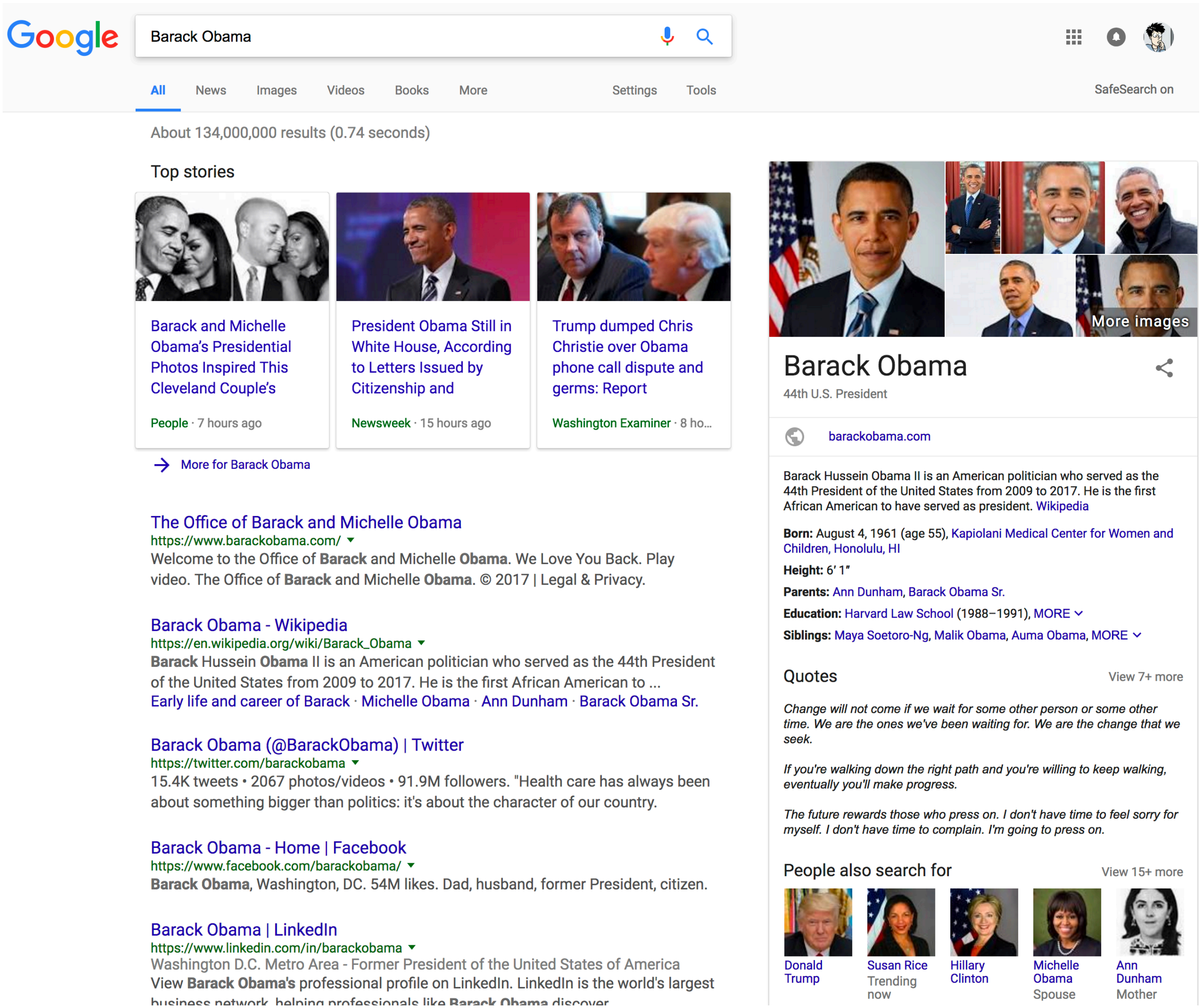
- Knowledge structured as graph
  - Each node = an entity
  - Each edge = a relation
- Fact: (head, relation, tail)
  - head = subject entity
  - relation = relation type
  - tail = object entity
- Typical KGs
  - Freebase: World KG



KG Application: Question Answering



KG Application: Search Engine



KG Application: Inference

Google

who is Barack Obama's wife's father

All News Images Videos Shopping More Settings Tools

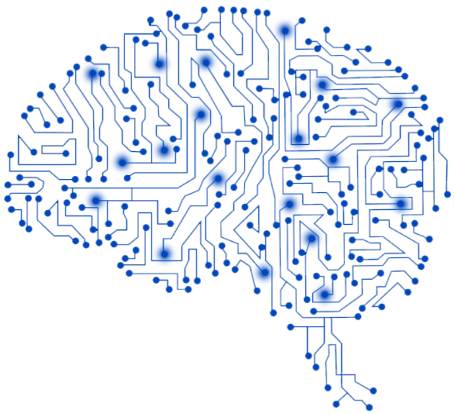
About 498,000 results (0.67 seconds)

Michelle Obama / Father

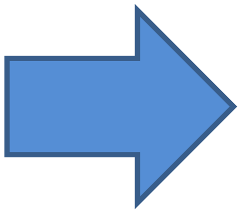
Fraser C. Robinson III

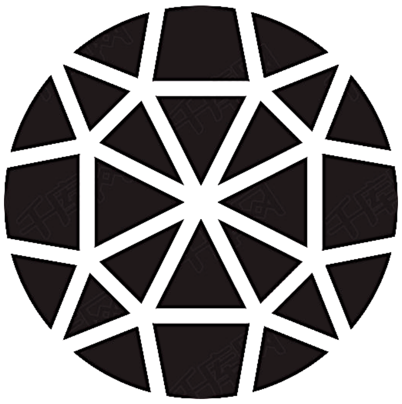
More about Fraser C. Robinson III

Feedback



Prediction





Data-Driven
DL

Symbol-based
World Knowledge

Knowledge Acquisition (KA)

- Goal: Automatically Extract knowledge from Web

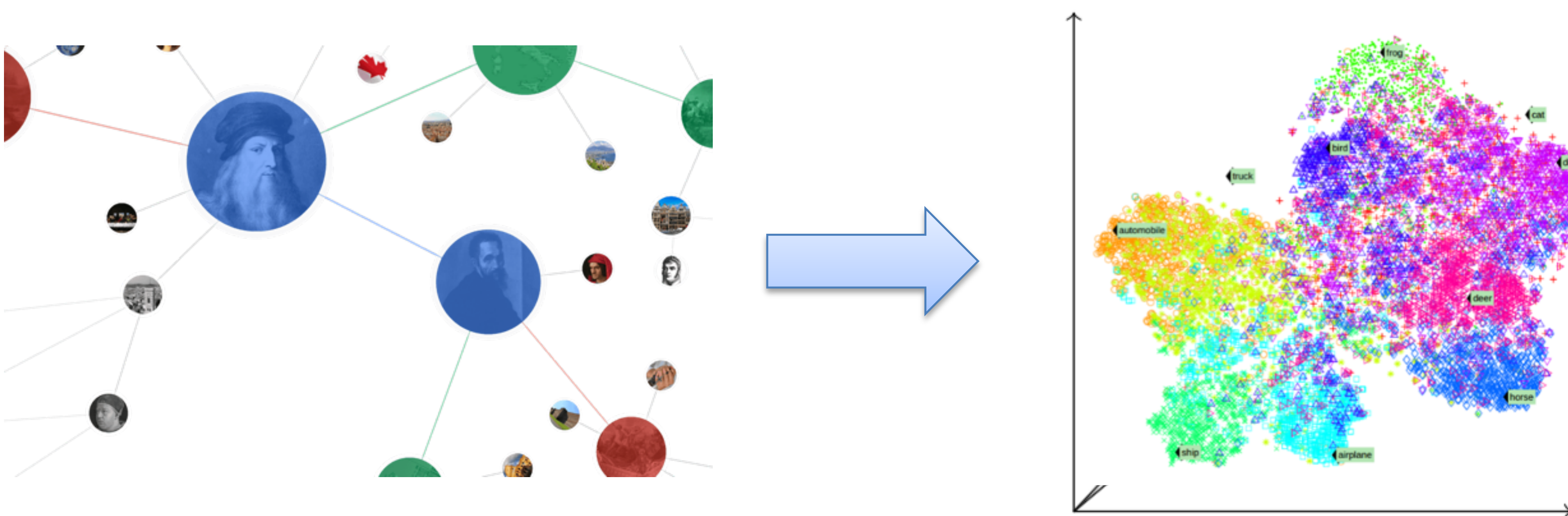
Knowledge Acquisition

Relation Prediction
From **KG**

Relation Extraction
From **Text**

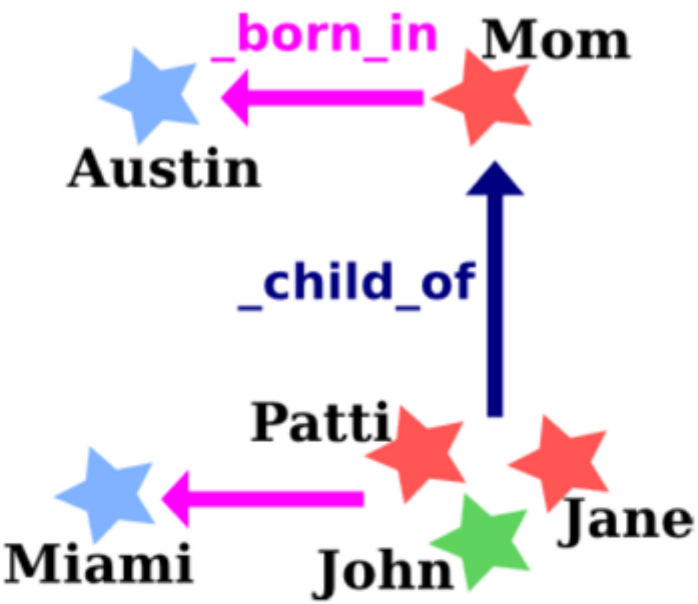
Relation Prediction from KG

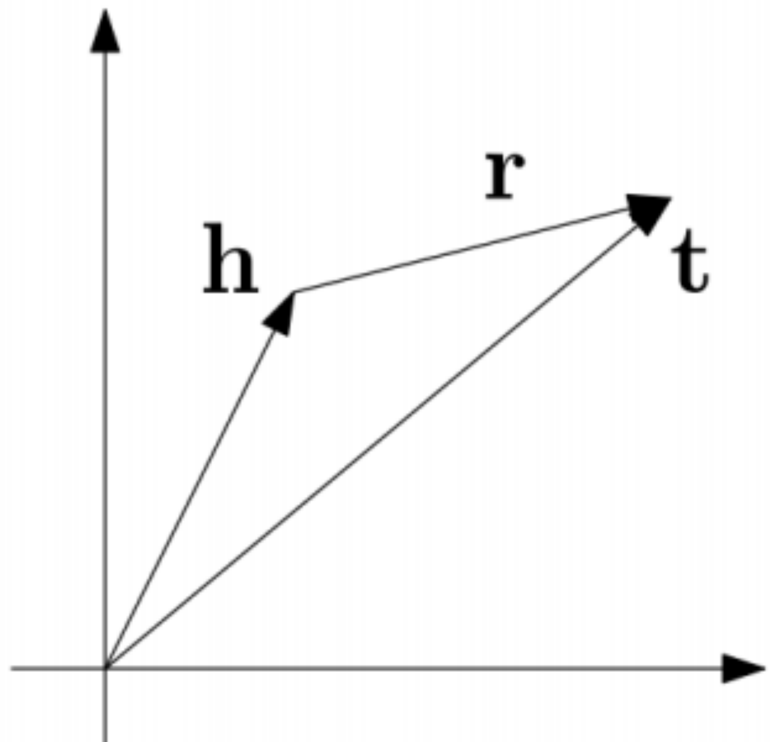
- Typical representations for KG
  - Symbolic triples (RDF)
  - Cannot efficiently measure semantic relatedness of entities
- How: Encode KGs into low-dimensional vector spaces



Distributed Learning for World Knowledge

- TransE regards the tail entity as a relation translation from the head entity for each (head, relation, tail)





Objective: **h + r = t**

Objective Function

- Energy Function
  - For correct (h, r, t)， requires **h** + **r** = **t**

$$f(h, r, t) = \|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|$$

Objective Function

Energy Function $f(h,r,t) = \|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|$

Objective Function $\sum_{(h,r,t)\in\Delta} \sum_{(h',r,t')\in\Delta'} [\gamma + f(h,r,t) - f(h',r,t')]_+$

where $[x]_+ = \max(0,x)$

w.r.t $\|\boldsymbol{h}\| \leq 1, \|\boldsymbol{t}\| \leq \boldsymbol{1}$

Δ triple sets in KG

Δ' negative triple sets not in KG

Entity/Relation Prediction with TransE

WALL-E _has_genre ?

h + r = ?

Entity/Relation Prediction with TransE

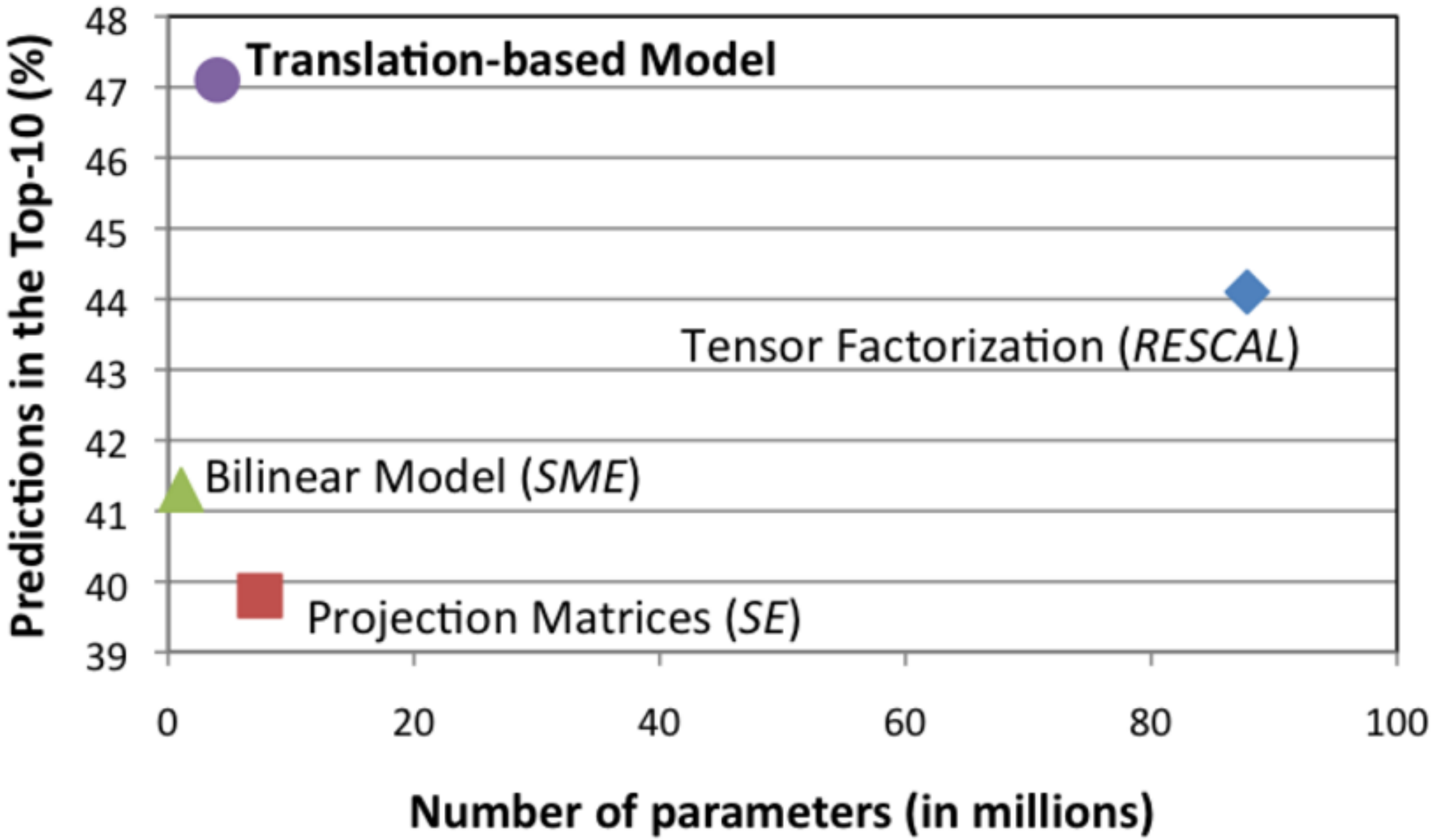
WALL-E _has_genre

Animation
Computer animation
Comedy film
Adventure film
Science Fiction
Fantasy
Stop motion
Satire
Drama
Connecting

h + r = t

Prediction Performance

Dataset: Freebase15K



TransE Examples

Entity	**Tsinghua_University**	**A.C._Milan**
1	University_of_Victoria	Inter_Milan
2	St._Stephen's_College,_Delhi	Celtic_F.C.
3	University_of_Ottawa	FC_Barcelona
4	University_of_British_Columbia	Genoa_C.F.C.
5	Peking_University	Udinese_Calcio
6	Utrecht_University	Real_Madrid_C.F.
7	Dalhousie_University	FC_Bayern_Munich
8	Brasenose_College,_Oxford	Bolton_Wanderers_F.C.
9	Cardiff_University	Borussia_Dortmund
10	Memorial_University_of_Newfoundland	Hertha_BSC_Berlin

TransE Examples

Entity	China	Barack_Obama	Apple
1	Japan	George_W._Bush	Onion
2	Taiwan	Nancy_Pelosi	Strawberries
3	South_Korea	John_Kerry	Avocado
4	Argentina	Hillary_Rodham_Clinton	Pear
5	North_Korea	Al_Gore	Cabbage
6	Hungary	George_H._W._Bush	Broccoli
7	Israel	John_McCain	Egg
8	Australia	Colin_Powell	Cheese
9	Iceland	Bill_Clinton	Bread
10	Hong_Kong	Charles_B._Rangel	Tomato

TransE Examples

Relation	/people/person/nationality	/location/location/contains
1	/people/person/places_lived	/base/aareas/schema/administrative_area/administrative_children
2	/people/person/place_of_birth	/location/country/administrative_divisions
3	/people/person/spouse_s	/location/country/first_level_divisions
4	/base/popstra/celebrity/vacations_in	/location/country/capital
5	/government/politician/government_positions_held	/award/award_nominee/award_nominations
6	/people/deceased_person/place_of_death	/location/administrative_division/capital
7	/olympics/olympic_athlete/country	/location/us_county/county_seat
8	/olympics/olympic_athlete/medals_won	/base/aareas/schema/administrative_area/capital
9	/music/artist/origin	/location/us_county/hud_county_place
10	/people/person/employment_history	/award/award_winner/awards_won

TransE Examples

Head	China	Barack_Obama
Relation	**/location/location/adjoin**	**/education/education/institution**
1	Japan	Harvard_College
2	Taiwan	Massachusetts_Institute_of_Technology
3	Israel	American_University
4	South_Korea	University_of_Michigan
5	Argentina	Columbia_University
6	France	Princeton_University
7	Philippines	Emory_University
8	Hungary	Vanderbilt_University
9	North_Korea	University_of_Notre_Dame
10	Hong_Kong	Texas_A&M_University

TransE Examples

Head	Stanford_University	Apple	Titanic
Relation	**/education/educational_institution/students_graduates**	**/food/food/nutrients**	**/film/film/genre**
1	Steven_Spielberg	Lipid	War_film
2	Ron_Howard	Protein	Period_piece
3	Stan_Lee	Valine	Drama
4	Barack_Obama	Tyrosine	History
5	Milton_Friedman	Serine	Biography
6	Walter_F._Parkes	Iron	Film_adaptation
7	Michael_Cimino	Cystine	Adventure_Film
8	Gale_Anne_Hurd	Pantothenic_acid	Action_Film
9	Bryan_Singer	Vitamin_A	Political_drama
10	Aaron_Sorkin	Sugar	Costume_drama

Challenges

1. How to deal with **complicated relations** for KA?

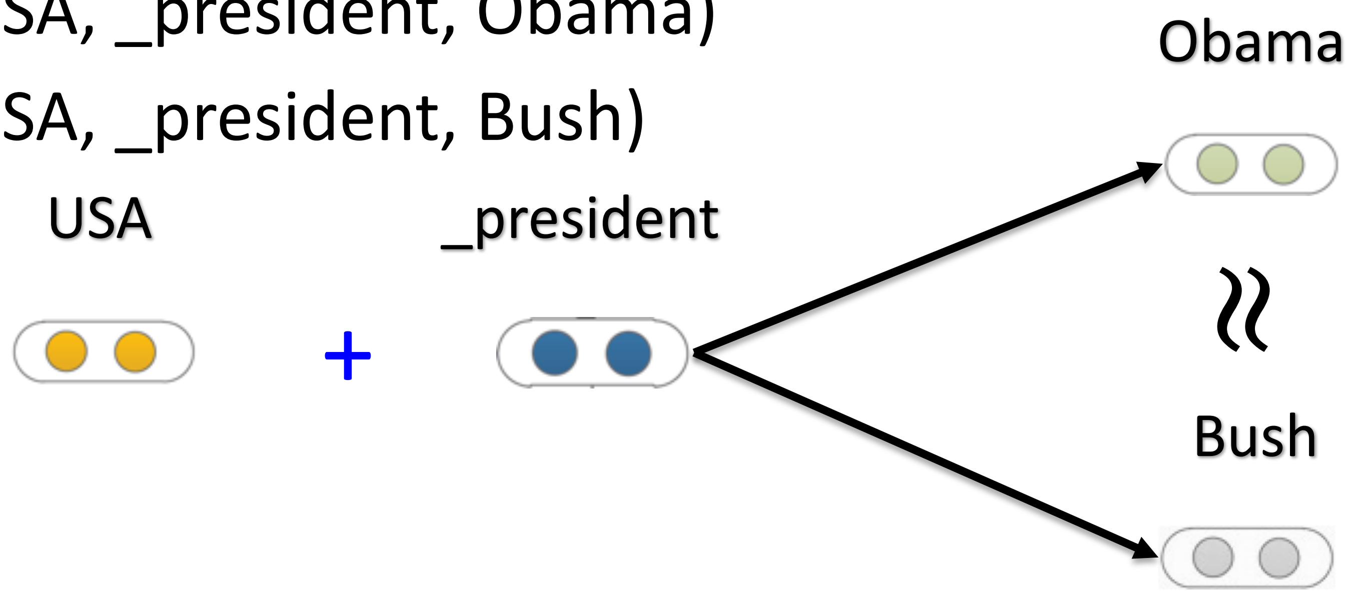
2. How to deal with **relation paths** for KA?

