Common sense mining

常识挖掘

常识

- 常识是与大多数人通常共享的日常情况和事件相关的实践知识和推理
- 数量多、维度广、变化的、默认的

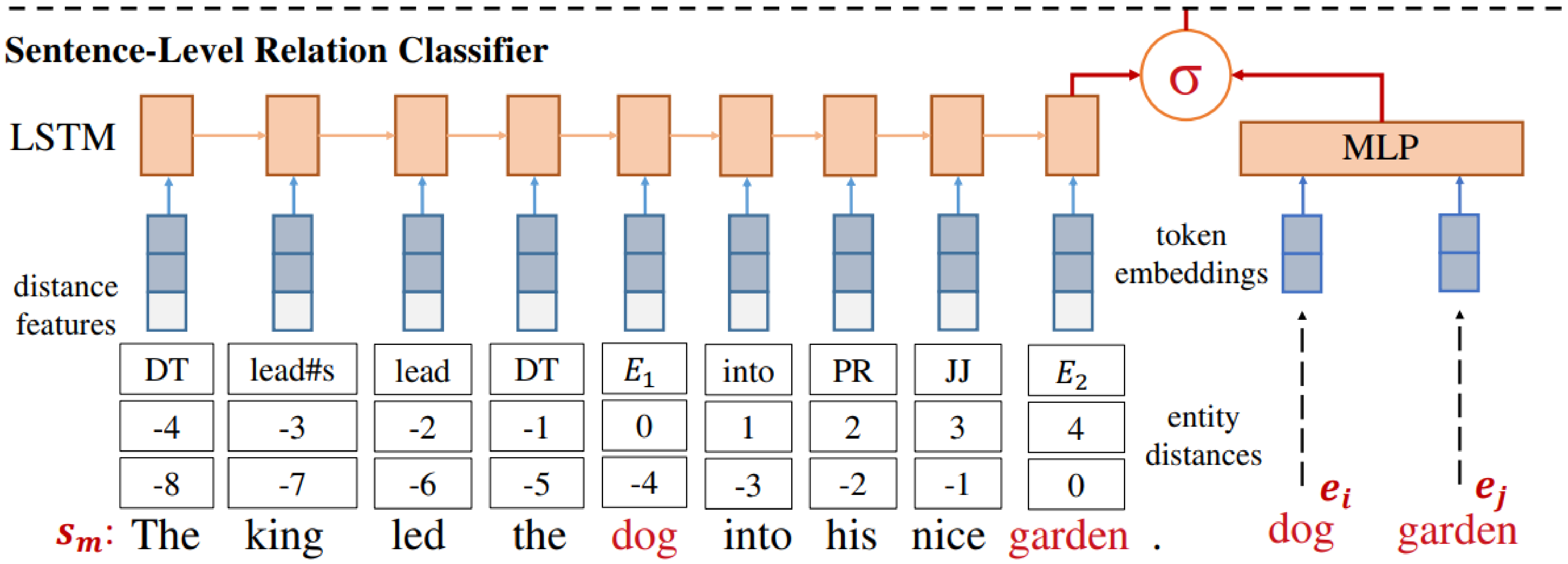
Automatic Extraction of Commonsense LocatedNear Knowledge

Frank F. Xu* **Bill Yuchen Lin*** **Kenny Q. Zhu**
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
{frankxu, yuchenlin}@sjtu.edu.cn, kzhu@cs.sjtu.edu.cn

ACL 2018