3. How to **incorporate KG with external info** for KA?

4. How to model **noisy KG with confidence** for KA?

Challenge I: Complicated Relations

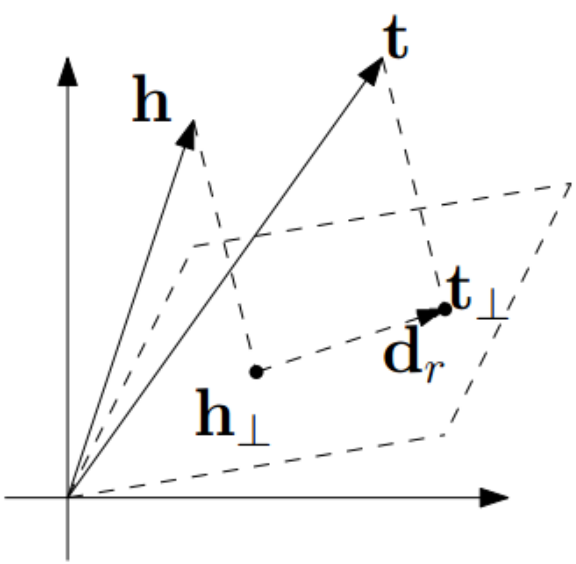
- TransE cannot well deal with complicated 1-to-N, N-to-1 and N-to-N relations
- 1-to-N relation example
  - (USA, _president, Obama)
  - (USA, _president, Bush)



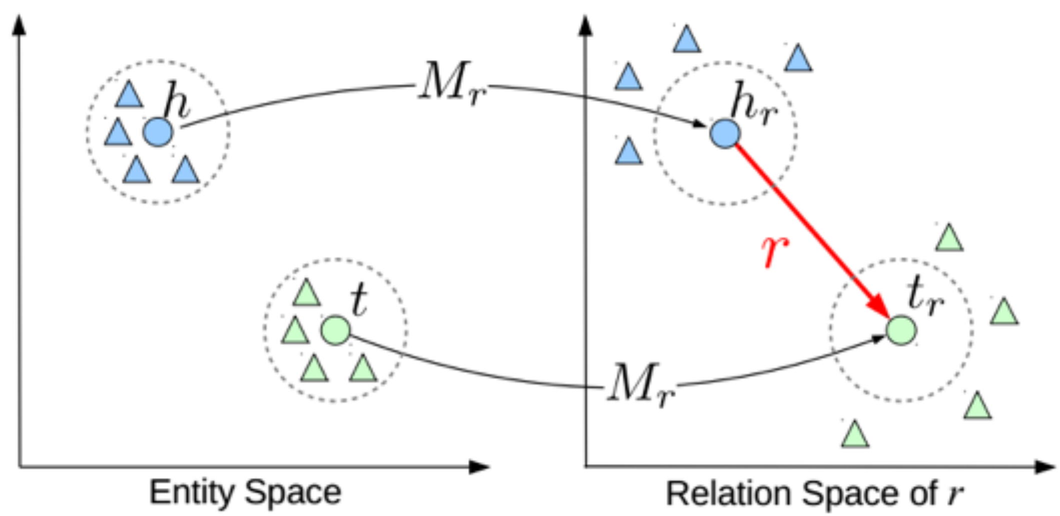
Problem: TransE tends to lose discriminability of entities due to complicated relations

Complex Relations

- Build relation-specific entity embeddings



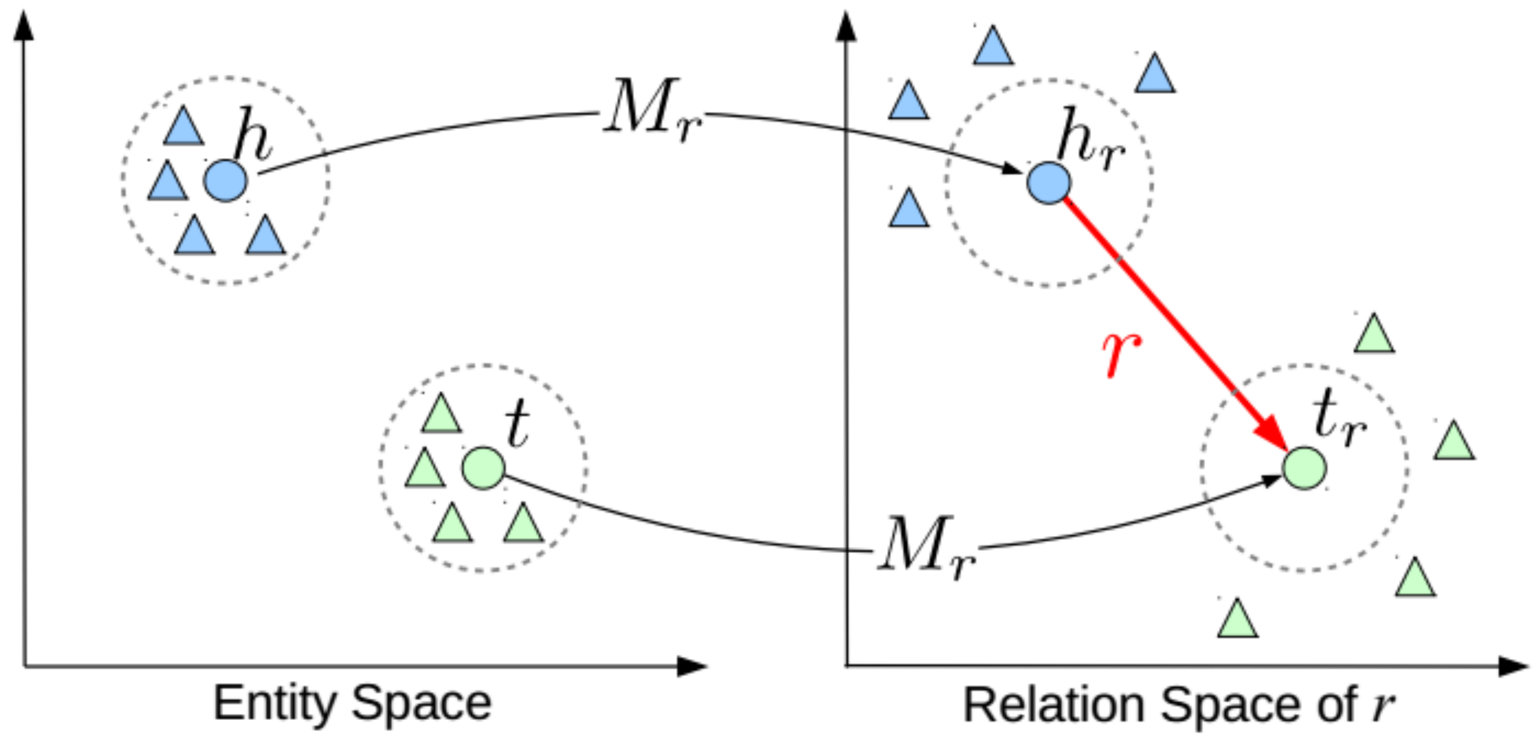
TransH



TransR

Wang, et al. (2014). Knowledge graph embedding by translating on hyperplanes. AAAI.
Lin, et al. (2015). Learning entity and relation embeddings for knowledge graph completion. AAAI.

TransR



$$\mathbf{h}_r = \mathbf{h}\mathbf{M}_r, \quad \mathbf{t}_r = \mathbf{t}\mathbf{M}_r$$

$$f_r(h,t) = |\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r|_{L1/L2}$$

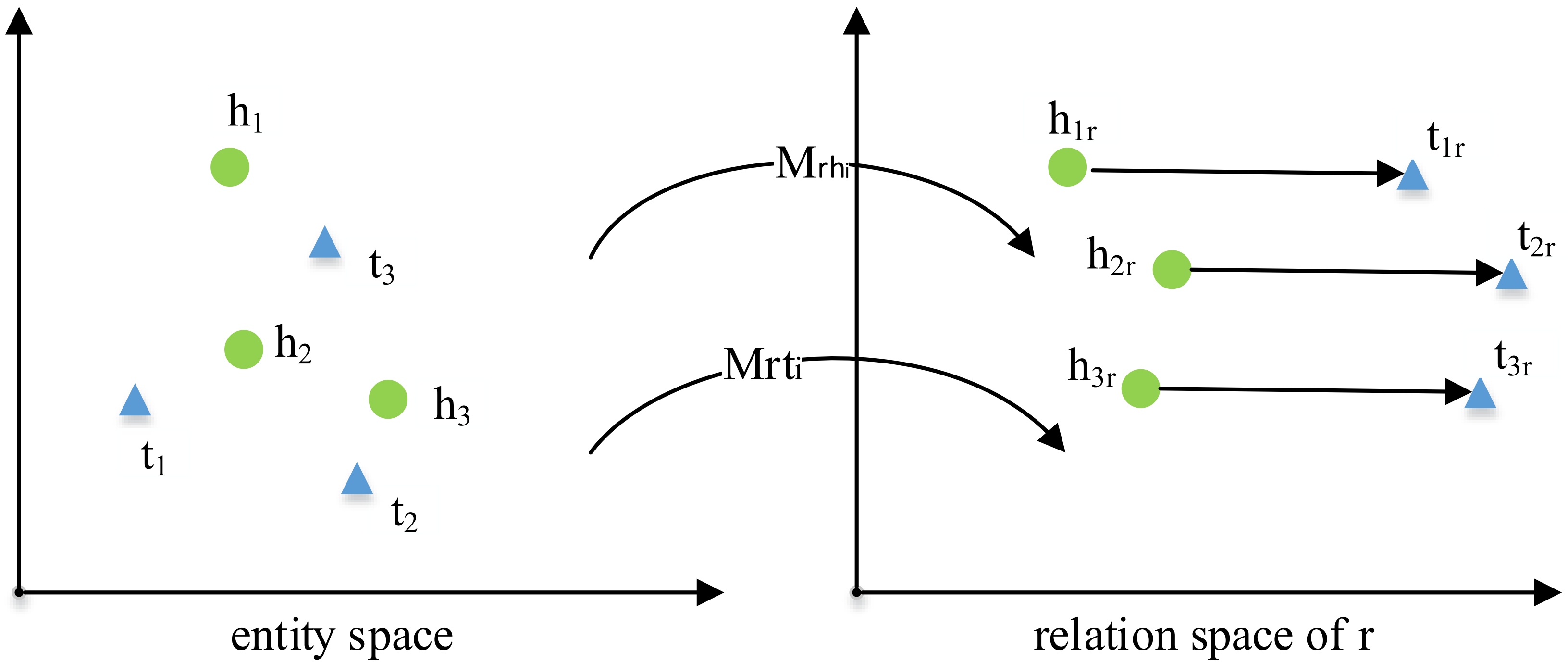
Evaluation Results

- TransR significantly improves prediction performance
- Comparison: Entity prediction on WordNet and Freebase

Data Sets	WN18				FB15K			
Metric	Mean Rank		Hits@10 (%)		Mean Rank		Hits@10 (%)	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickel, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) (Bordes et al. 2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif) (Wang et al. 2014)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern) (Wang et al. 2014)	401	388	73.0	82.3	212	87	45.7	64.4
TransR (unif)	232	219	78.3	91.7	226	78	43.8	65.5
TransR (bern)	238	225	**79.8**	92.0	**198**	77	48.2	68.7
CTransR (unif)	243	230	78.9	**92.3**	233	82	44	66.3
CTransR (bern)	**231**	**218**	79.4	**92.3**	199	**75**	**48.4**	**70.2**

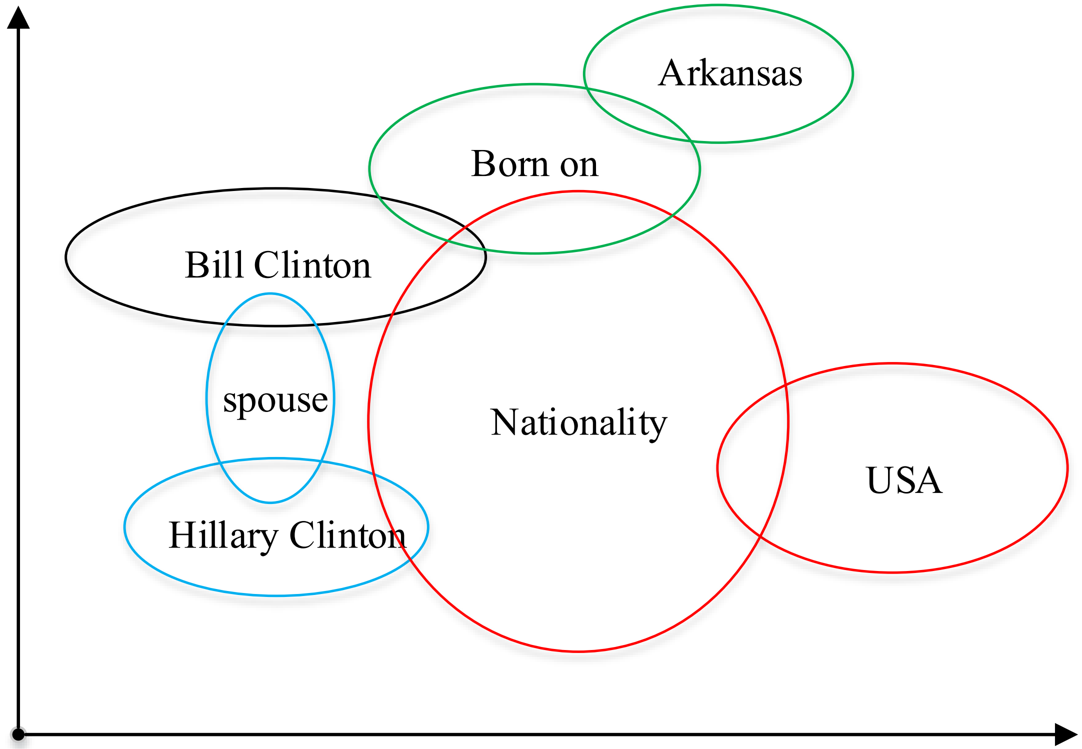
Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, Xuan Zhu. **Learning Entity and Relation Embeddings for Knowledge Graph Completion**. AAAI 2015.

TransD



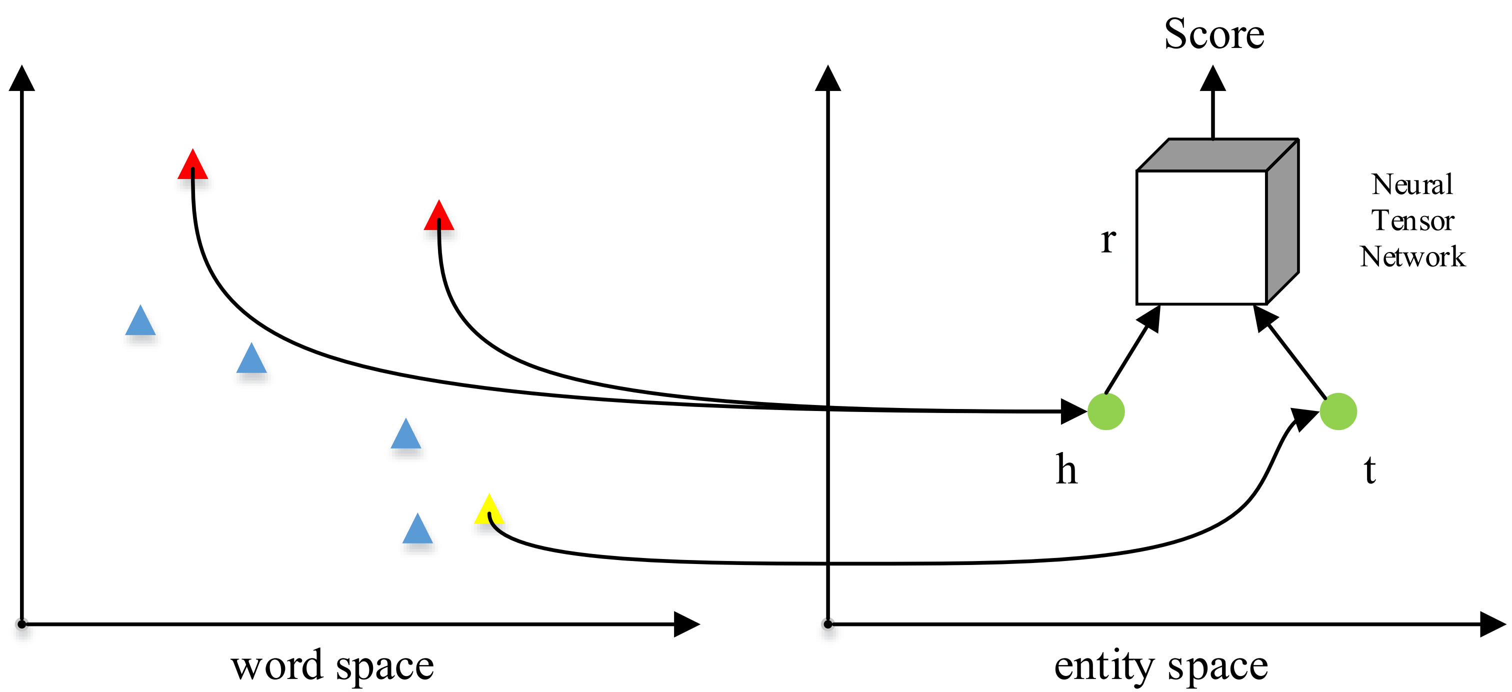
- TransD builds relation-specific entity embeddings with projection matrices related not only to relation but also head/tail entities.

KG2E



- To specially consider the (un)certainties of entities and relations , KG2E models each relations and entities with Gaussian Distribution.

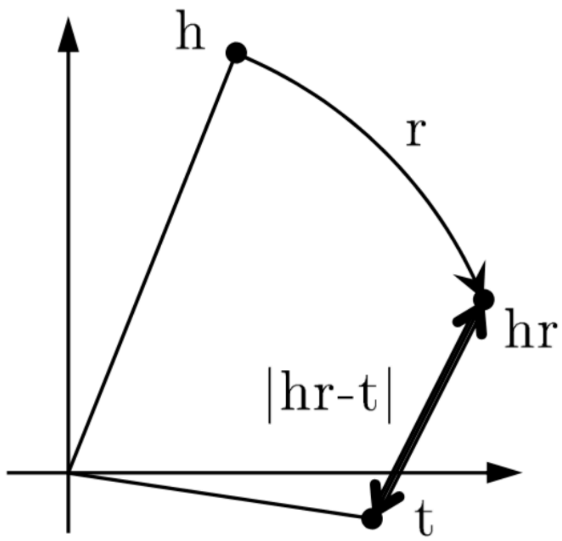
NTN



- NTN models KG with a Neural Tensor Network and represents entities via word vectors.

$$f_r(h,t) = \mathbf{u}_r^\top \tanh(\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{M}_{r,1}\mathbf{h} + \mathbf{M}_{r,2}\mathbf{t} + \mathbf{b}_r)$$

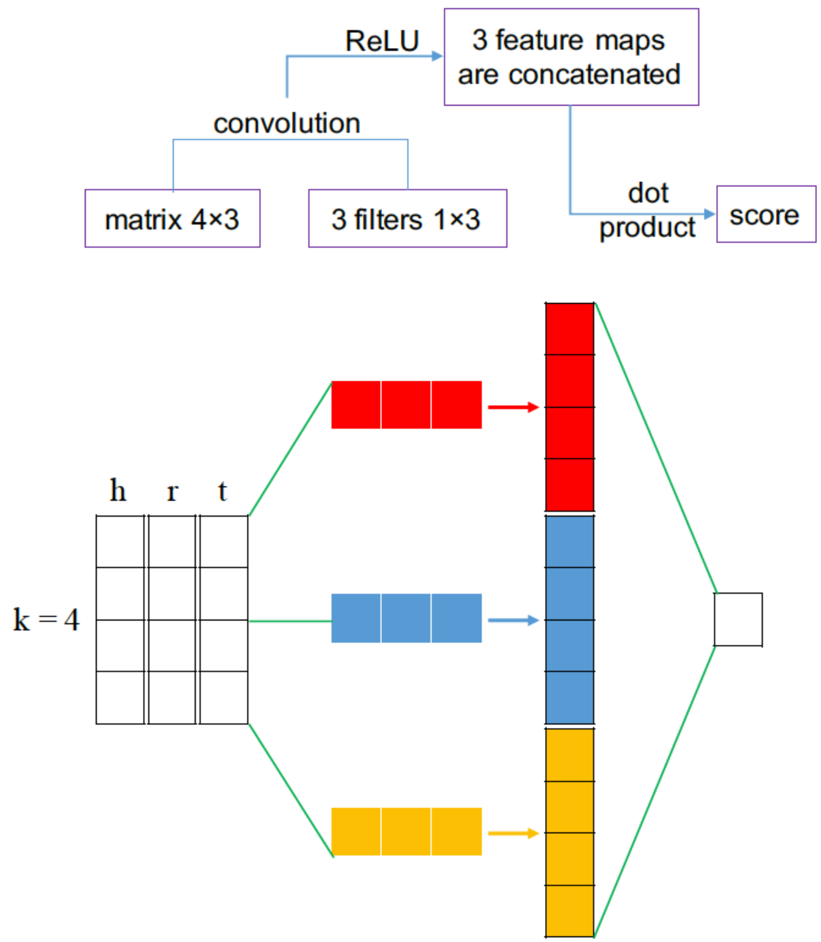
RotateE



- Model each relation as a rotation from the head entity to the tail entity in the vector space

$$f_r(h,t) = |\mathbf{h} \circ \mathbf{r} - \mathbf{t}|_{L1/L2}$$

ConvKB



- Model relational facts with CNN

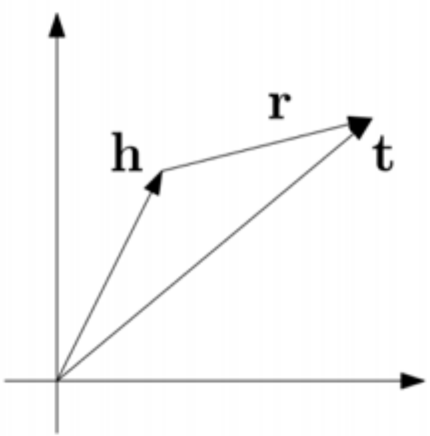
Other Models

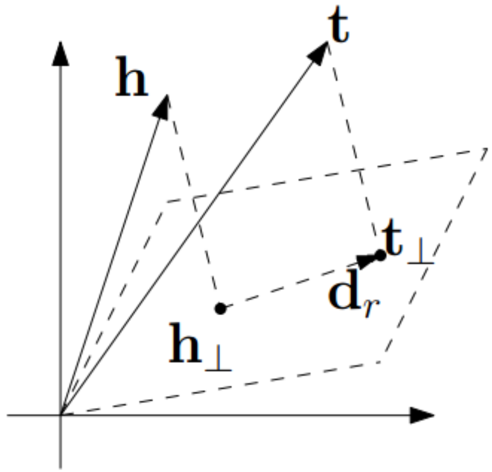
- **TranSparse** uses sparse projection matrices to deal with the issue of entities and relations are heterogeneous and unbalanced
- **Holographic Embeddings** uses the circular correlation to combine the expressive power of the tensor product with the efficiency and simplicity of TransE.
- **Complex Embeddings** employs eigenvalue decomposition model which makes use of complex valued embeddings.

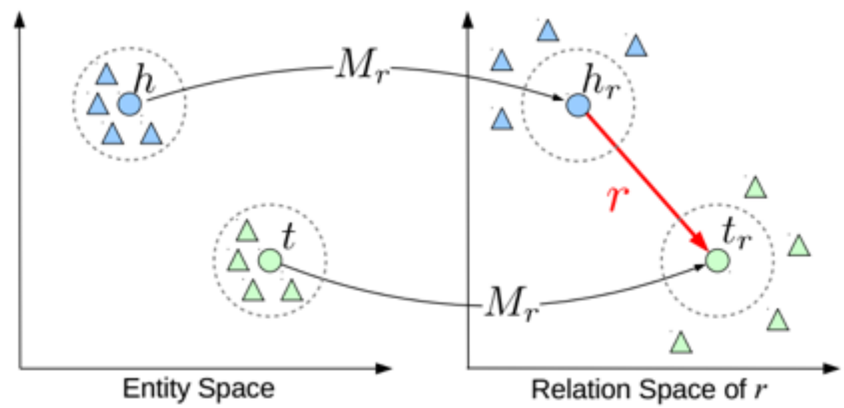
Ji, et al. (2016) Knowledge Graph Completion with Adaptive Sparse Transfer Matrix. AAAI.
Nichkel, et al. (2015) Holographic Embeddings of Knowledge Graphs. Arxiv.
Trouillon, et al. (2016) Complex embeddings for simple link prediction. Arxir.

Summary

- TransE is too simple to handle complex relations well
  - 1-N, N-1, N-N
  - TransA, TransD, TransE, TransG, TransH, TransR, KG2E, TranSparse, Hole

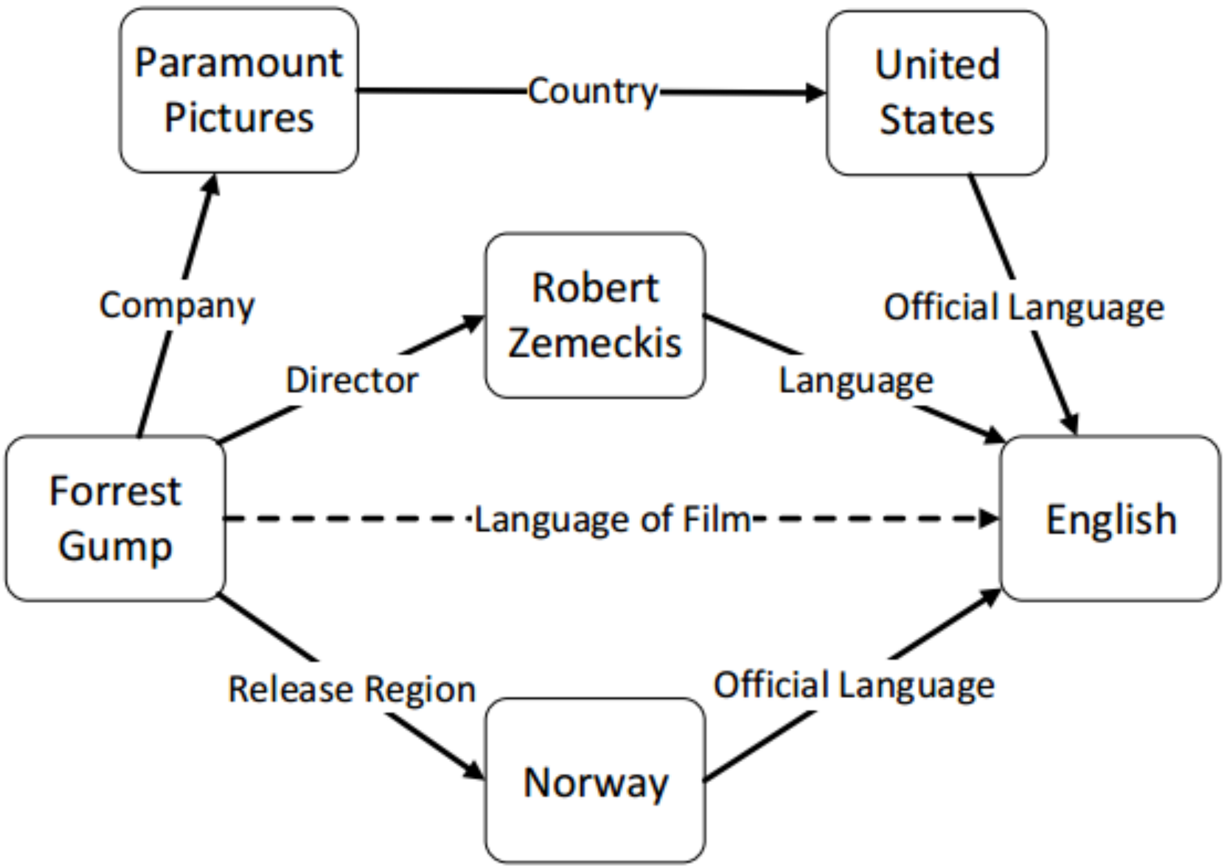






Challenge II: Relation Paths

- Relation paths can be used to predict relations between two entities



Relation Paths for Relation Prediction

- Path Ranking Algorithm

ID	PRA Path (Comment)
	athletePlaysForTeam
1	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{leaguePlayers}} c \xrightarrow{\text{athletePlaysForTeam}} c$ (teams with many players in the athlete's league)
2	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{leagueTeams}} c \xrightarrow{\text{teamAgainstTeam}} c$ (teams that play against many teams in the athlete's league)
	athletePlaysInLeague
3	$c \xrightarrow{\text{athletePlaysSport}} c \xrightarrow{\text{players}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (the league that players of a certain sport belong to)
4	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (popular leagues with many players)
	athletePlaysSport
5	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{athletePlaysSport}} c$ (popular sports of all the athletes)
6	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{superpartOfOrganization}} c \xrightarrow{\text{teamPlaysSport}} c$ (popular sports of a certain league)
	stadiumLocatedInCity
7	$c \xrightarrow{\text{stadiumHomeTeam}} c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the stadium with the same team)
8	$c \xrightarrow{\text{latitudeLongitude}} c \xrightarrow{\text{latitudeLongitudeOf}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the stadium with the same location)
	teamHomeStadium
9	$c \xrightarrow{\text{teamPlaysInCity}} c \xrightarrow{\text{cityStadiums}} c$ (stadiums located in the same city with the query team)
10	$c \xrightarrow{\text{teamMember}} c \xrightarrow{\text{athletePlaysForTeam}} c \xrightarrow{\text{teamHomeStadium}} c$ (home stadium of teams which share players with the query)
	teamPlaysInCity
11	$c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the team's home stadium)
12	$c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumHomeTeam}} c \xrightarrow{\text{teamPlaysInCity}} c$ (city of teams with the same home stadium as the query)
	teamPlaysInLeague
13	$c \xrightarrow{\text{teamPlaysSport}} c \xrightarrow{\text{players}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (the league that the query team's members belong to)
14	$c \xrightarrow{\text{teamPlaysAgainstTeam}} c \xrightarrow{\text{teamPlaysInLeague}} c$ (the league that the query team's competing team belongs to)
	teamPlaysSport
15	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{teamPlaysSport}} c$ (sports played by many teams)
16	$c \xrightarrow{\text{teamPlaysInLeague}} c \xrightarrow{\text{leagueTeams}} c \xrightarrow{\text{teamPlaysSport}} c$ (the sport played by other teams in the league)

Lao, et al. (2011). Random walk inference and learning in a large scale knowledge base. EMNLP.

PTransE：Path-based TransE

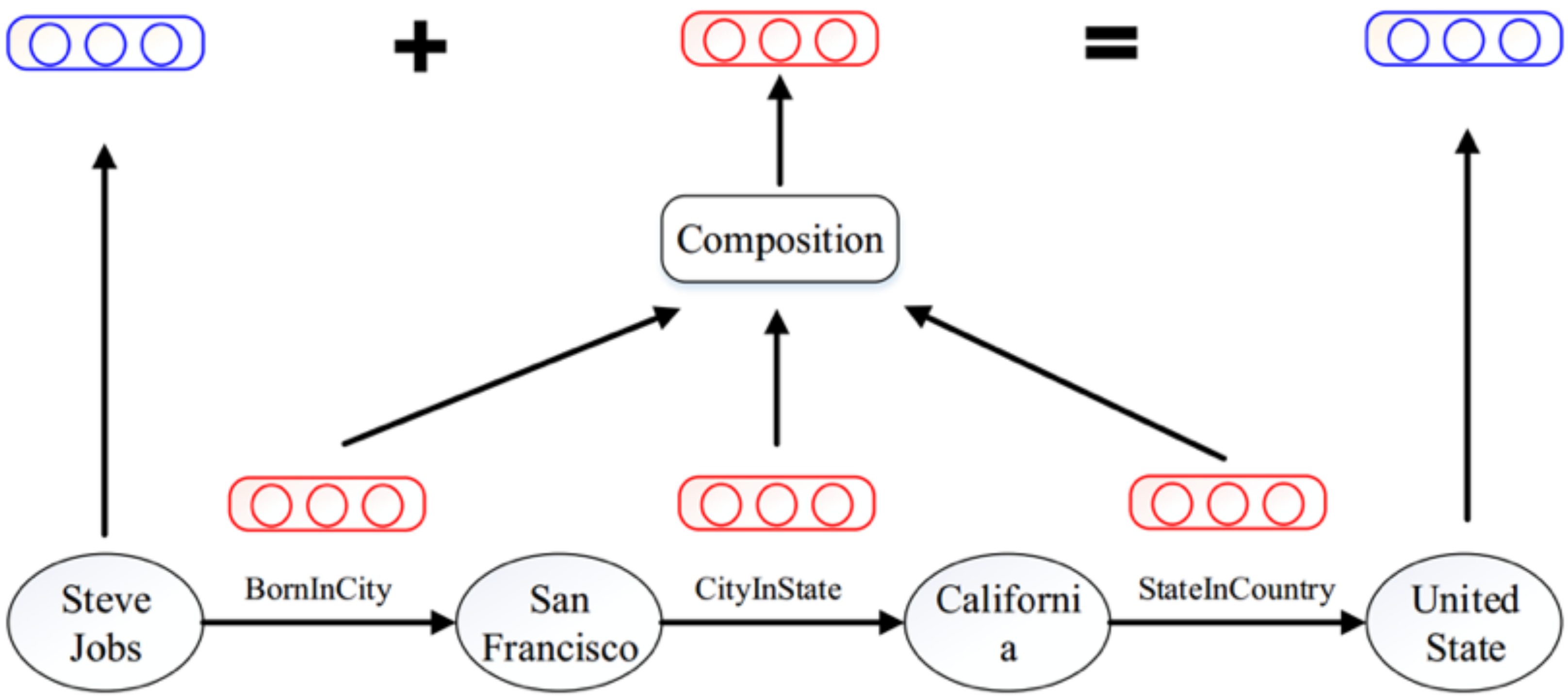
- Key Idea: Consider the information of relation paths in knowledge representation learning

	TransE	PTransE
KB	$h \xrightarrow{r} t$	$h \xrightarrow{r_1} e_1 \xrightarrow{r_2} t$
Triples	(h, r, t)	(h, r_1, e_1) (e_1, r_2, t) $(h, r_1 \circ r_2, t)$
Objectives	$\mathbf{h} + \mathbf{r} = \mathbf{t}$	$\mathbf{h} + \mathbf{r}_1 = \mathbf{e}_1$ $\mathbf{e}_1 + \mathbf{r}_2 = \mathbf{t}$ $\mathbf{h} + (\mathbf{r}_1 \circ \mathbf{r}_2) = \mathbf{t}$

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, Song Liu. **Modeling Relation Paths for Representation Learning of Knowledge Bases**. EMNLP 2015.

PTransE ： Path-based TransE

- Path Representation: Semantic composition of relations within the path



Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, Song Liu. **Modeling Relation Paths for Representation Learning of Knowledge Bases**. EMNLP 2015.

Evaluation Results

- PTransE significantly improves prediction performance
- Comparison: Relation prediction on Freebase

Metric	Mean Rank		Hits@1 (%)	
	Raw	Filter	Raw	Filter
TransE	2.8	2.5	65.1	$\boxed{84.3}$
+Rev	2.6	2.3	67.1	86.7
+Rev+Path	2.4	1.9	65.2	89.0
PTransE (ADD, 2-step)	**1.7**	**1.2**	69.5	93.6
-TransE	135.8	135.3	51.4	78.0
-Path	2.0	1.6	**69.7**	89.0
PTransE (MUL, 2-step)	2.5	2.0	66.3	89.0
PTransE (RNN, 2-step)	1.9	1.4	68.3	93.2
PTransE (ADD, 3-step)	1.8	1.4	68.5	$\boxed{\mathbf{94.0}}$

+10%

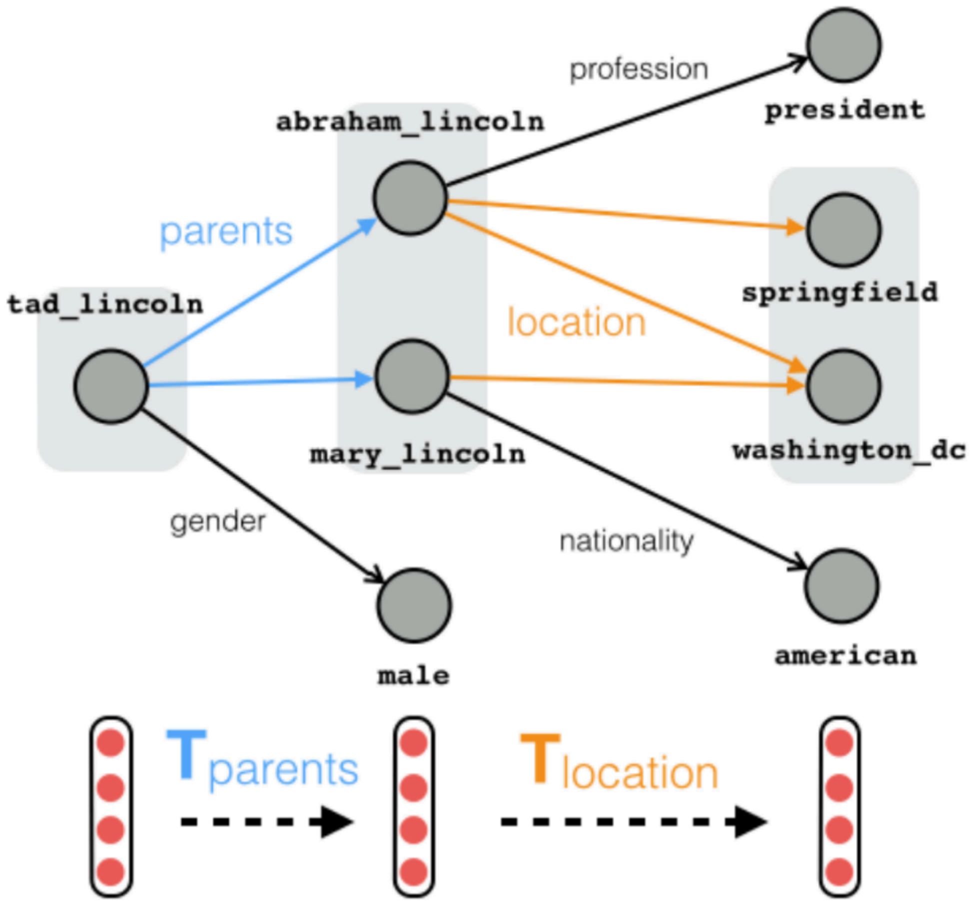
Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, Song Liu. **Modeling Relation Paths for Representation Learning of Knowledge Bases**. EMNLP 2015.

Examples: Relation Composition

Relation1	**/people/person/place_of_birth**
Relation2	**/location/administrative_division/country**
1	/people/person/nationality
2	/people/person/places_lived./people/place_lived/location
3	/people/person/place_of_birth
4	/music/artist/origin
5	/olympics/olympic_athlete_affiliation/country
6	/government/politician/government_positions_held
7	/base/popstra/vacation_choice/location
8	/people/deceased_person/place_of_death
9	/government/political_appointer/appointees
10	/location/administrative_division/country

Path Query Answering

- Query
  - Where are Tad Lincoln's parents located



Guu, et al. (2015). Traversing Knowledge Graphs in Vector Space. EMNLP.

Summary

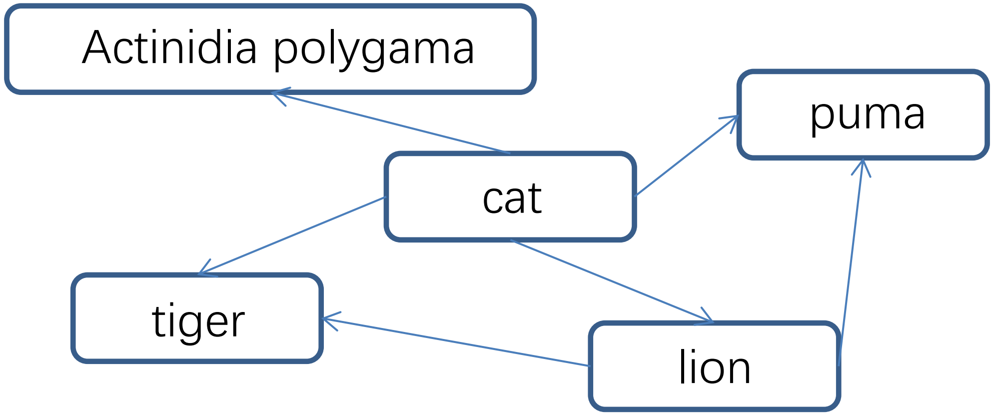
- Relation paths contain rich inference patterns about knowledge

- More complex inference patterns should be taken into consideration

(Obama, _president, USA)

(Obama, _is, American)

Challenge III: Rich Information



The domestic cat (Latin: Felis catus) or the feral cat (Latin: Felis silvestris catus) is a small, typically furry, carnivorous mammal. They are often called house cats when kept as indoor pets or simply cats when there is no need to distinguish them from other felids and felines

Entity Descriptions

- KG contains rich information besides network structure

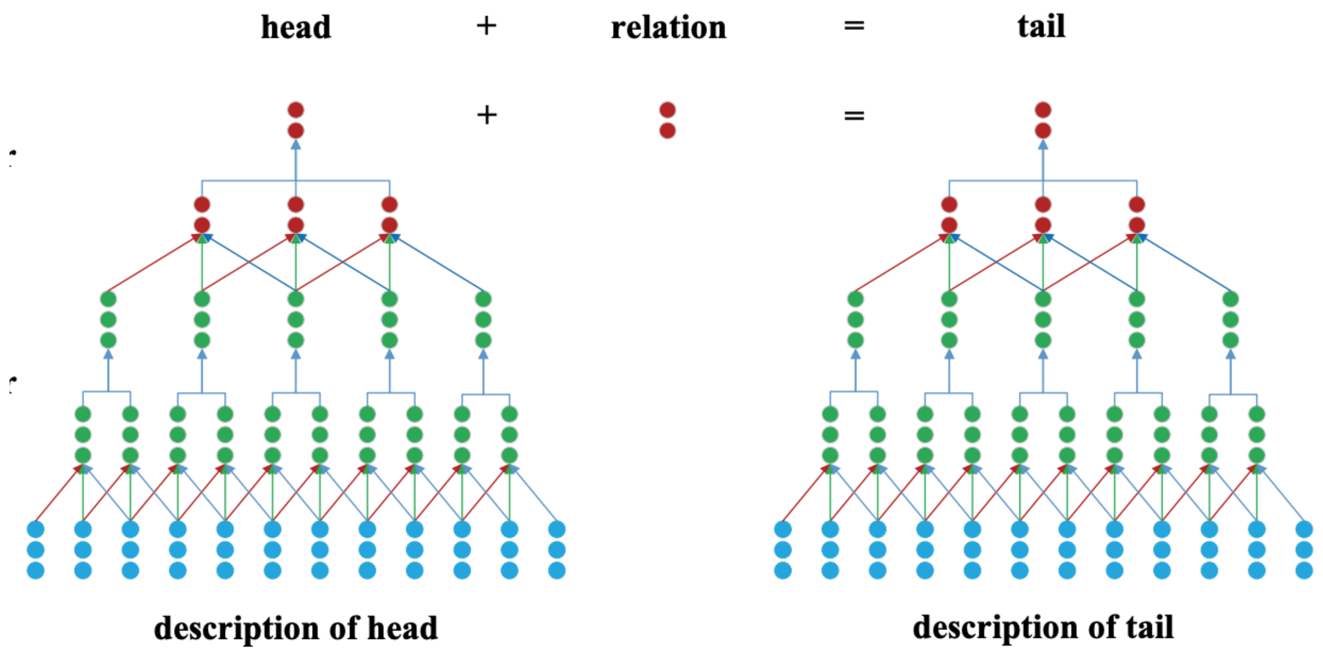
(***William Shakespeare***, book/author/works_written, ***Romeo and Juliet***)

William Shakespeare was an English poet, playwright, and actor, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. ...

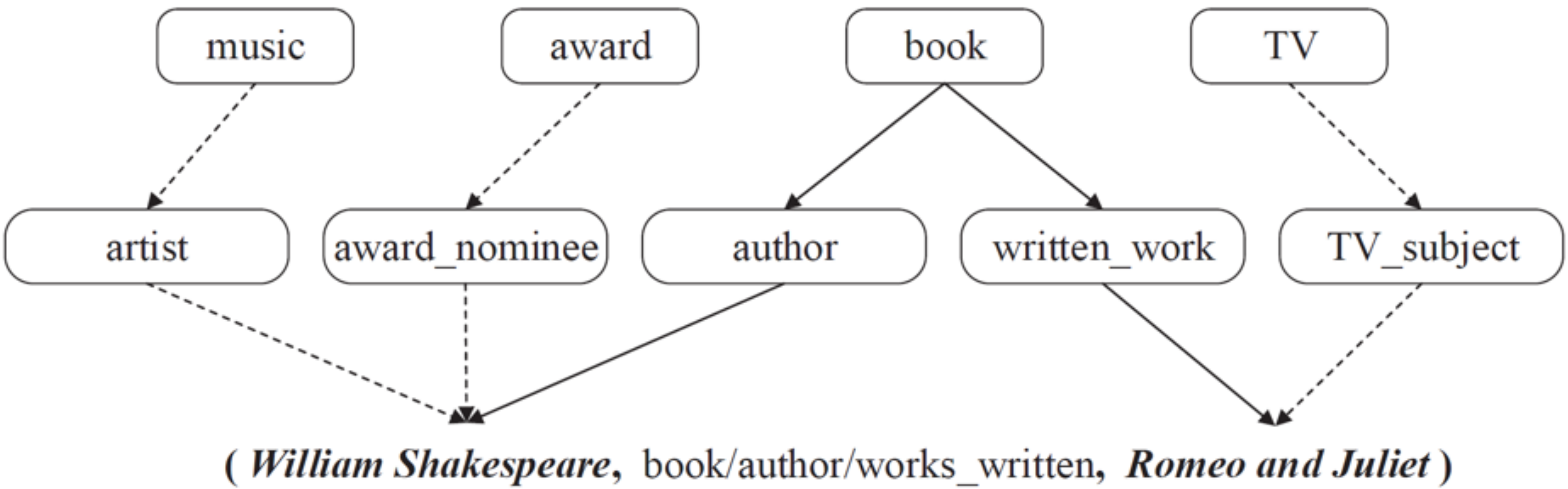
Romeo and Juliet is a tragedy written by William Shakespeare early in his career about two young star-crossed lovers whose deaths ultimately reconcile their feuding families. ...

DKRL: Incorporate Entity Description

- Entity description has rich information about entities
  - DKRL is a combination of conventional models and entity description. It aligns text space and knowledge space with a description encoder.

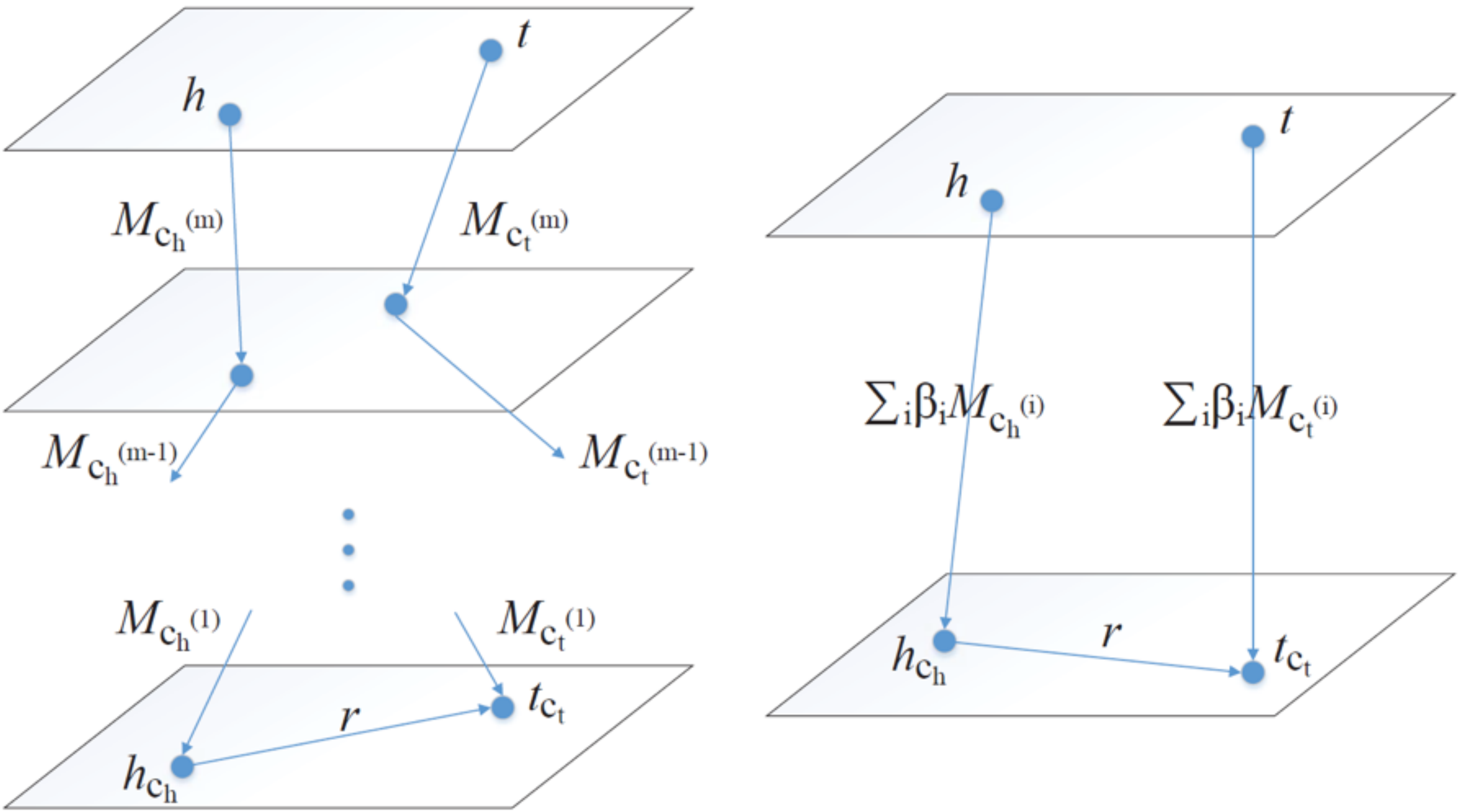


Entity Types



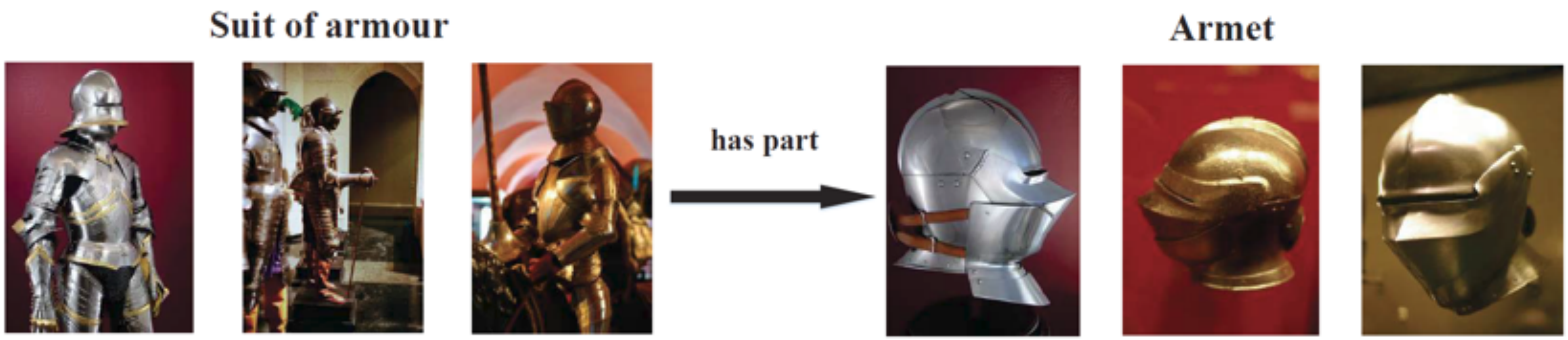
Xie, et al. (2016). Representation Learning of Knowledge Graphs with Hierarchical Types. IJCAI.

TKRL: Incorporate Entity Types



(a) RHE (b) WHE

Visual Information in Images

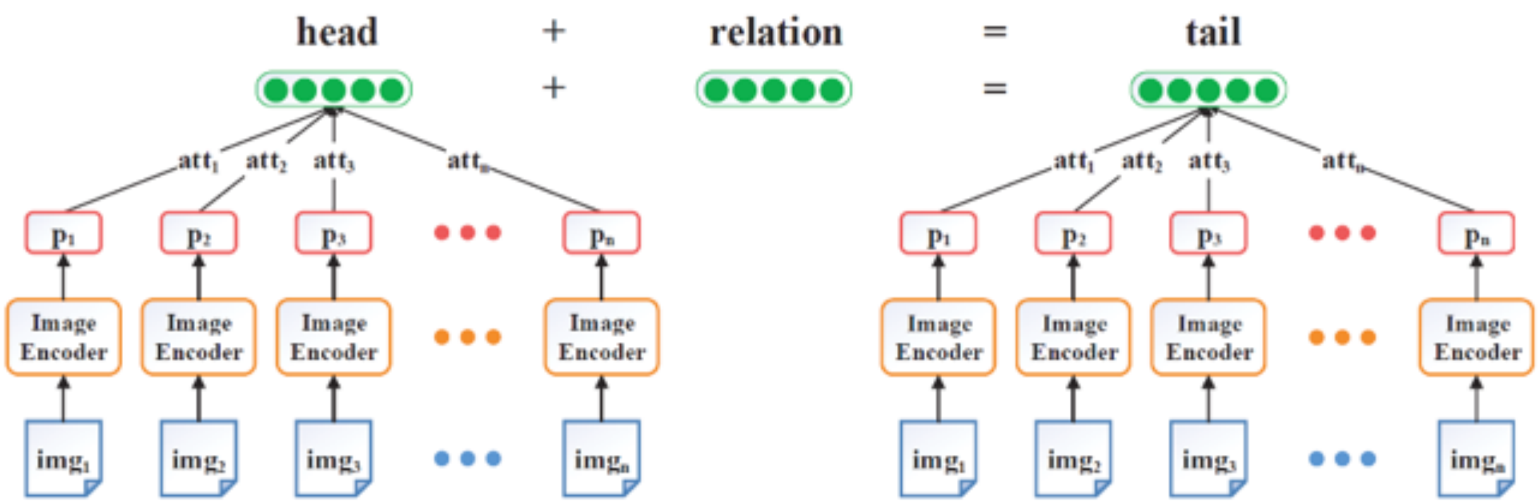


(*Suit of armour*, has_part, *Armet*)

Xie, et al. (2017). Image-embodied Knowledge Representation Learning. IJCAI.

IKRL: Introduce Entity Images

- Images are everywhere and images have rich information
- IKRL takes entity images as inputs, and with the image encoder (representation module and projection module), gets the entity embeddings



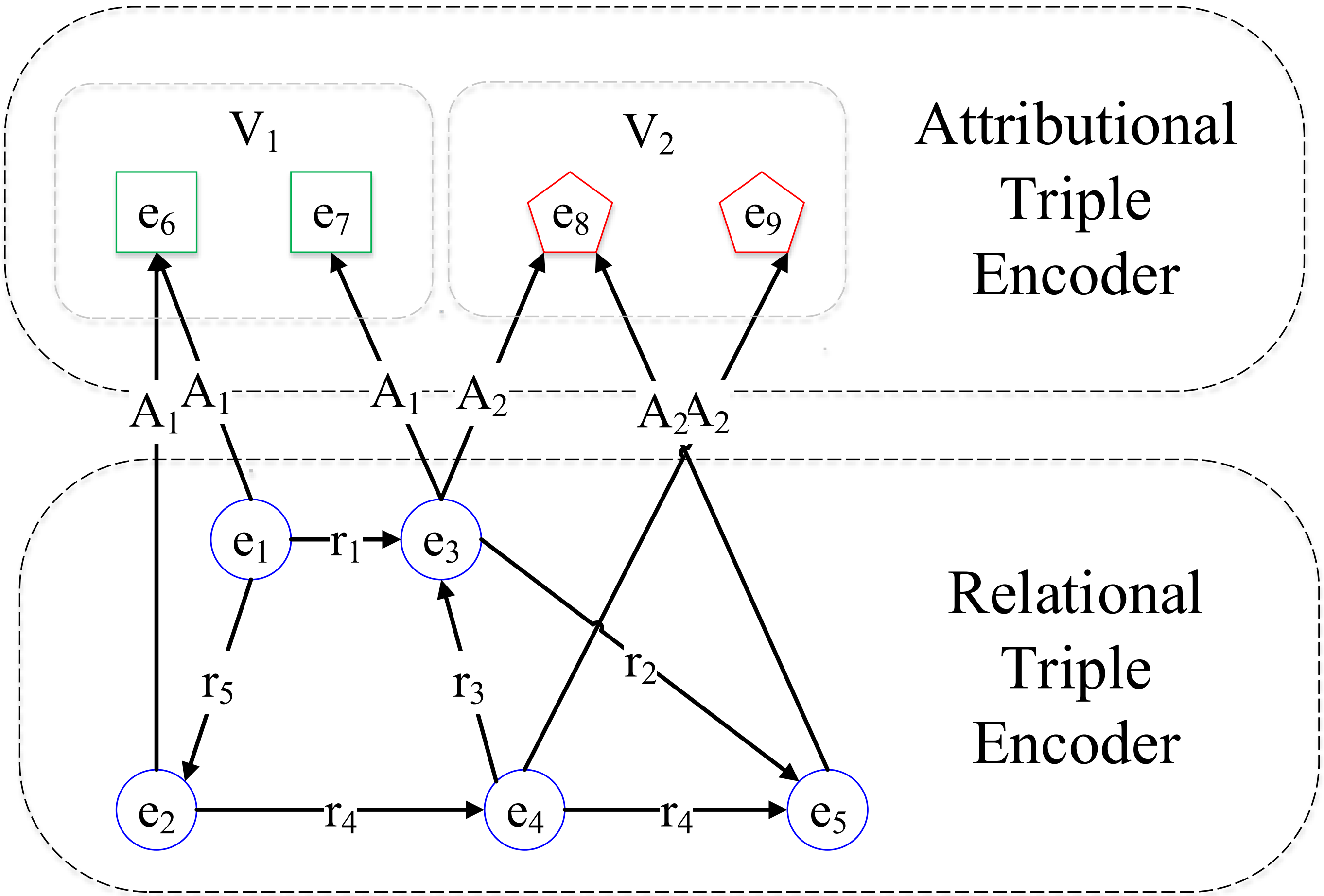
Entity Attributes

- Relational Facts in Knowledge Graph
  - Entity Attributes
  - Relations between Entities

Relation Type	Relation	E_t	E_h
Attributes	nationality	1.05	1,551.90
	gender	1.00	637,333.33
	ethnicity	1.12	41.52
	religion	1.09	107.40
Relations	parents	1.58	1.67
	capital	1.29	1.42
	author	1.02	2.17
	founder	1.37	1.31

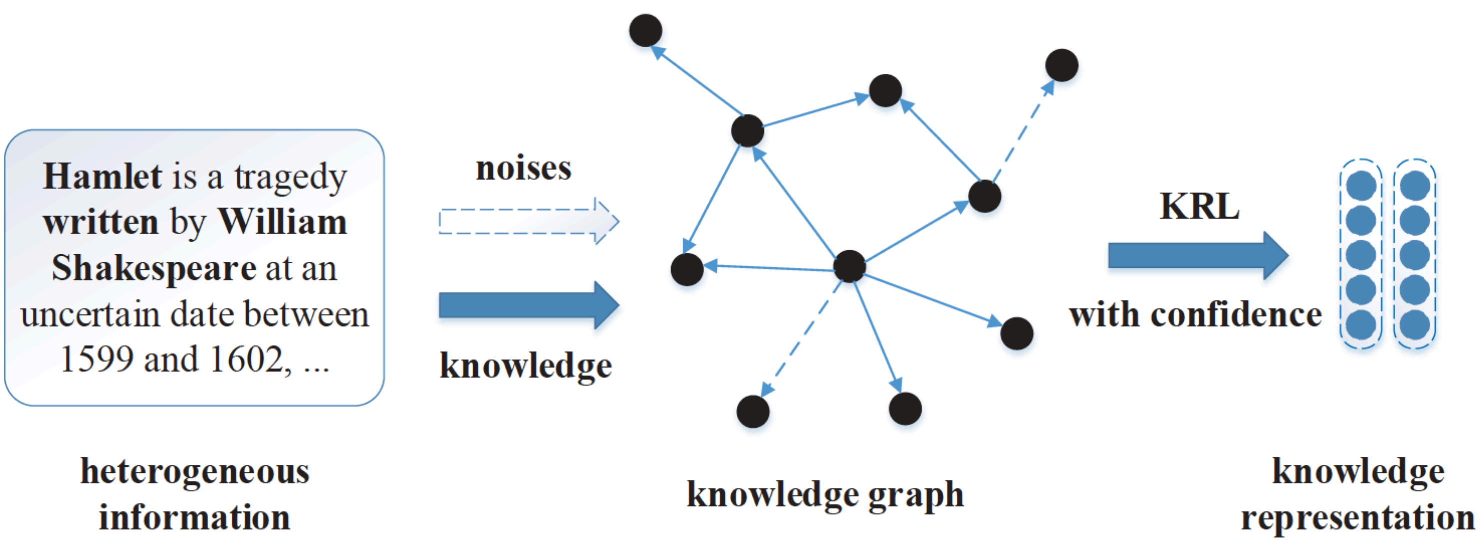
KR-EAR

- Relational Triple Encoder
- Attributional Triple Encoder



Challenge IV: KG Noise

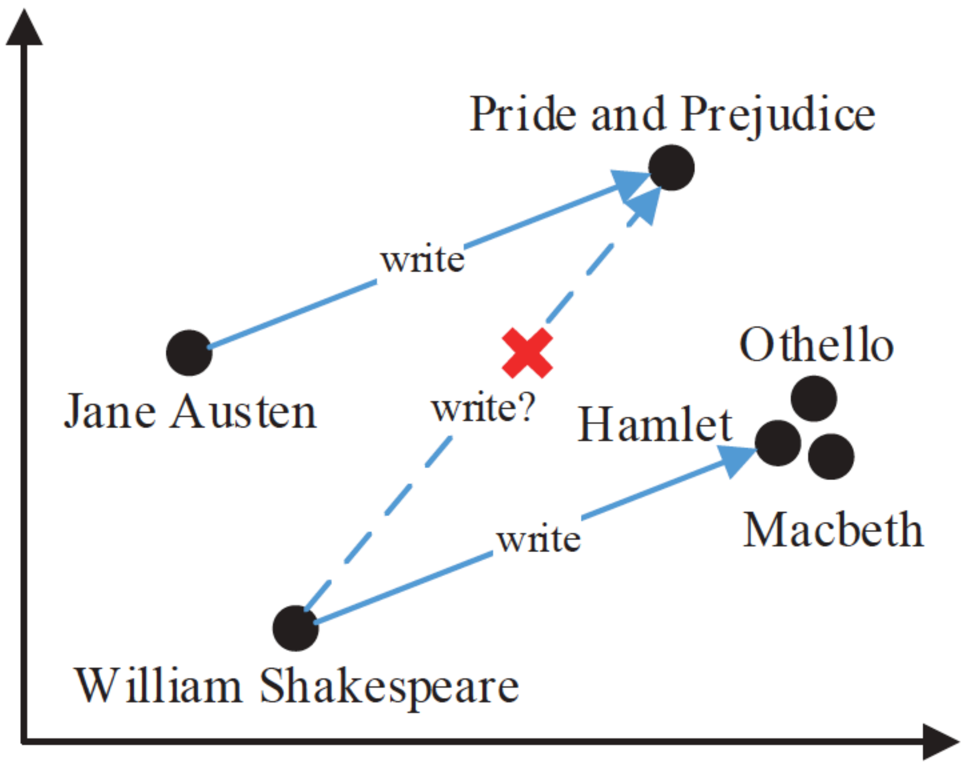
- Automatic mechanism and crowdsourcing lead to noise in KG
  - relation extraction model on benchmark achieves only around 60% precision when the recall is 20%



Xie, et al. (2018). Does William Shakespeare REALLY Write Hamlet? Knowledge RepresentationLearning with Confidence. AAAI.

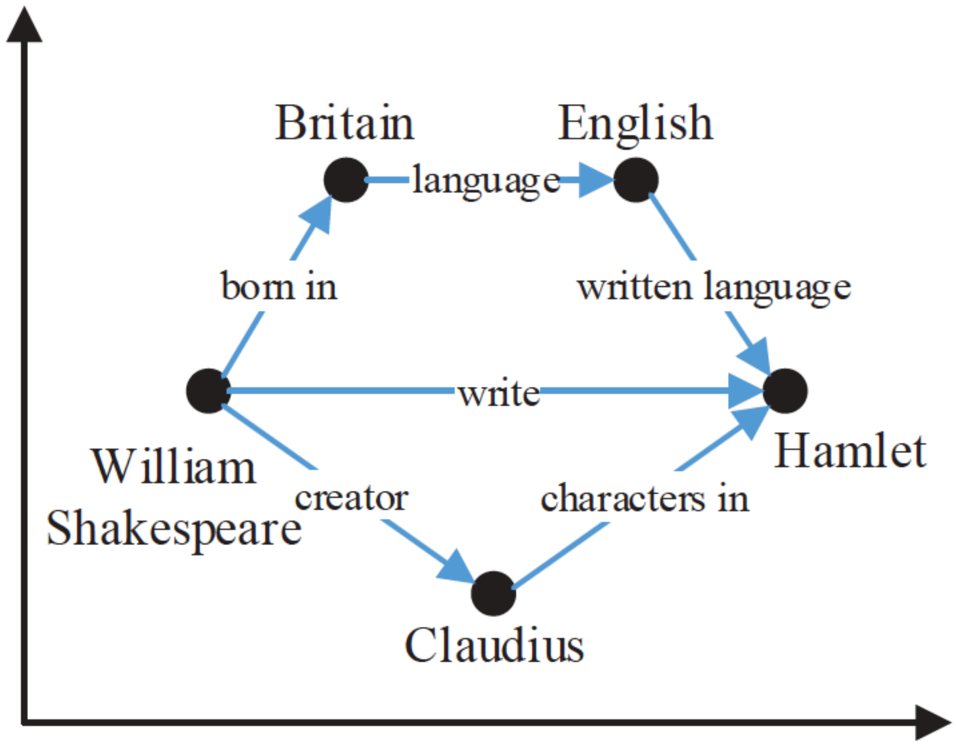
Confidence-aware KRL

- Local triple confidence



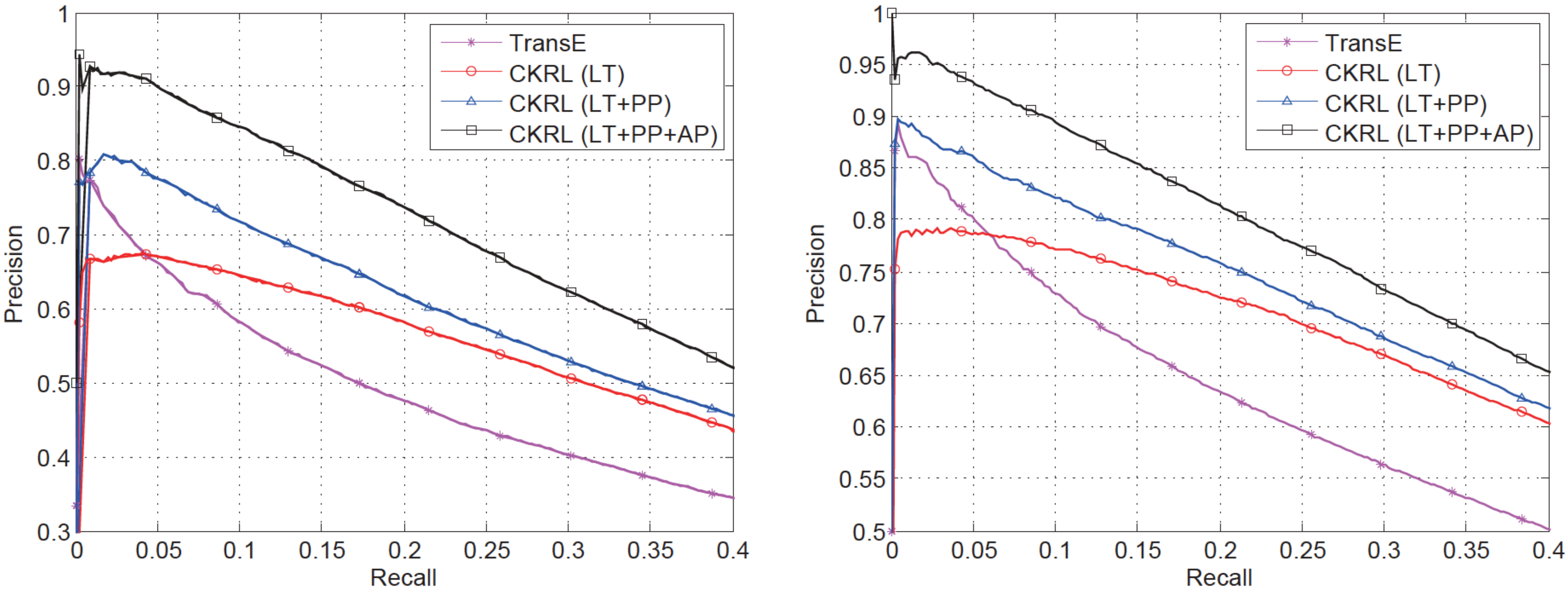
Confidence-aware KRL

- Global path confidence



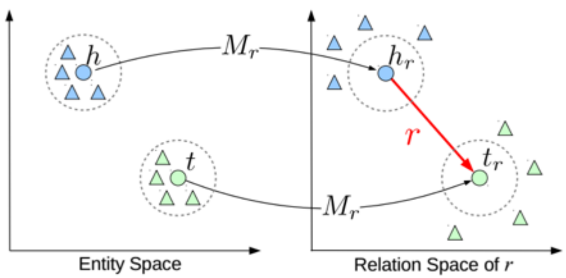
Noise Detection

- CKRL outperform TransE in different proportions of noise

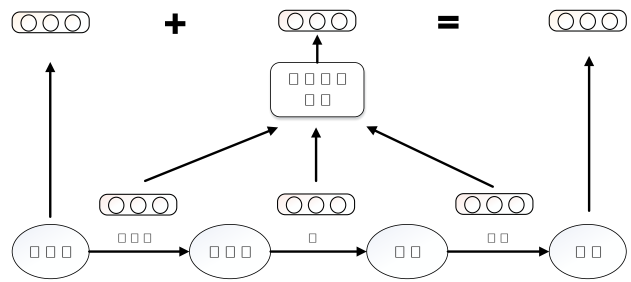


Summary

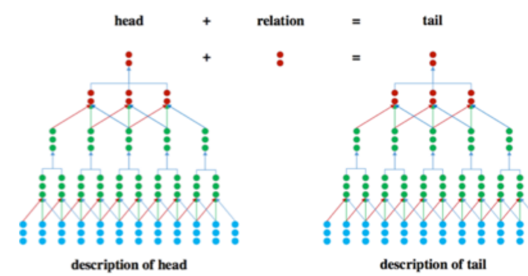
- Incorporate Complex Information in KRL



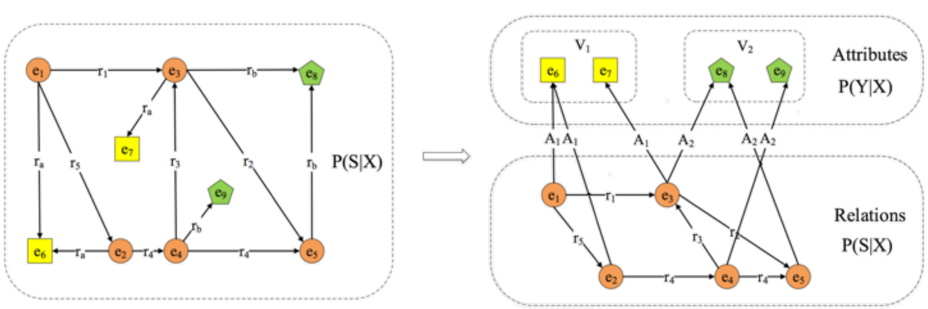
+Complex relations
TransR (AAAI 2015)



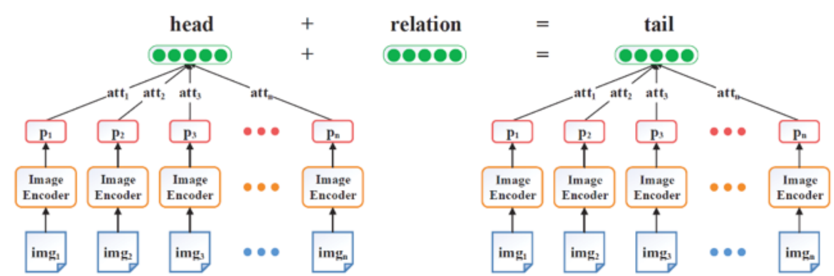
+Relation paths
PTransE (EMNLP 2015)



+Entity description
DKRL (AAAI 2016)



+Attributes and relations
KR-EAR (IJCAI 2016)



+Entity images
IKRL (IJCAI 2017)

Knowledge Representation Learning Paper List

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.

Knowledge Acquisition (KA)

- Goal: Automatically Extract knowledge from Web

Knowledge Acquisition

Relation Prediction

From **KG**

Relation Extraction

From **Text**

Relation Extraction

Problem Extract world knowledge from text

Founder

Bill Gates → Microsoft

Bill Gates | was the co-founder and CEO of | Microsoft

Challenges

1. How to **denoise** training data of distant supervision?

2. How to deal with **few-shot instances in** relation extraction?

3. How to extract relational facts at the **document level**?

4. How to extract **complicated event facts**?

Challenge I: Noise in Distant Supervision

- Build large-scale datasets with distant supervision

Bill Gates | Founder | Microsoft

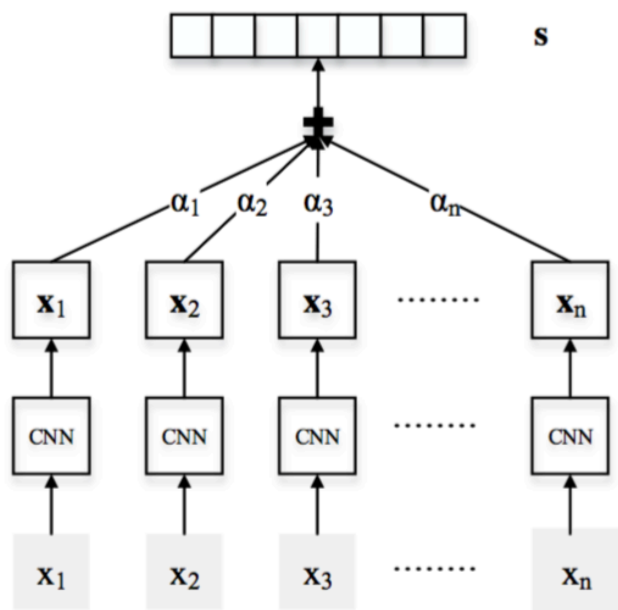
Bill Gates | was the co-founder and CEO of | Microsoft ✓

Bill Gates | announced to retire from | Microsoft ✗

Bill Gates | and Paul Allen co-founded the IT giant | Microsoft ✓

Sentence-Level Attention

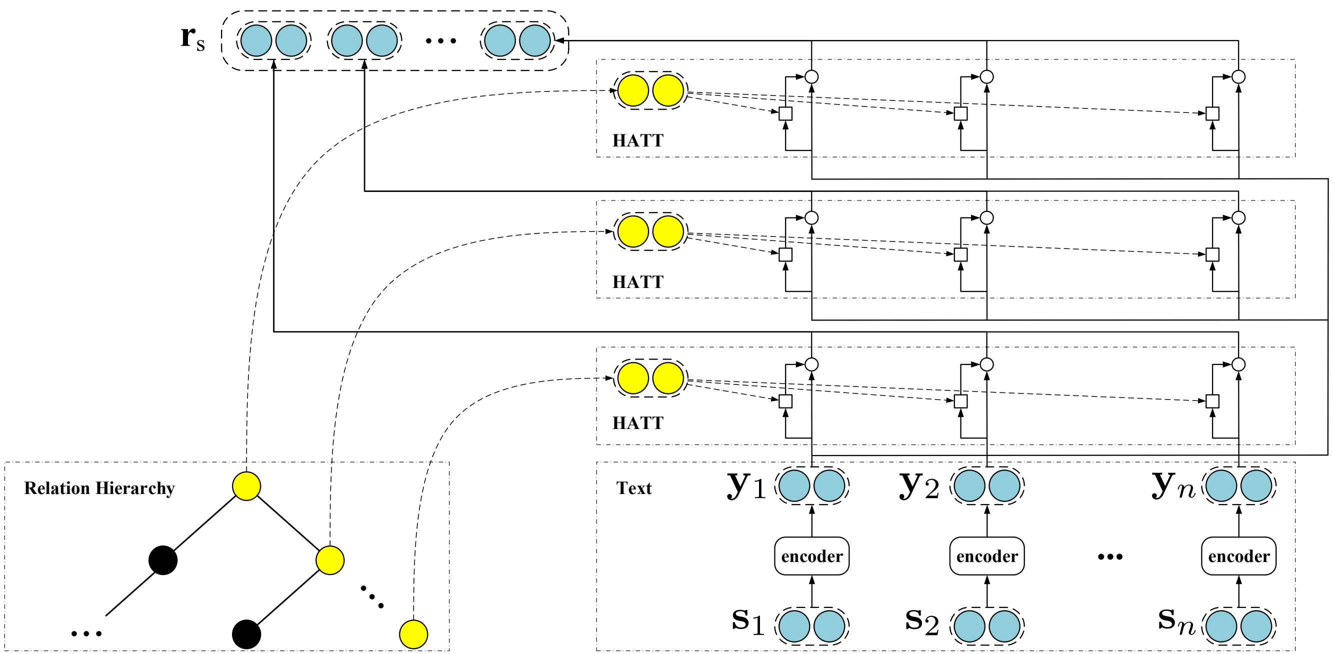
- Denoise in distant supervision
  - Multi-instance multi-label learning
  - We propose sentence-level attention mechanism to learn the credibility of automatic annotations



Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, Maosong Sun. Neural Relation Extraction with Selective Attention over Instances. ACL 2016.

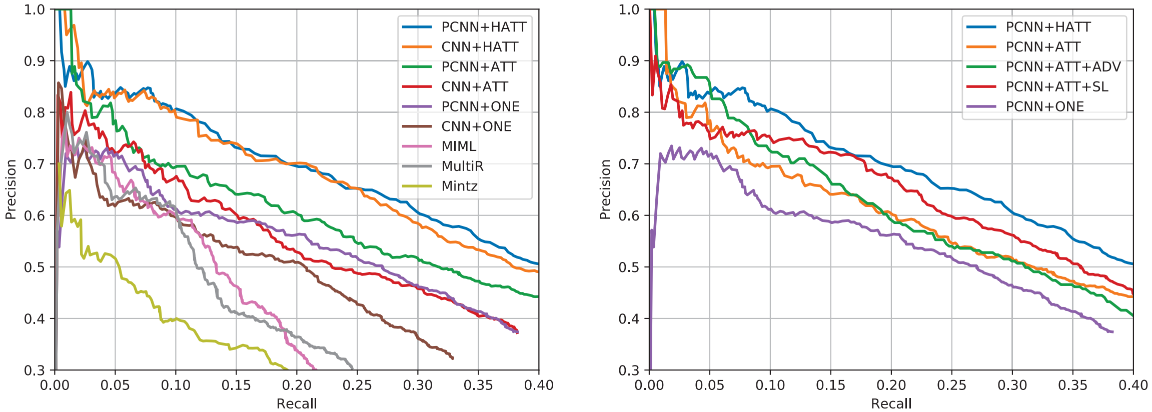
Hierarchical Attention

- Denoise in distant supervision
  - Utilize the hierarchy of relations to build selective attention
  - Relation hierarchy can both serve as an extra information and improve the performance on long-tail relations



Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun, Peng Li. Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention. EMNLP 2018.

Hierarchical Attention



Significant improvement

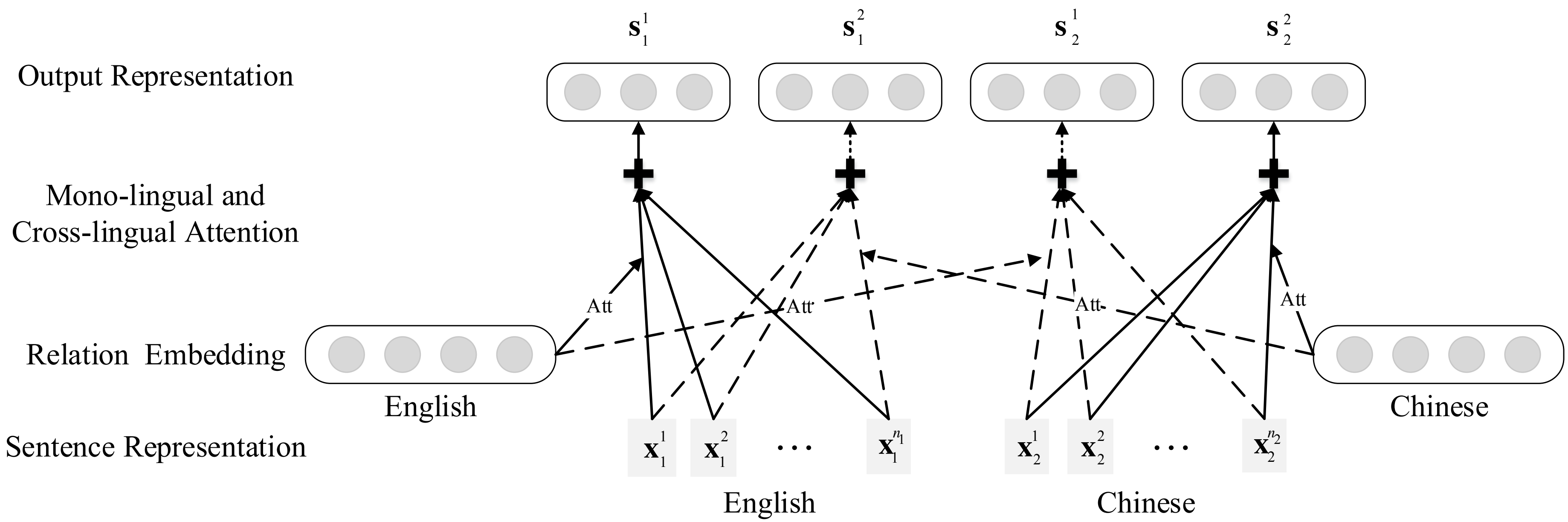
Training Instances Hits@K (Micro)		<100 10	15	20	<200 10	15	20
CNN	+ATT	<5.0	<5.0	21.1	<5.0	30.0	50.0
	+HATT	**5.3**	**36.8**	**52.6**	**40.0**	**60.0**	**70.0**
PCNN	+ATT	<5.0	10.5	47.4	33.3	43.3	66.7
	+HATT	**31.6**	**52.6**	**63.2**	**53.3**	**70.0**	**76.7**
Training Instances Hits@K (Macro)		**<100 10**	**15**	**20**	**<200 10**	**15**	**20**
CNN	+ATT	<5.0	<5.0	18.5	<5.0	16.2	33.3
	+HATT	**5.6**	**31.5**	**57.4**	**22.7**	**43.9**	**65.1**
PCNN	+ATT	<5.0	7.4	40.7	17.2	24.2	51.5
	+HATT	**29.6**	**51.9**	**61.1**	**41.4**	**60.6**	**68.2**

Improve long tail relation performance

Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun, Peng Li. Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention. EMNLP 2018.

Multilingual knowledge extraction

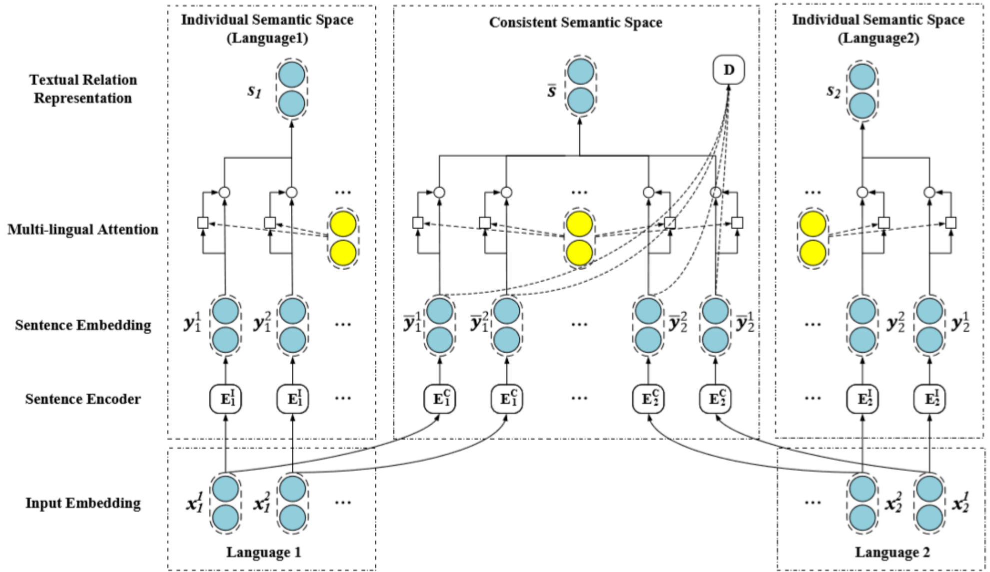
- Multilingual knowledge extraction
  – Use mono-lingual attention to utilize language-specific features
  – Use cross-lingual attention to extract unified features across languages.



Yankai Lin, Zhiyuan Liu, Maosong Sun. Neural Relation Extraction with Multi-lingual Attention. ACL 2017.

Multilingual knowledge extraction

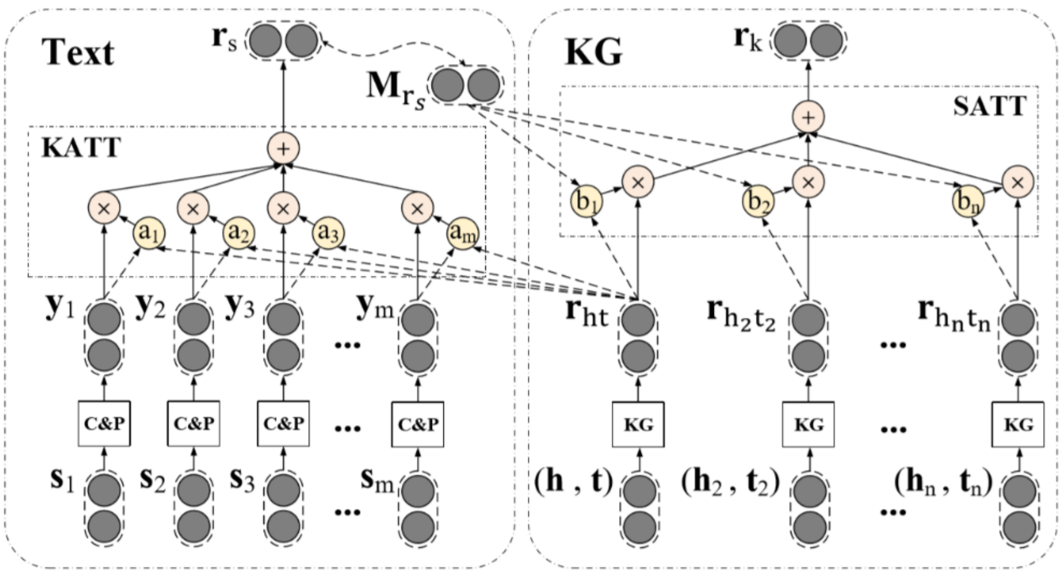
- Multilingual knowledge extraction
  - Introduce adversarial training from domain adaptation
  - Adversarial training could better disentangle language features in the representation level



Xiaozhi Wang, Xu Han, Yankai Lin, Zhiyuan Liu, Maosong Sun. Adversarial Multi-lingual Neural Relation Extraction. COLING 2018.

KG + Text Mutual Attention

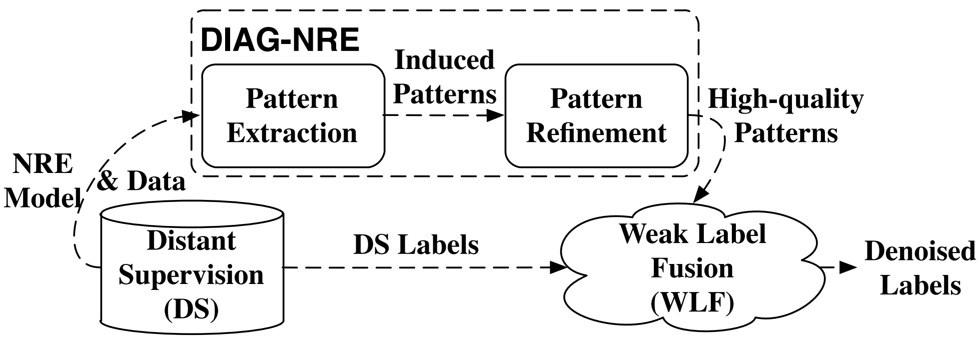
- Joint extraction with knowledge graphs
  - Representation learning and relation extraction in one semantic space
  - Use mutual attention mechanism to improve both models

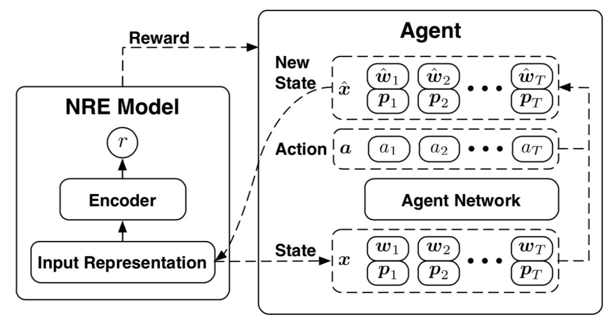


Xu Han, Zhiyuan Liu, Maosong Sun. Neural Knowledge Acquisition via Mutual Attention between Knowledge Graph and Text. AAAI 2018.

RE with Human-in-The-Loop

- Introduce human annotations
  - DIAG-NRE: Human in the loop
  - Use reinforcement learning to extract relation patterns
  - Human annotators verify those patterns
  - Litter human intervention. Significant improvement





Pattern-induction Example						
Entities	PER				CITY	
Tokens	*Joachim_Fest*	was	born	in	*Berlin*	.
Actions	0	1	0	0	*0*	1
Pattern	ENTITY1:PER PAD{1,3} born in ENTITY2:CITY					

DIAG-NRE: A Neural Pattern Diagnosis Framework for Distantly Supervised Neural Relation Extraction. ACL 2019.

Challenge II: Few-Shot Instances

- Learn how to extract relations with few samples

Supporting Set

capital of

London is the capital of the ***U.K***

capital of

Washington is the capital of the ***U.S.A***

member of

Newton served as the president of ***the Royal Society***

member of

Leibniz was a member of ***the Prussian Academy of Sciences***

birth name

Samuel Langhorne Clemens, better known by his pen name ***Mark Twain***

birth name

Alexei Maximovich Peshkov, primarily known as ***Maxim Gorky***

Query

?

Euler was elected a foreign member of ***the Royal Swedish Academy of Sciences***

Few-Shot Relation Extraction

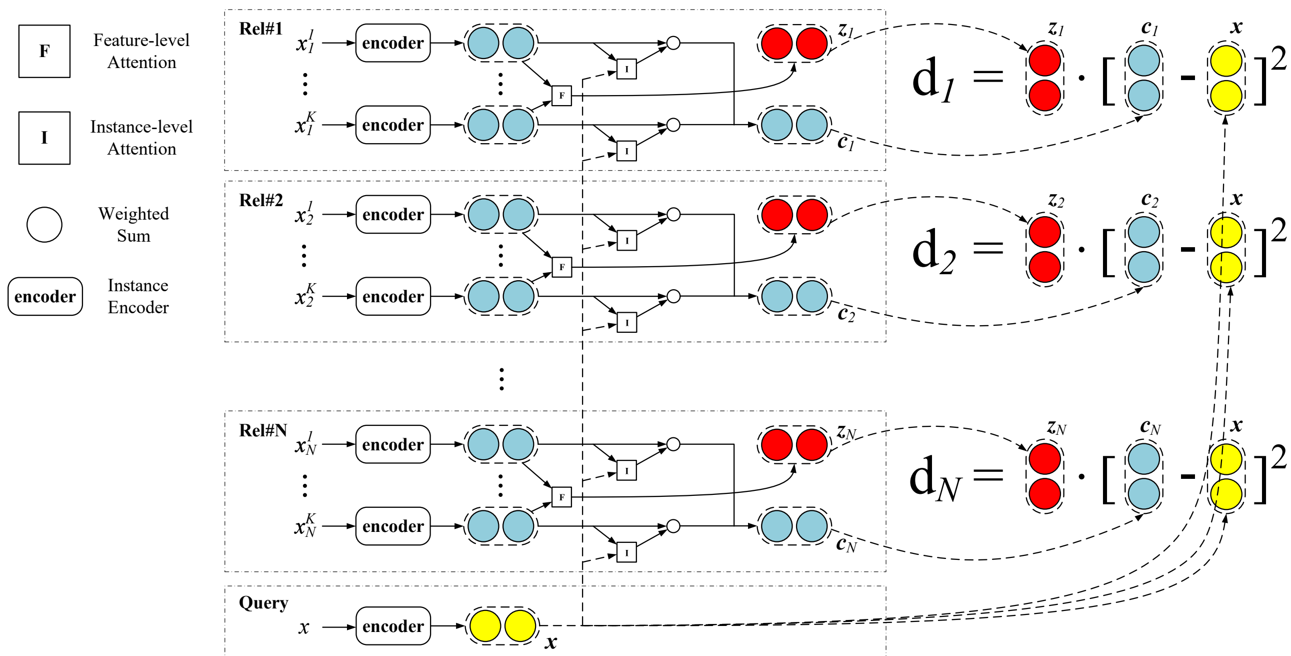
- Build few-shot benchmark: FewRel
  - Transfer existing few-shot algorithms to relation extraction
  - Most few-shot models are designed for images and perform poorly on text
  - NLP domain needs its own few-shot models

Model	5 Way 1 Shot	5 Way 5 Shot	10 Way 1 Shot	10 Way 5 Shot
Finetune (CNN)	44.21 ± 0.44	68.66 ± 0.41	27.30 ± 0.28	55.04 ± 0.31
Finetune (PCNN)	45.64 ± 0.62	57.86 ± 0.61	29.65 ± 0.40	37.43 ± 0.42
kNN (CNN)	54.67 ± 0.44	68.77 ± 0.41	41.24 ± 0.31	55.87 ± 0.31
kNN (PCNN)	60.28 ± 0.43	72.41 ± 0.39	46.15 ± 0.31	59.11 ± 0.30
Meta Network (CNN)	64.46 ± 0.54	80.57 ± 0.48	53.96 ± 0.56	69.23 ± 0.52
GNN (CNN)	66.23 ± 0.75	81.28 ± 0.62	46.27 ± 0.80	64.02 ± 0.77
SNAIL (CNN)	67.29 ± 0.26	79.40 ± 0.22	53.28 ± 0.27	68.33 ± 0.25
Prototypical Network (CNN)	69.20 ± 0.20	84.79 ± 0.16	56.44 ± 0.22	75.55 ± 0.19
Human performance	92.22 ± 5.53	-	85.88 ± 7.40	-

Xu Han et al. FewRel: A Large-Scale Supervised Few-shot Relation Classification Dataset with State-of-the-Art Evaluation. EMNLP 2018.

Few-Shot Relation Extraction

- Text is noisy
- Use instance-level and feature-level attention to improve Prototypical Networks in the noise setting



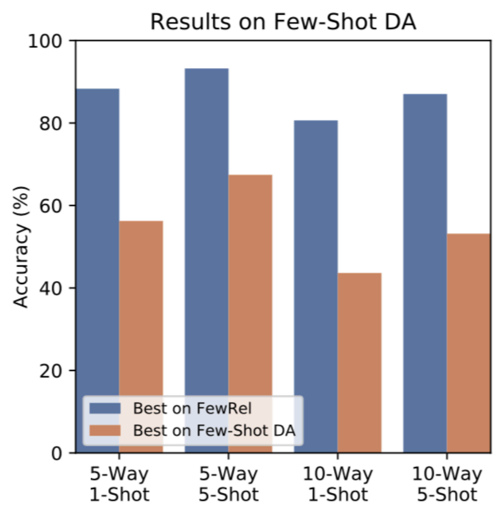
Tianyu Gao, Xu Han, Zhiyuan Liu, Maosong Sun. Hybrid Attention-based Prototypical Networks for Noisy Few-Shot Relation Classification. AAAI 2019.

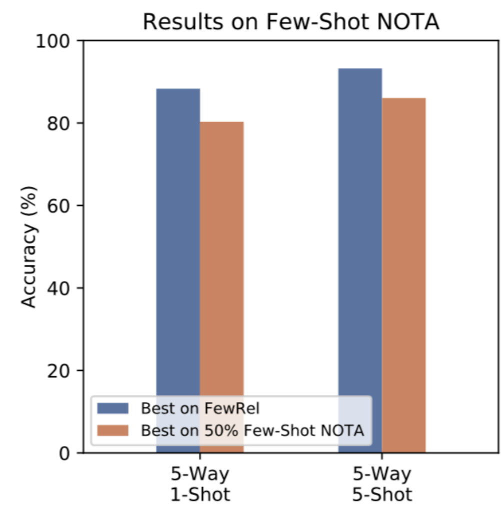
Few-Shot Relation Extraction

- A more challenging benchmark: FewRel 2.0
  - Two new challenges: domain adaptation and none-of-the-above answers
  - Experiments show that existing models cannot handle the two challenges well

Supporting Set	
(A) capital_of	(1) *London* is the capital of *the U.K.* (2) *Washington* is the capital of *the U.S.A.*
(B) member_of	(1) *Newton* served as the president of *the Royal Society*. (2) *Leibniz* was a member of *the Prussian Academy of Sciences*.
(C) birth_name	(1) *Samuel Langhorne Clemens*, better known by his pen name *Mark Twain*, was an American writer. (2) *Alexei Maximovich Peshkov*, primarily known as *Maxim Gorky*, was a Russian and Soviet writer.
Test Instance	
(A) or (B) or (C)	*Euler* was elected a foreign member of *the Royal Swedish Academy of Sciences*.

Table 1: An example for a 3 way 2 shot scenario. Different colors indicate different entities, blue for head entity, and red for tail entity.





Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu , Peng Li, Maosong Sun, Jie Zhou. FewRel 2.0: Towards More Challenging Few-Shot Relation Classification. EMNLP 2019.

Challenge III: Beyond Sentences

- Conventional relation extraction focuses on single sentences and cannot handle more complex knowledge structures

Sentence-level RE

Ernest Hemingway was raised in ***Oak Park, Illinois*** ⇒ *[Ernest Hemingway]* $\xrightarrow{\text{place of birth}}$ *[Oak Park, Illinois]*

Bag-level RE

In 1921, ***Ernest Hemingway*** married ***Hadley Richardson***, the first of his four wives

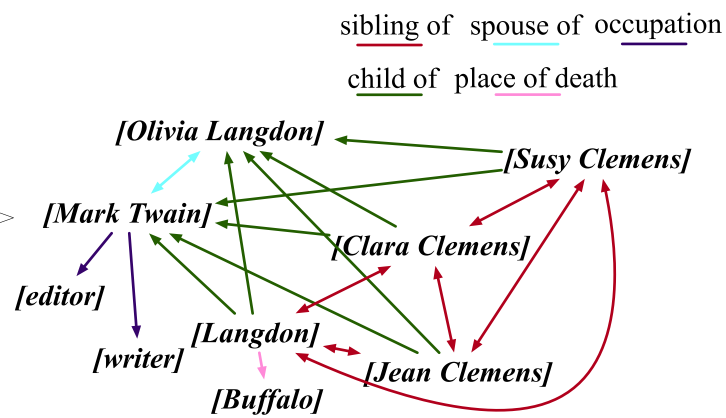
Hadley Richardson was the first wife of American author ***Ernest Hemingway***

… … …

⇒ *[Ernest Hemingway]* $\xrightarrow{\text{spouse}}$ *[Hadley Richardson]*

Document-level RE

Mark Twain and ***Olivia Langdon*** corresponded throughout 1868. She rejected his first marriage proposal, but they were married in Elmira, New York in February 1870. Then, Twain owned a stake in the Buffalo Express newspaper and worked as an ***editor*** and ***writer***. While they were living in ***Buffalo***, their son ***Langdon*** died of diphtheria at the age of 19 months. They had three daughters: ***Susy Clemens***, ***Clara Clemens***, and ***Jean Clemens***.



Document-Level Relation Extraction

- Large-scale document-level RE benchmark: DocRED
  - The largest relation extraction dataset ever
  - To solve DocRED, a model needs to do pattern recognition, logical reasoning, coreference resolution and common-sense reasoning

Dataset	# Doc	# Word	# Sent	# Ent	# Rel	# Inst	# Fact
SemEval-2010 Task 8	-	205k	10,717	21,434	9	8,853	8,383
ACE 2003-2004	-	297k	12,783	46,108	24	16,771	16,536
TACRED	-	1,823k	53,791	152,527	41	21,773	5,976
FewRel	-	1,397k	56,109	72,124	100	70,000	55,803
BC5CDR	1,500	282k	11,089	29,271	1	3,116	2,434
DocRED (Human-Annotated)	5,053	1,002k	40,348	129,767	96	71,046	62,762
DocRED (Distantly Supervised)	101,997	21,368k	870,605	2,771,037	96	1,678,864	760,917

Reasoning Types	%	Examples
Pattern recognition	38.9	[1] ***Me Musical Nephews*** is a **1942** one-reel animated cartoon directed by Seymour Kneitel and animated by Tom Johnson and George Germanetti. [2] Jack Mercer and Jack Ward wrote the script. ... **Relation:** `publication_date` **Supporting Evidence: 1**
Logical reasoning	26.6	[1] "Nisei" is the ninth episode of the third season of the American science fiction television series The X-Files. ... [3] It was directed by David Nutter, and written by **Chris Carter**, Frank Spotnitz and Howard Gordon. ... [8] The show centers on FBI special agents ***Fox Mulder*** (David Duchovny) and Dana Scully (Gillian Anderson) who work on cases linked to the paranormal, called X-Files. ... **Relation:** `creator` **Supporting Evidence: 1, 3, 8**
Coreference reasoning	17.6	[1] ***Dwight Tillery*** is an American politician of the Democratic Party who is active in local politics of Cincinnati, Ohio. ... [3] He also holds a law degree from the **University of Michigan Law School**. [4] ***Tillery*** served as mayor of Cincinnati from 1991 to 1993. **Relation:** `educated_at` **Supporting Evidence: 1, 3**
Common-sense reasoning	16.6	[1] ***William Busac*** (1020-1076), son of William I, Count of Eu, and his wife Lesceline. ... [4] ***William*** appealed to King Henry I of France, who gave him in marriage **Adelaide**, the heiress of the county of Soissons. [5] **Adelaide** was daughter of Renaud I, Count of Soissons, and Grand Master of the Hotel de France. ... [7] ***William*** and **Adelaide** had four children: ... **Relation:** `spouse` **Supporting Evidence: 4, 7**

Document-Level Relation Extraction

- Experiment
  - Existing models perform poorly on DocRED

Model	Dev Ign F1	Dev Ign AUC	Dev F1	Dev AUC	Test Ign F1	Test Ign AUC	Test F1	Test AUC
Supervised Setting								
CNN	44.55	41.18	47.57	45.24	43.04	40.68	46.35	44.70
LSTM	47.15	44.99	49.99	49.08	46.14	44.25	49.53	48.98
BiLSTM	48.31	46.75	51.16	50.98	46.77	45.83	50.09	50.21
Context-Aware	**48.46**	**46.78**	**51.57**	**51.07**	**47.87**	**46.27**	**51.26**	**50.71**
Weakly Supervised Setting								
CNN	35.70	27.47	46.40	43.90	35.41	26.59	46.38	43.05
LSTM	39.79	27.74	51.92	46.97	36.64	26.97	51.33	46.30
BiLSTM	**41.95**	**28.63**	54.21	**48.54**	38.23	**27.59**	53.38	46.90
Context-Aware	41.75	27.88	**54.26**	47.74	**39.10**	27.14	**53.75**	**46.91**

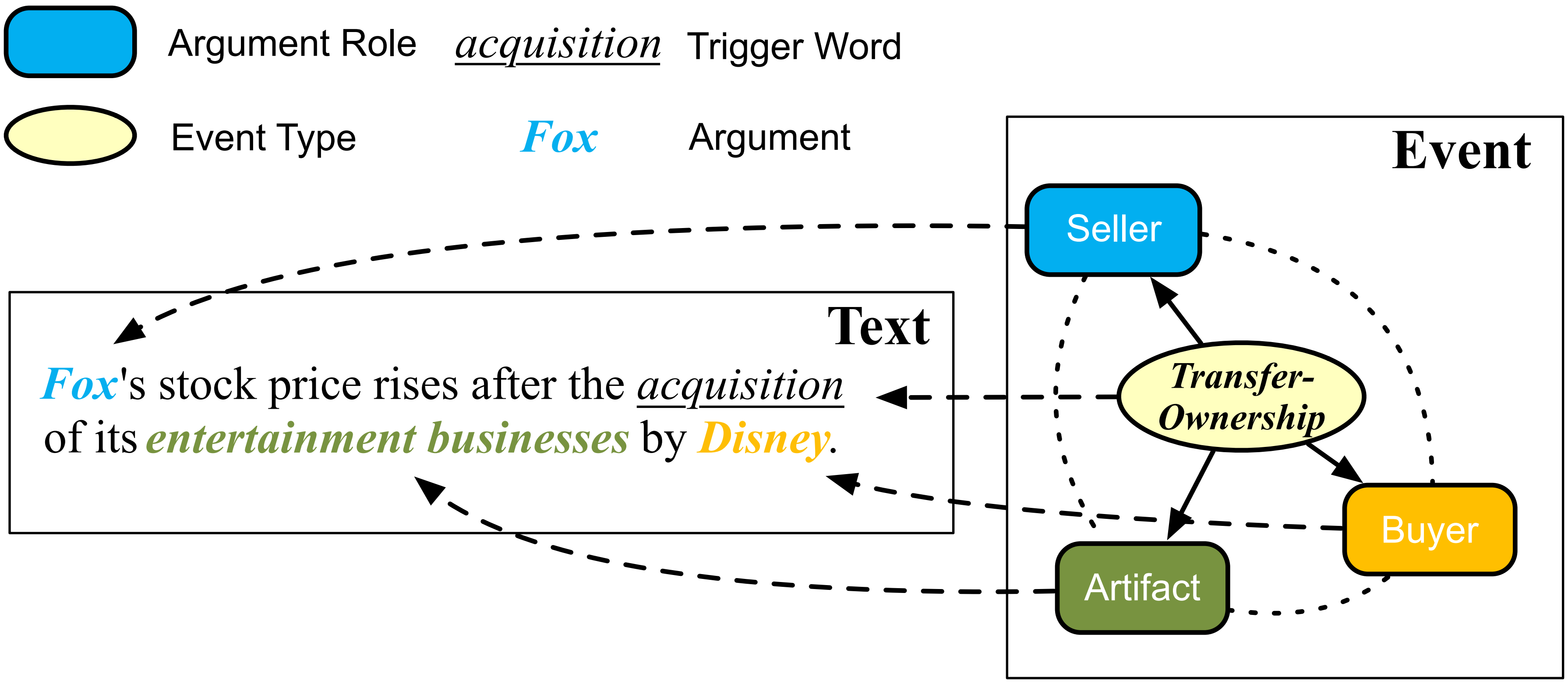
  - There are huge gaps between model performance and human performance

Method	RE P	RE R	RE F1	RE+Sup P	RE+Sup R	RE+Sup F1
Model	55.6	52.6	54.1	46.4	43.1	44.7
Human	**89.7**	**86.3**	**88.0**	**71.2**	**75.8**	**73.4**

Yuan Yao et al. DocRED: A Large-Scale Document-Level Relation Extraction Dataset. ACL 2019.

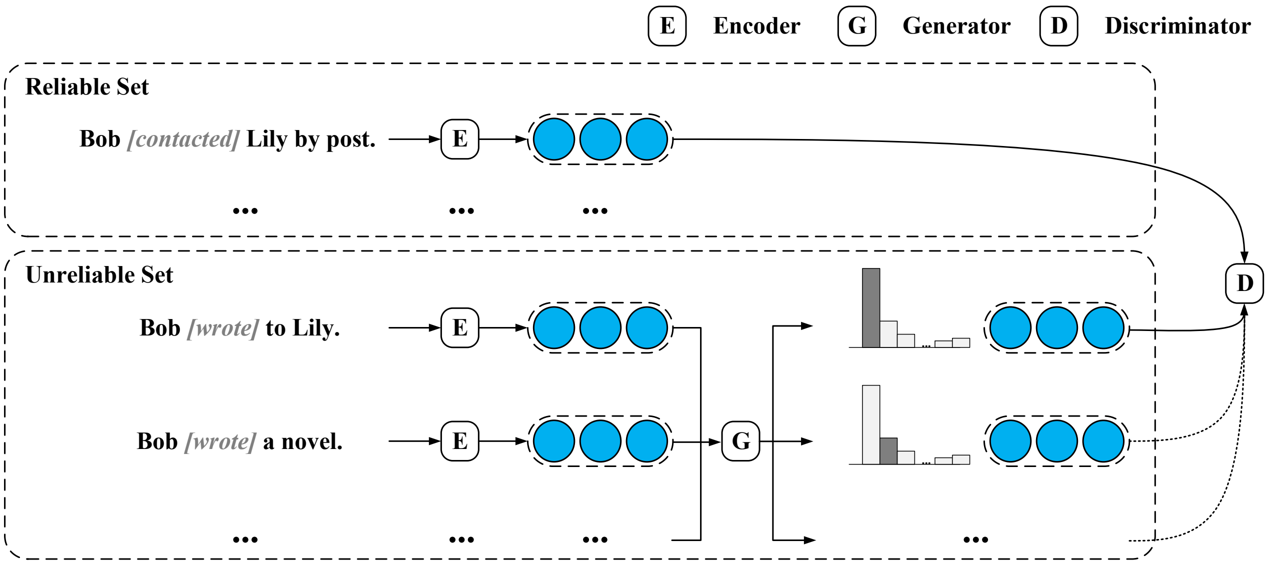
Challenge IV: Event Detection

- Event Detection



Event Detection

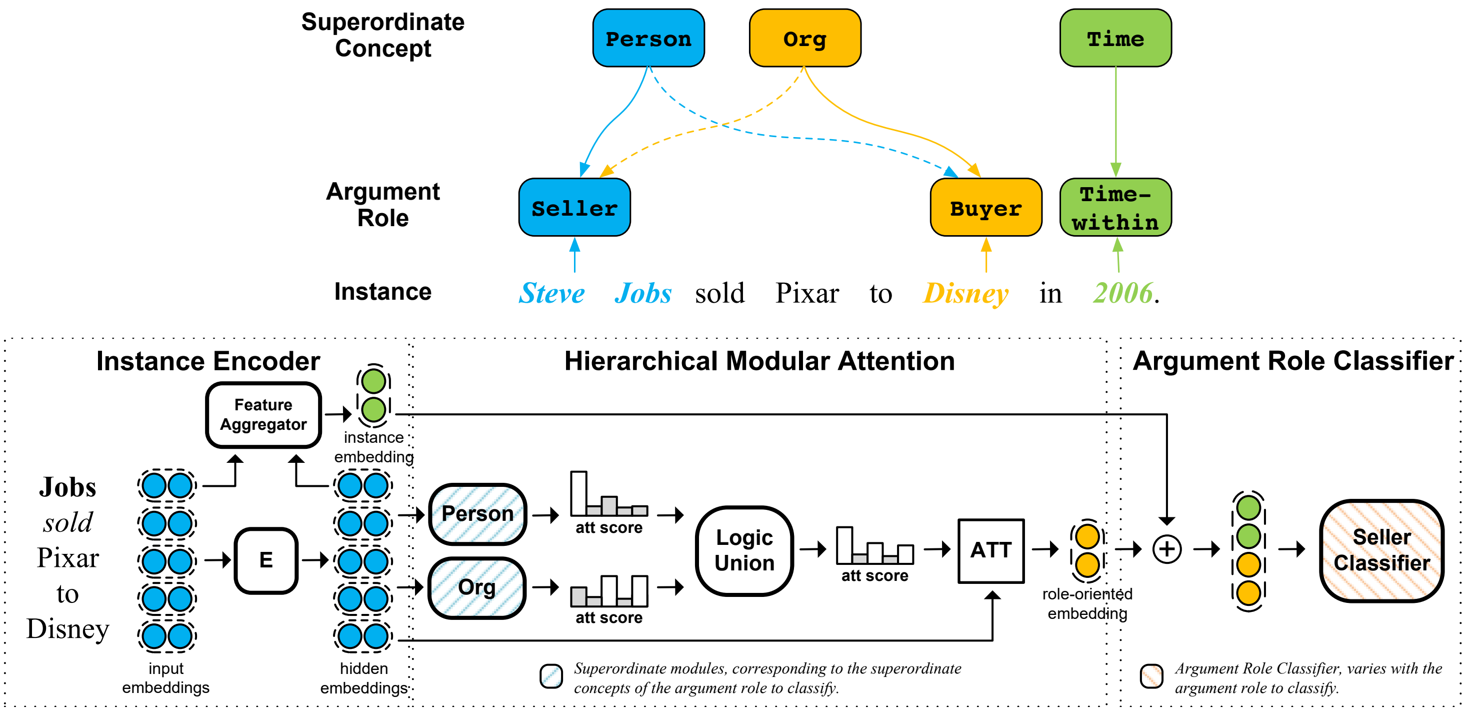
- Introduce large-scale unsupervised corpus to solve the data deficiency in event detection
- Introduce adversarial training to denoise data
  - Split the dataset as “reliable” and “unreliable”
  - Use adversarial training to filter new data from the “unreliable” set to the “reliable” set



Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, Peng Li. Adversarial Training for Weakly Supervised Event Detection. NAACL-HLT 2019.

Event Detection

- Introduce abstract concept of event arguments
  - Use modular networks and logical operators to extract event arguments. Then join the extraction results.



Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, Xiang Ren. HMEAE: Hierarchical Modular Event Argument Extraction. EMNLP 2019.

Summary

- More Data
  - Denoising for distant supervision
  - Learning for one/few shot instances
  - Multi-lingual relation extraction
- More Context
  - Document-level relation extraction
- More Structure
  - Complex event extraction
- More Type
  - Open Relation Extraction (for new types)

Knowledge Extraction Paper List

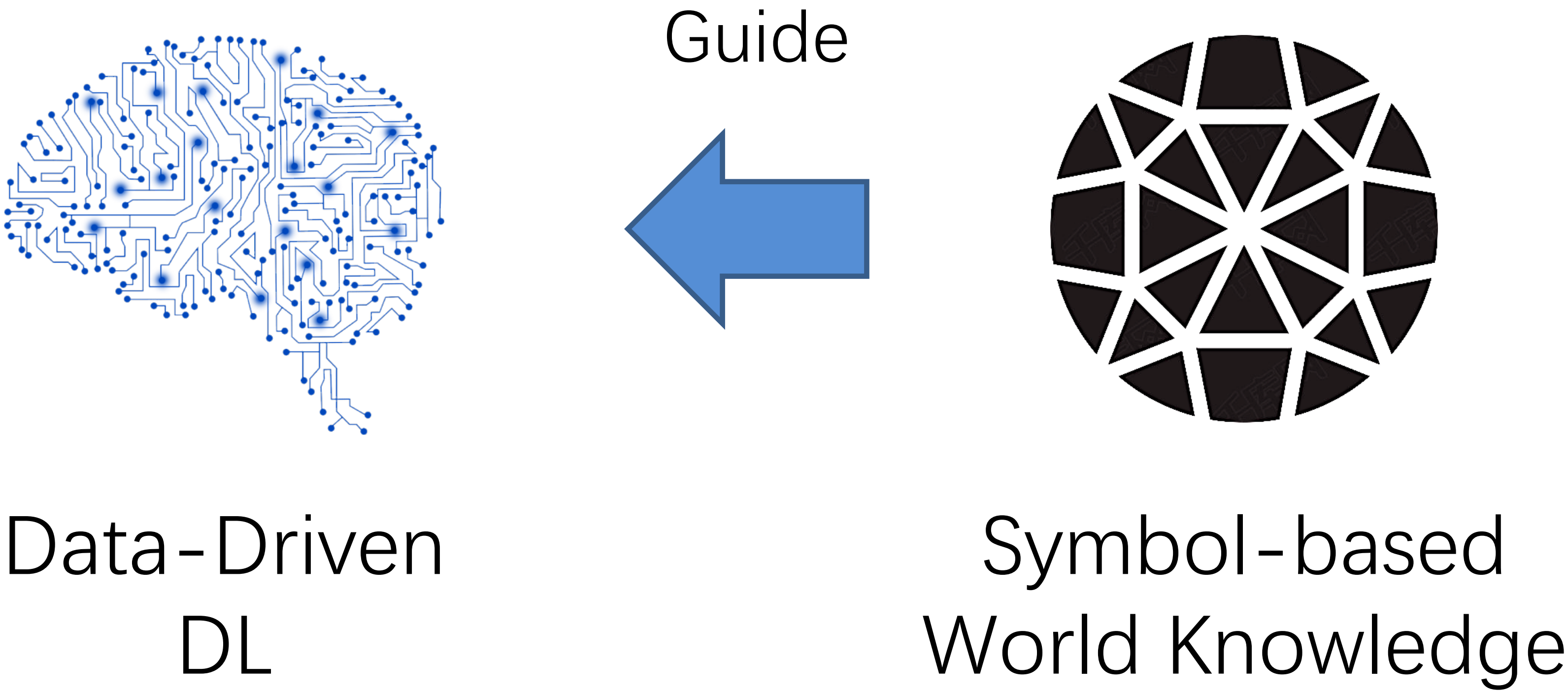
https://github.com/thunlp/NREPapers

- 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.

Knowledge Extraction Paper List

https://github.com/thunlp/NREPapers

- Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, Xiang Ren. **HMEAE: Hierarchical Modular Event Argument Extraction**. EMNLP 2019.
- Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, Peng Li. **Adversarial Training for Weakly Supervised Event Detection**. NAACL-HLT 2019.
- Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, Maosong Sun. **DocRED: A Large-Scale Document-Level Relation Extraction Dataset**. ACL 2019.
- Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu , Peng Li, Maosong Sun, Jie Zhou**. FewRel 2.0: Towards More Challenging Few-Shot Relation Classification**. EMNLP 2019.
- Tianyu Gao, Xu Han, Zhiyuan Liu, Maosong Sun. **Hybrid Attention-based Prototypical Networks for Noisy Few-Shot Relation Classification**. AAAI 2019. Long paper.
- Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, Maosong Sun. **FewRel: A Large-Scale Supervised Few-shot Relation Classification Dataset with State-of-the-Art Evaluation**. EMNLP 2018.
- Shun Zheng, Xu Han, Yankai Lin, Peilin Yu, Lu Chen, Ling Huang, Zhiyuan Liu, Wei Xu. **DIAG-NRE: A Neural Pattern Diagnosis Framework for Distantly Supervised Neural Relation Extraction**. ACL 2019.


Guide
Data-Driven
DL
Symbol-based
World Knowledge

Entity Typing

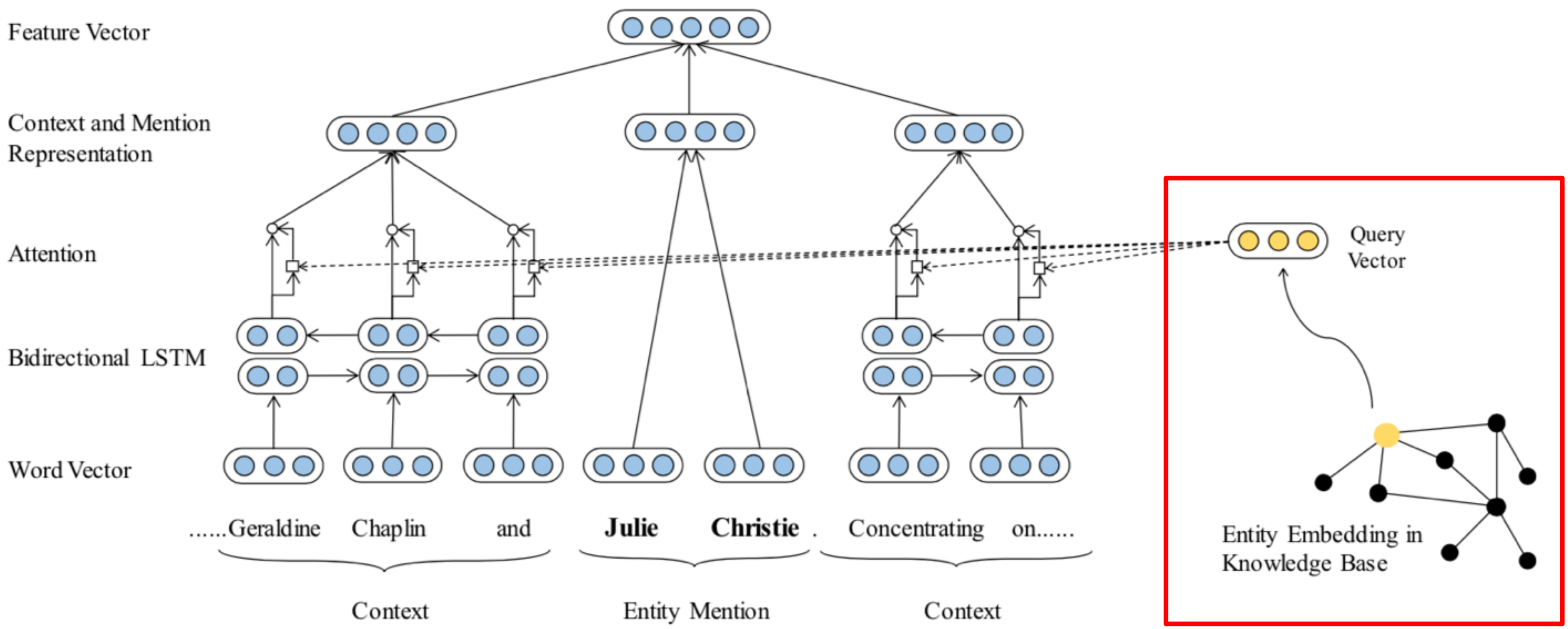
- **Entity Typing** is the task of labeling a given entity mention in text with several semantic labels.

Sentence with Target Entity	Entity Types
During the Inca Empire, {**the Inti Raymi**} was the most important of four ceremonies celebrated in Cusco.	event, festival, **ritual, custom, ceremony, party, celebration**
{**They**} have been asked to appear in court to face the charge.	person, **accused, suspect, defendant**
Ban praised Rwanda's commitment to the UN and its role in {**peacemaking operations**}.	event, **plan, mission, action**

Eunsol Choi, Omer Levy, Yejin Choi, Luke Zettlemoyer. Ultra-Fine Entity Typing. ACL 2018.

Knowledge-guided Entity Typing

- Fine-grained entity typing
- Based on KG embeddings, propose Knowledge attention for better context understanding



Ji Xin, Yankai Lin, Zhiyuan Liu, Maosong Sun. Improving Neural Fine-Grained Entity Typing with Knowledge Attention. The 32th AAAI Conference on Artificial Intelligence (AAAI 2018).

Knowledge-guided Entity Typing

- The key to utilize knowledge is to find out related entities in the KG
  - The alignments between mentions and entities are given during training
  - The alignments are unknown during inference
- Disambiguation
- Compute an approximate entity embedding $\hat{e}$
- Retrieve the candidate entity set based on the mention; select the candidate with the smallest L2 distance with $\hat{e}$ as the corresponding entity

Ji Xin, Yankai Lin, Zhiyuan Liu, Maosong Sun. Improving Neural Fine-Grained Entity Typing with Knowledge Attention. The 32th AAAI Conference on Artificial Intelligence (AAAI 2018).

Knowledge-guided Entity Typing

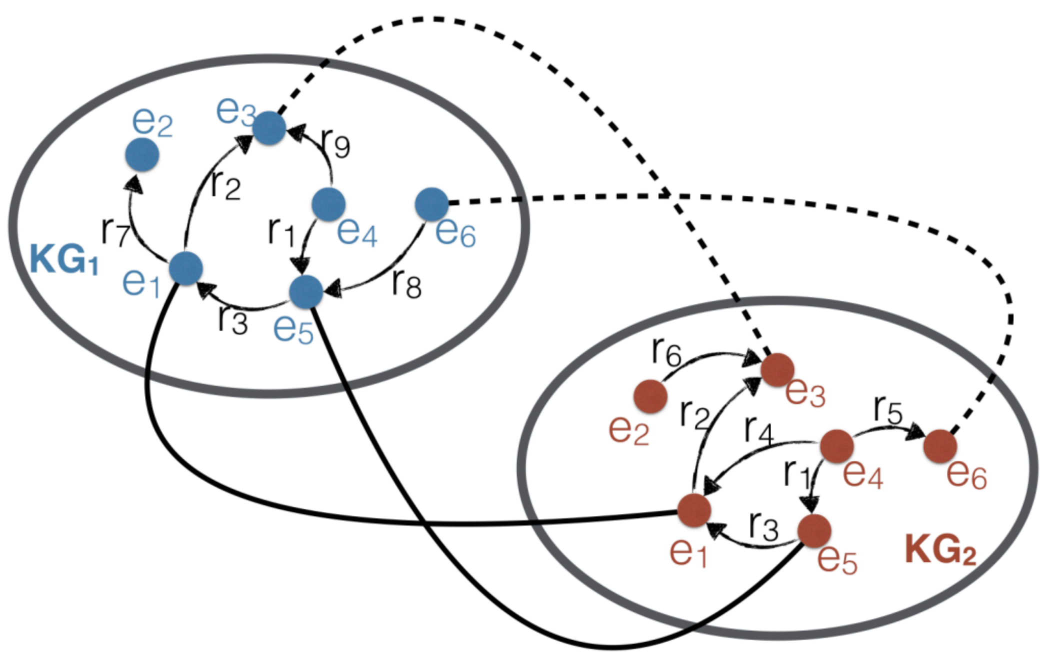
- KA and KA+D outperform all baselines, which indicates the effectiveness of knowledge
- KA+D means KA with Disambiguation

Dateset	WIKI-AUTO						
Metrics	Strict	Macro			Micro		
	Acc	Pre	Rec	F1	Pre	Rec	F1
AFET	20.32	67.00	45.82	54.75	69.29	42.40	52.61
KB-ONLY	35.12	69.65	71.35	70.49	54.85	74.99	63.36
HNM	34.88	68.09	61.03	64.37	72.80	64,48	68.39
SA	42.77	75.33	69.69	72.40	77.35	72.63	74.91
MA	41.58	73.64	71.71	72.66	75.94	75.52	75.72
KA	45.49	74.82	72.46	73.62	76.96	75.49	76.22
KA+D	**47.20**	**75.72**	**74.03**	**74.87**	**77.96**	**77.87**	**77.92**

Ji Xin, Yankai Lin, Zhiyuan Liu, Maosong Sun. Improving Neural Fine-Grained Entity Typing with Knowledge Attention. The 32th AAAI Conference on Artificial Intelligence (AAAI 2018).

Entity Alignment

- Goal: Align synonymous entity pairs from heterogeneous Knowledge Graphs
- Entity Alignment can be regarded as a graph matching problem



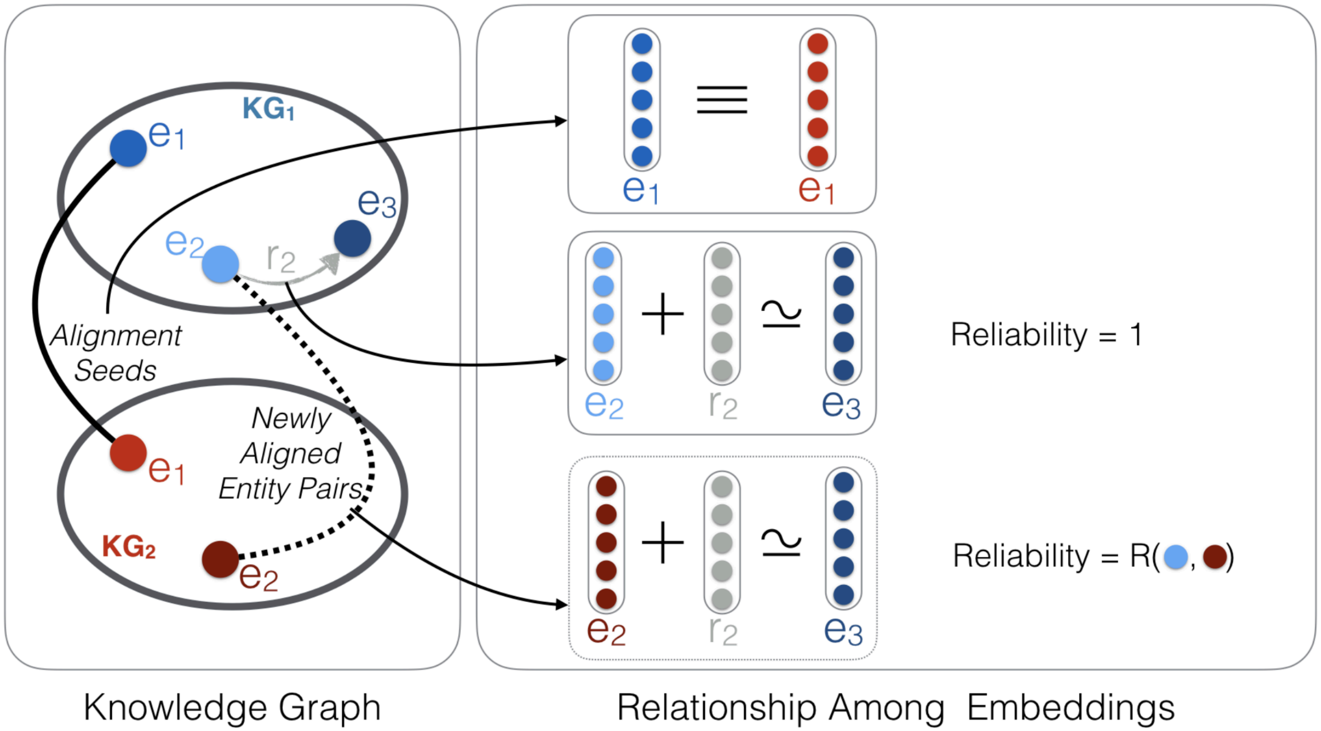
Knowledge-guided Entity Alignment

- Graph matching algorithms cannot be applied to knowledge graphs directly
- Propose to use KRL to solve this problem
- Method
  - KRL: Knowledge graph learning for graphs
  - Parameter sharing: Make those aligned entities share the same embeddings
  - Iterative Alignment: Add new entity alignments based on the distance and use them to find more entities to align

Hao Zhu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Iterative Entity Alignment via Joint Knowledge Embeddings. IJCAI 2017.

Knowledge-guided Entity Alignment

- The solid line and the dashed line between KGs denote alignment seeds and newly aligned entity pairs during iterative learning



Hao Zhu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Iterative Entity Alignment via Joint Knowledge Embeddings. IJCAI 2017.

Knowledge-guided Entity Alignment

- Build three datasets based on FB15K (DFB-1,2,3)
- Knowledge-guided Entity Alignment achieves the best performance

指标	DFB-1			DFB-2			DFB-3		
	Hits@1 (%)	Hits@10 (%)	MR	Hits@1 (%)	Hits@10 (%)	MR	Hits@1 (%)	Hits@10 (%)	MR
MTransE (LT)	38.9	61.0	237.7	12.3	33.8	419.2	6.5	22.0	699.8
MTransE (TB)	13.6	35.1	547.7	13.9	35.4	675.7	4.5	16.1	1255.5
TransE + PS	61.9	79.2	105.2	41.1	67.0	154.9	12.2	34.6	431.9
ITransE (HA)	62.6	78.9	100.0	41.2	66.9	151.9	12.3	33.7	432.3
ITransE (SA)	**67.1**	**83.1**	**80.1**	**57.7**	**77.7**	**109.3**	**16.2**	**40.9**	**367.2**
PTransE + PS	65.8	83.4	62.9	46.3	72.1	96.8	15.8	40.2	346.9
IPTransE (HA)	66.1	83.3	59.1	46.2	72.6	94.2	15.1	39.7	337.6
IPTransE (SA)	**71.7**	**86.5**	**49.0**	**63.5**	**82.2**	**67.5**	**20.4**	**47.4**	**281.0**

Hao Zhu, Ruobing Xie, Zhiyuan Liu, Maosong Sun. Iterative Entity Alignment via Joint Knowledge Embeddings. IJCAI 2017.

Knowledge-guided Neural Ranking

- Queries and documents often match based on knowledge

Query: Obama family tree

Documents:

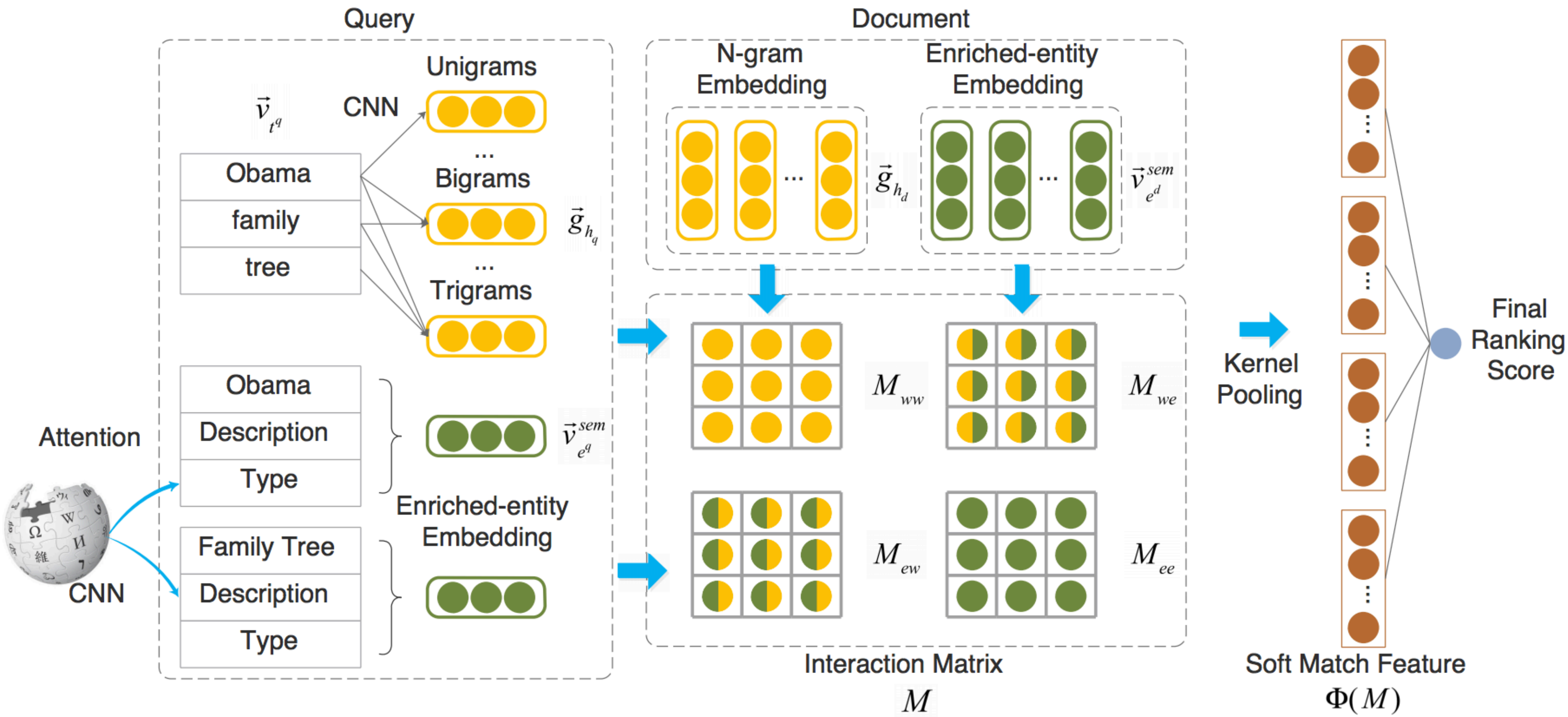
Ranking	Document	Score
1	The family of **Barack Obama**, the 44th President of the **United States**, and his wife **Michelle Obama**...	12.0
2	**Barack Hussein Obama** is an American politician who served as the 44th ...	5.0

- Study the effectiveness of knowledge graph semantics in state-of-the-art neural ranking models

Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval. ACL 2018.

Knowledge-guided Neural Ranking

- Introduce world knowledge from KGs into KNRM



Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval. ACL 2018.

Knowledge-guided Neural Ranking

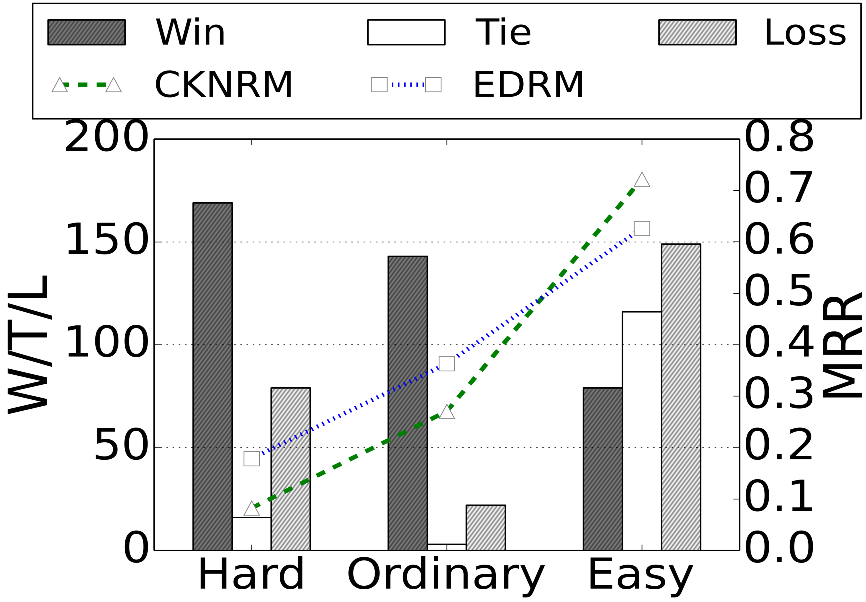
- Knowledge-guided models achieve significant improvement on KNRM

	Testing-SAME				Testing-DIFF				Testing-RAW	
Method	**NDCG@1**		**NDCG@10**		**NDCG@1**		**NDCG@10**		**MRR**	
BM25	0.142	-46%	0.287	-32%	0.163	-46%	0.325	-23%	0.228	-34%
RankSVM	0.146	-45%	0.309	-26%	0.170	-43%	0.352	-17%	0.224	-35%
Coor-Ascent	0.159	-40%	0.355	-15%	0.209	-30%	0.378	-11%	0.242	-30%
DRMM	0.137	-48%	0.313	-25%	0.213	-29%	0.359	-15%	0.234	-32%
CDSSM	0.144	-46%	0.333	-21%	0.183	-39%	0.353	-16%	0.231	-33%
MP	0.218	-17%	0.379	-10%	0.197	-34%	0.345	-18%	0.240	-30%
K-NRM	0.265	–	0.420	–	0.300	–	0.423	–	0.345	–
Conv-KNRM	0.336	27%	0.481	15%	0.338	13%	0.432	2%	0.358	4%
EDRM-KNRM	0.310	17%	0.455	8%	0.333	11%	0.434	3%	0.362	5%
EDRM-CKNRM	**0.340**	**28%**	**0.482**	**15%**	**0.371**	**24%**	**0.451**	**7%**	**0.389**	**13%**

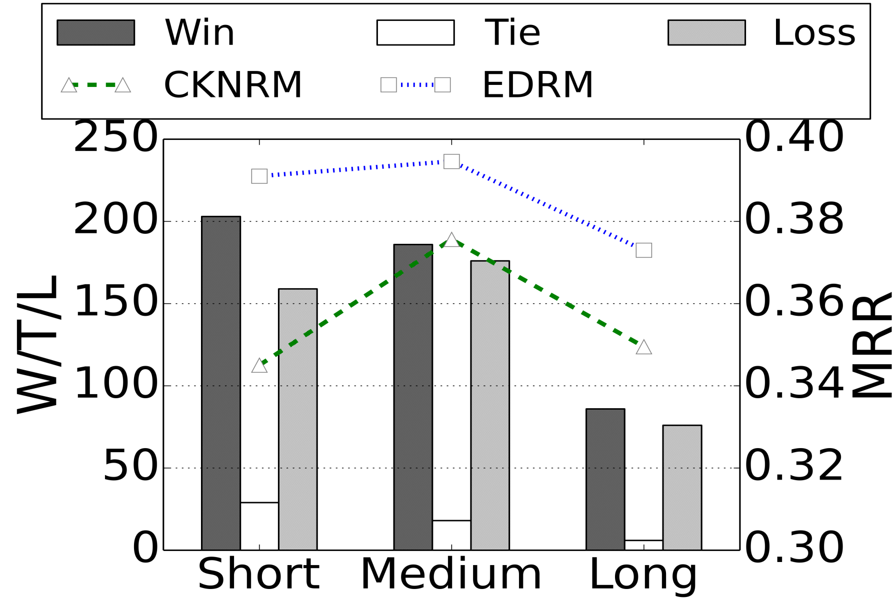
Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval. ACL 2018.

Knowledge-guided Neural Ranking

- More improvement on short and hard queries
- Knowledge are more important for the limited query text



Query Difficulty Scenario



Query Length Scenario

Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval. ACL 2018.

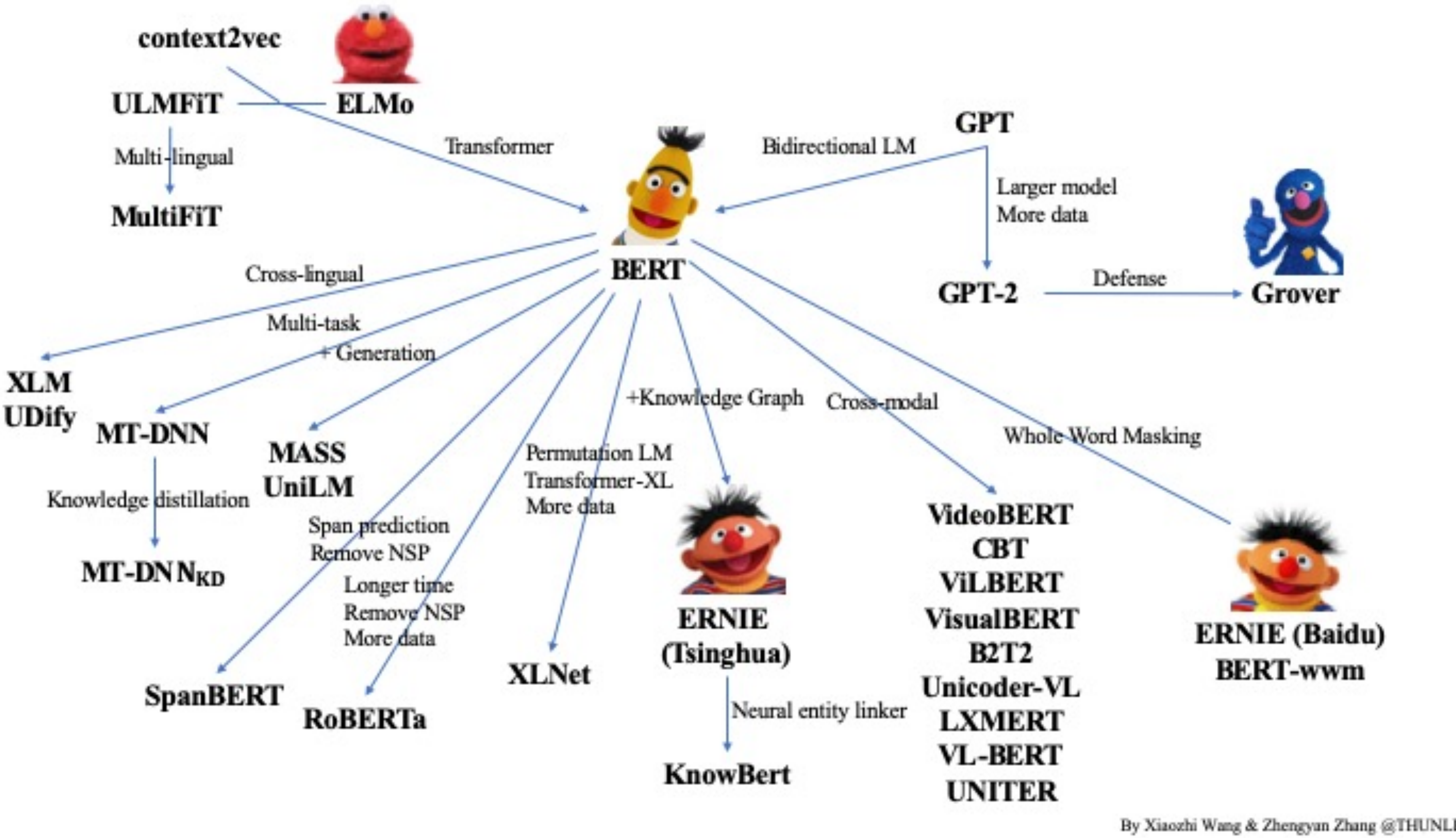
Pretrained Language Model

- Impressive progress of deep learning on unsupervised text corpora

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

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

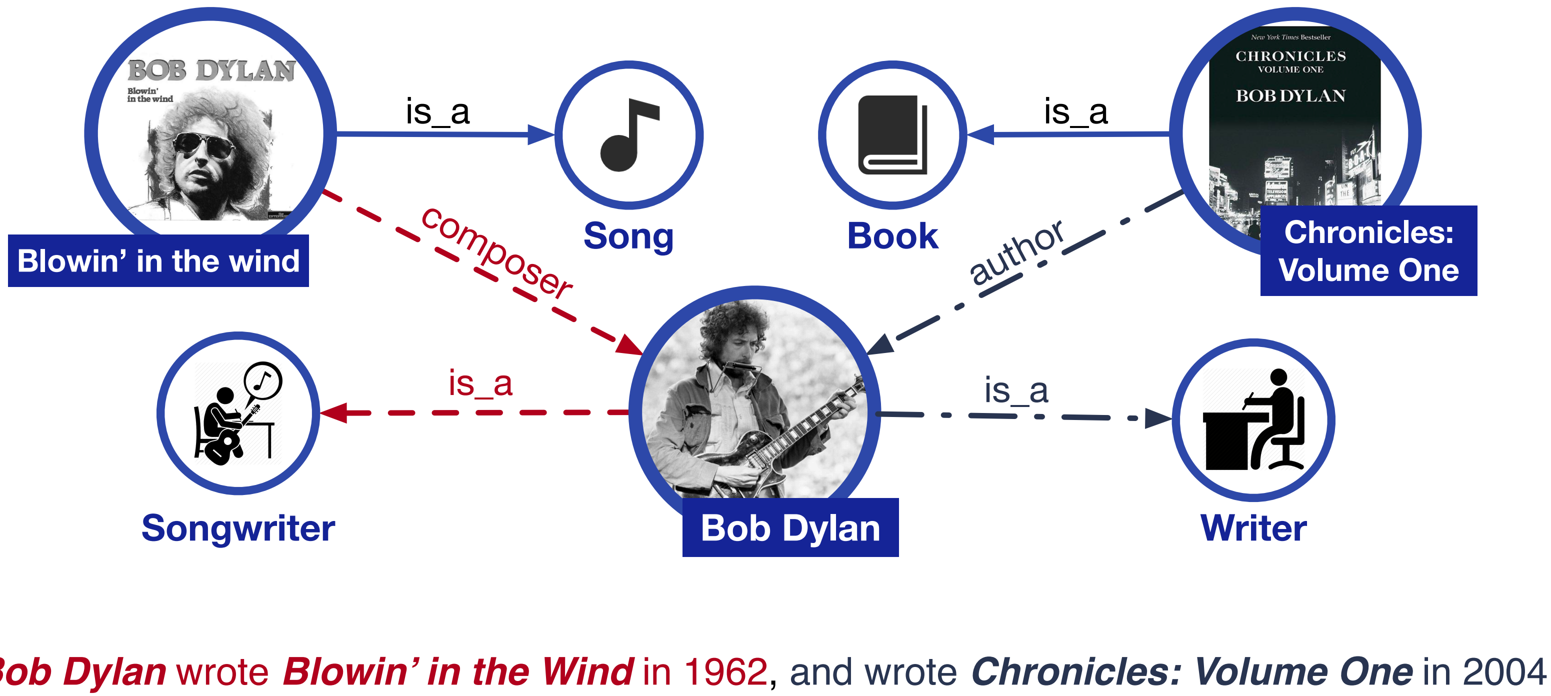
Pretrained Language Model



https://github.com/thunlp/PLMpapers

Knowledge-guided PLM

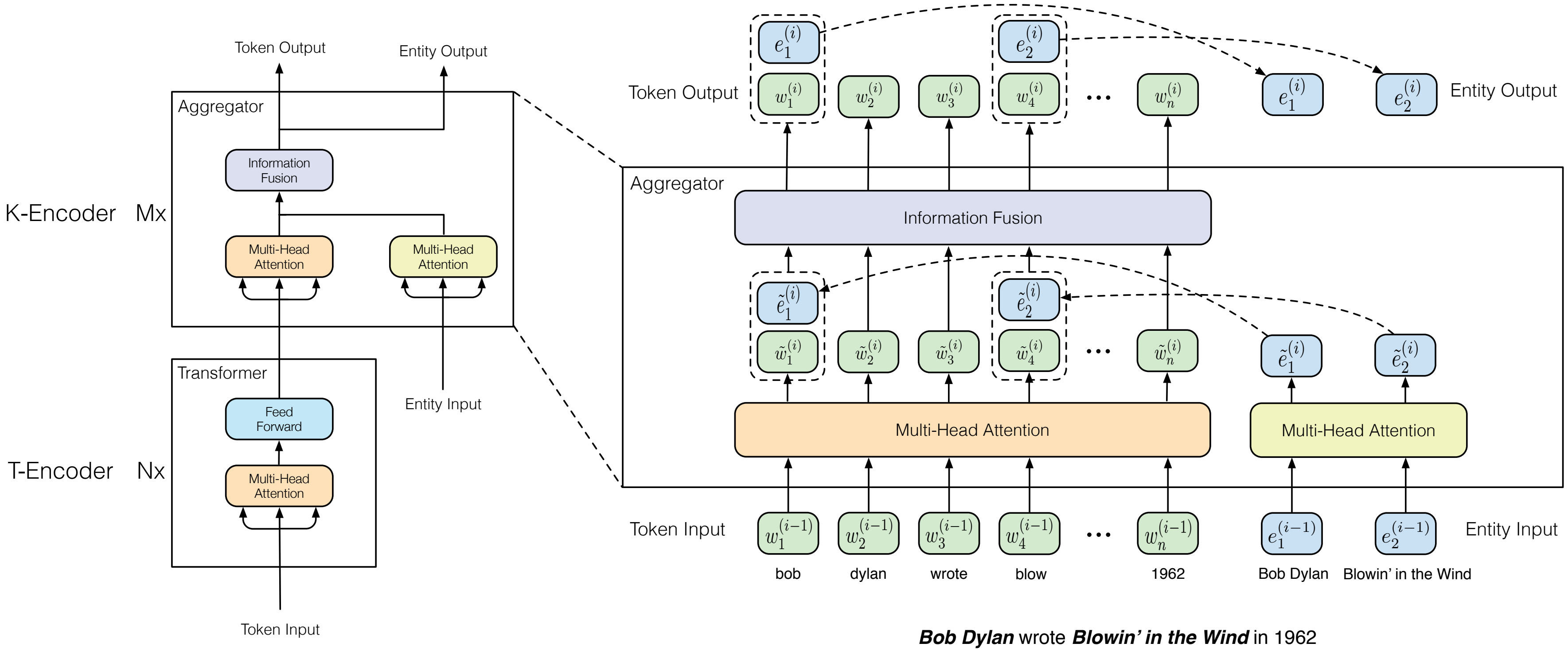
- Intuitively, external knowledge information can effectively benefit language understanding
  - Low resource entities
  - Implicit background knowledge



Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, Qun Liu. ERNIE: Enhanced Language Representation with Informative Entities. ACL 2019.

Knowledge-guided PLM

- The architecture of ERNIE
  - Lower layers for text
  - Higher layers for knowledge integration



(a) Model Achitecture

(b) Aggregator

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, Qun Liu. ERNIE: Enhanced Language Representation with Informative Entities. ACL 2019.

World Knowledge Guided NLP Paper List

- 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.

PART THREE

Legal Domain Knowledge

Zhiyuan Liu

Tsinghua NLP

Legal Intelligence

- What is Legal Intelligence?
  - Use large scaled legal documents
  - Combined with AI and NLP technology
  - Solve tasks for legal system

- Several tasks of Legal Intelligence：
  - Judgment Prediction: Predict judgment results from facts
  - Similar Case Matching: Match the best case to ensure fairness
  - Legal Question Answering: Provide legal consultation service
  - ……

Legal Intelligence

- Several challenges and difficulties：
- Interpretability and reliability:
  - Unreasonable or unexplained results cannot be applied to real legal systems
  - Real legal systems emphasize fairness and justice
- The fusion of legal knowledge:
  - The task of legal intelligence involves a large number of professional concepts and knowledge
  - How to integrate legal professional knowledge into legal tasks is essential for legal intelligence

Pretrained Legal Language Model

- PLM under common corpus cannot be directly transferred to the legal text, because the legal text has many professional vocabulary and different writing habits
- In order to provide a PLM for legal intelligence tasks, we use the legal corpus to pretrain the legal language models for helping downstream legal tasks

名称	基础模型	数据来源	训练数据大小	词表大小	模型大小
民事文书BERT	bert-base	全部民事文书	2654万篇文书	22554	370MB
刑事文书BERT	bert-base	全部刑事文书	663万篇文书	22554	370MB

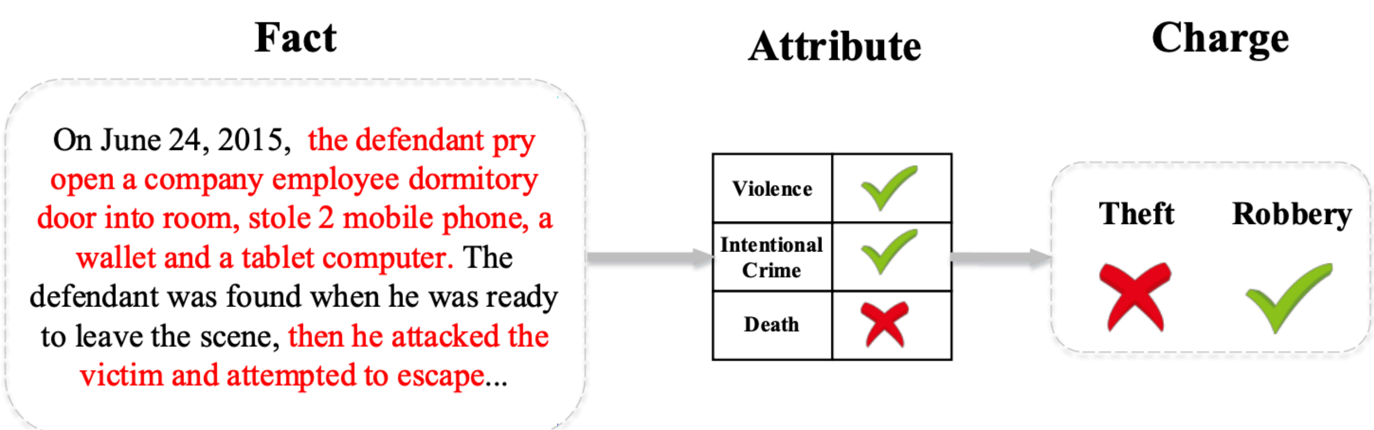
https://github.com/thunlp/OpenCLaP

Legal Judgment Prediction

- Legal Judgment Prediction aims to predict the judgment result based on the facts of a case
- Long Tail Distribution: the case number of various charges are highly imbalanced
- Confusing Charges: there exist many confusing charge pairs, such as (misappropriation of funds, embezzlement), whose definitions are very similar to each other

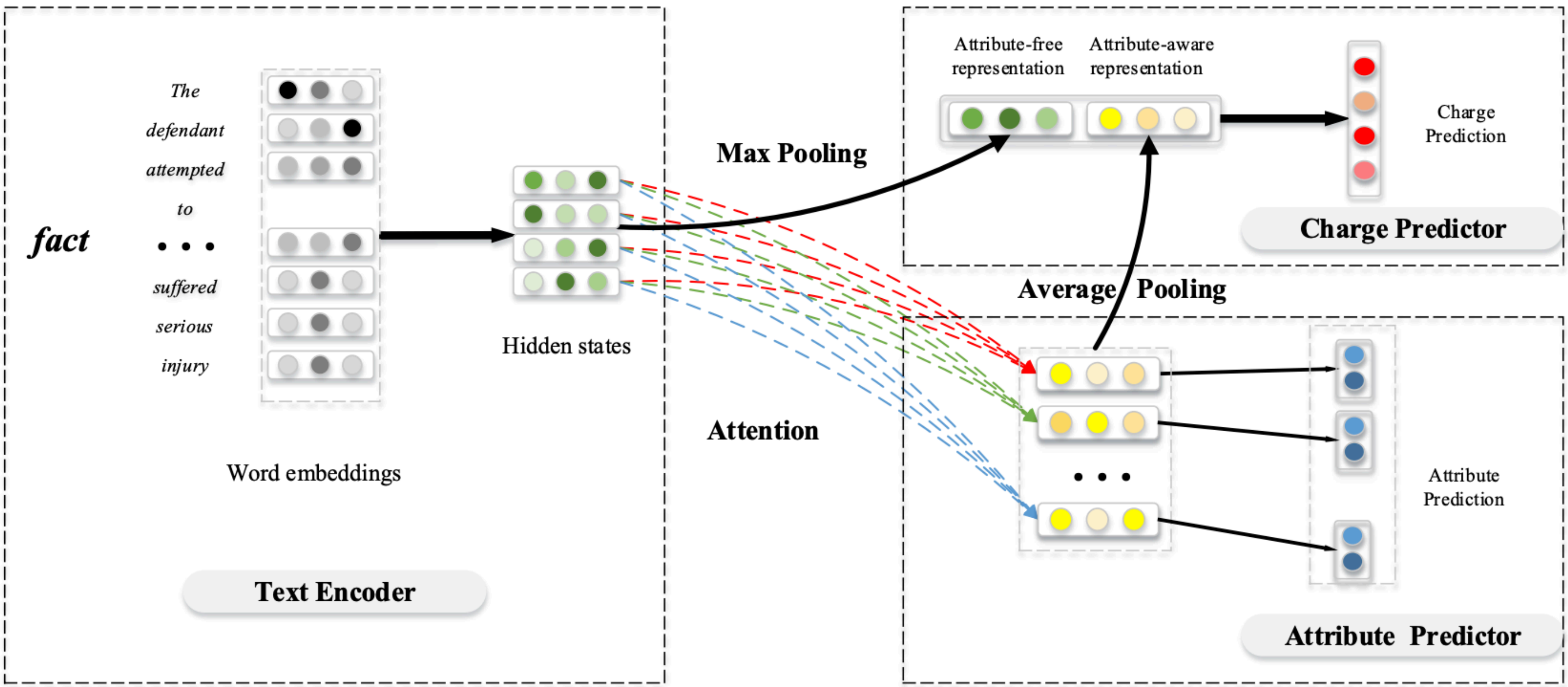
Charge Prediction with Discriminative Legal Attributes

- Attribute-based Judgement: a trial method that conducts trials around the basic attribute of the case
- Following the attribute-based judgment method, the knowledge of legal attribute is introduced to predict the judgment result



Hu, Zikun, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. Few-shot charge prediction with discriminative legal attributes. COLING 2018.

Charge Prediction with Discriminative Legal Attributes



Hu, Zikun, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. Few-shot charge prediction with discriminative legal attributes. COLING 2018.

Experiment Results

- Attribute-based model can outperform existing legal judgment prediction methods

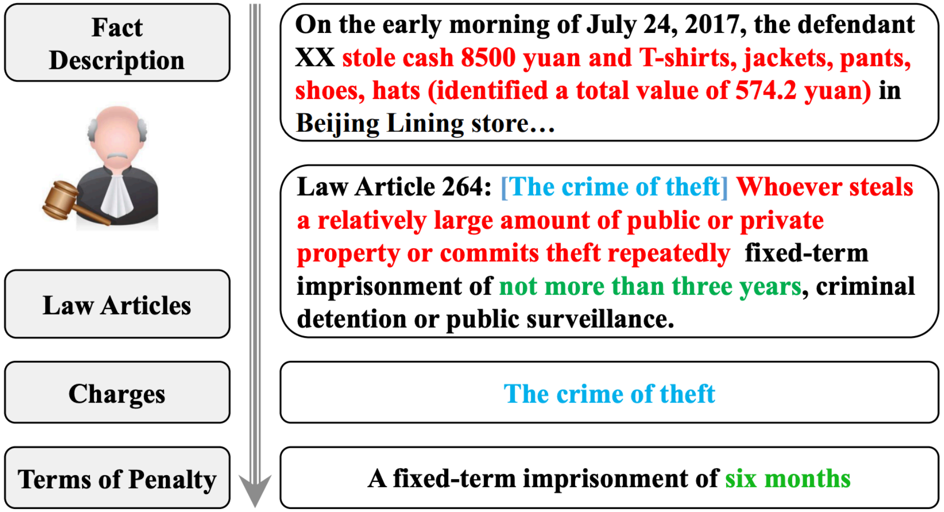
Datasets	Criminal-S				Criminal-M				Criminal-L			
Metrics	Acc.	MP	MR	F1	Acc.	MP	MR	F1	Acc.	MP	MR	F1
TFIDF+SVM	85.8	49.7	41.9	43.5	89.6	58.8	5 0.1	52.1	91.8	67.5	54.1	57.5
CNN	91.9	50.5	44.9	46.1	93.5	57.6	48.1	50.5	93.9	66.0	50.3	54.7
CNN-200	92.6	51.1	46.3	47.3	92.8	56.2	50.0	50.8	94.1	61.9	50.0	53.1
LSTM	93.5	59.4	58.6	57.3	94.7	65.8	63.0	62.6	95.5	69.8	67.0	66.8
LSTM-200	92.7	60.0	58.4	57.0	94.4	66.5	62.4	62.7	95.1	72.8	66.7	67.9
Fact-Law Att.	92.8	57.0	53.9	53.4	94.7	66.7	60.4	61.8	95.7	73.3	67.1	68.6
Our Model	**93.4**	**66.7**	**69.2**	**64.9**	**94.4**	**68.3**	**69.2**	**67.1**	**95.8**	**75.8**	**73.7**	**73.1**

- The model achieves great improvements than baseline method for the low-frequency charges

Charge Type	Low frequency	Medium frequency	High frequency
Charge Number	49	51	49
LSTM-200	32.6	55.0	83.3
Our Model	**49.7 (↑ 17.1%)**	**60.0 (↑ 5.0%)**	**85.2 (↑ 1.9%)**

Legal Judgment Prediction via Topological Learning

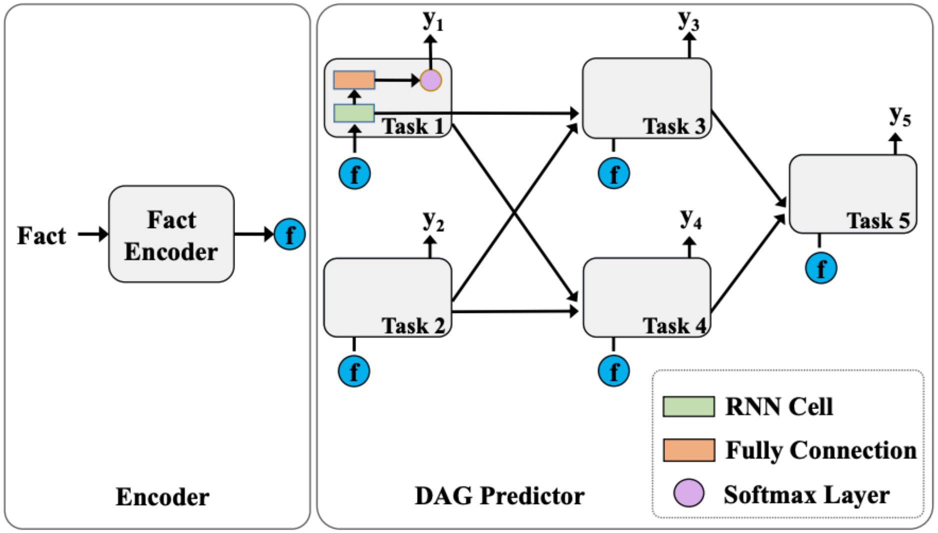
- Legal judgment usually consists of multiple subtasks, such as the decisions of applicable law articles, charges, and the term of penalty
- There exists topological dependencies among these subtasks



Zhong, Haoxi, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. Legal judgment prediction via topological learning. EMNLP 2018.

Legal Judgment Prediction via Topological Learning

- We formalize the dependencies among subtasks as a Directed Acyclic Graph (DAG) and propose a topological multi-task learning framework, which incorporates the DAG dependencies into judgment prediction



Zhong, Haoxi, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. Legal judgment prediction via topological learning. EMNLP 2018.

Experiment Results

- We evaluate the performance on three LJP subtasks, including law articles, charges and the terms of penalty. The proposed model outperforms other baselines significantly on most subtasks

	Tasks	Law Articles				Charges				The Term of Penalty			
	Metrics	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1
Single	TFIDF+SVM	82.4	45.5	26.7	30.2	82.2	47.4	27.9	31.3	48.5	36.0	16.7	16.5
	CNN	92.5	46.9	38.4	40.0	92.3	41.2	32.3	33.7	57.4	35.6	22.2	22.7
	HLSTM	91.4	38.6	37.3	36.9	91.8	37.8	36.0	35.2	56.1	22.5	25.0	23.3
Multi	Fact-Law Att.	93.5	50.9	45.6	45.9	93.4	47.2	41.4	41.5	56.3	31.3	26.4	26.7
	PM	93.7	51.9	44.1	44.9	93.6	45.5	39.1	39.3	58.2	38.2	24.9	26.8
	CNN-MTL	94.3	53.0	46.0	46.9	94.1	48.5	41.7	42.5	58.7	39.9	28.8	29.4
	HLSTM-MTL	92.4	45.5	41.4	41.0	92.3	41.9	36.6	35.9	54.9	30.6	26.6	26.4
Ours	**TopJudge**	**94.4**	**53.9**	**47.3**	**48.2**	**94.9**	**53.9**	**48.2**	**49.1**	**58.8**	**40.2**	**32.9**	**32.8**

Zhong, Haoxi, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. Legal judgment prediction via topological learning. EMNLP 2018.

Summary

- Legal domain has large-scale and well-formatted text for NLP, requiring specific domain knowledge

http://cail.cipsc.org.cn/

CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction

Chaojun Xiao[1*] Haoxi Zhong[1*] Zhipeng Guo[1] Cunchao Tu[1] Zhiyuan Liu[1] Maosong Sun[1] Yansong Feng[2] Xianpei Han[3] Zhen Hu[4] Heng Wang[4] Jianfeng Xu[5]
[1]Department of Computer Science and Technology, Tsinghua University, China
[2]Institute of Computer Science and Technology, Peking University, China
[3]Institute of Software, Chinese Academy of Sciences, China
[4]China Justice Big Data Institute
[5]Supreme People's Court, China

Legal Judgment Prediction

中国法研杯 CAIL 2019
司法人工智能挑战赛
（中国电科“X+AI”系列挑战赛）

Legal Reading Comprehension

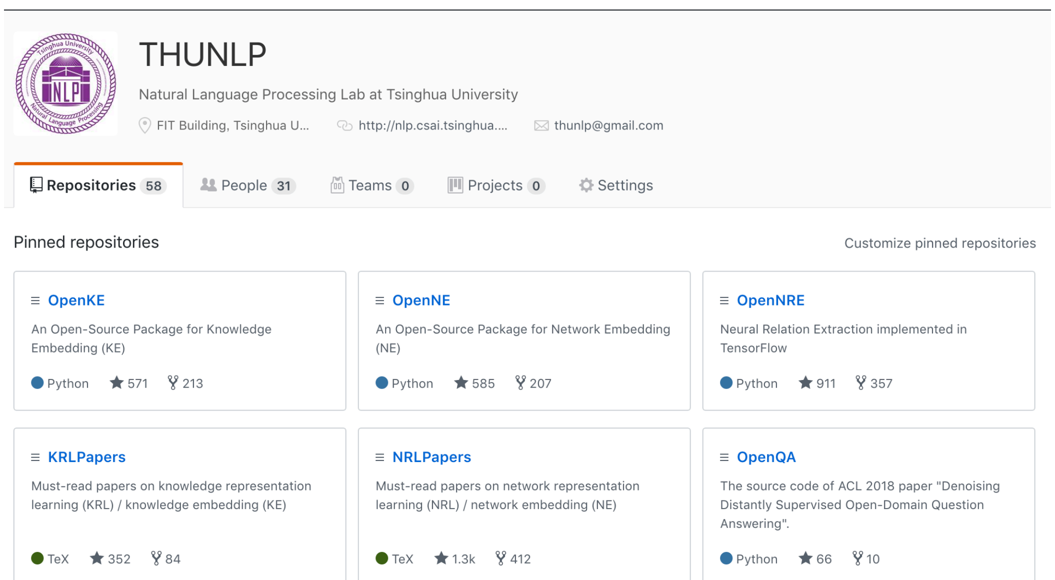
Legal Semantic Matching

Legal Information Extraction

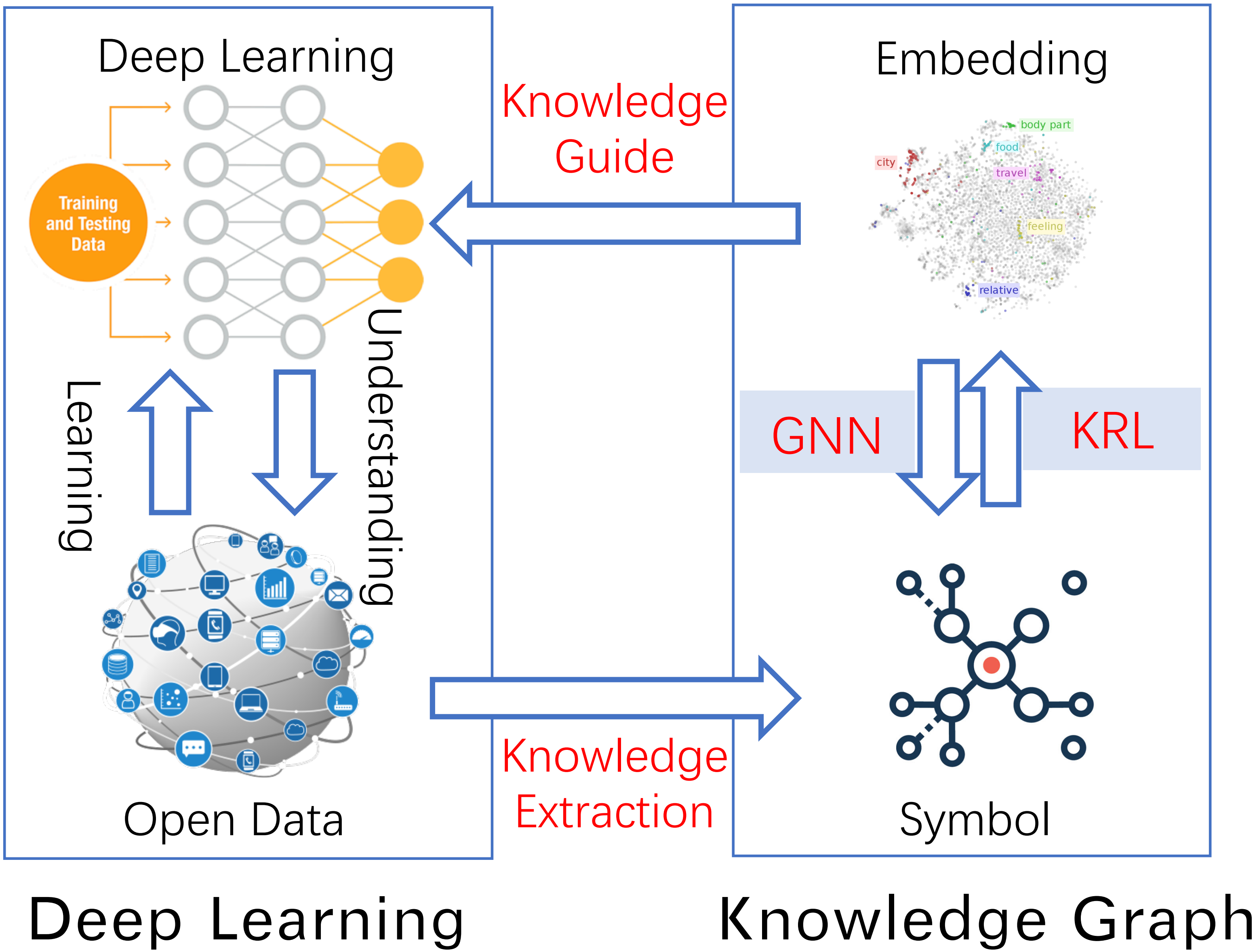
Open Source

- Packages for representation and acquisition of linguistic and world knowledge
- The projects obtain 23000+ stars on GitHub

https://github.com/thunlp



Overall Summary



Thanks!

http://nlp.csai.tsinghua.edu.cn/~lzy/

liuzy@Tsinghua.edu.cn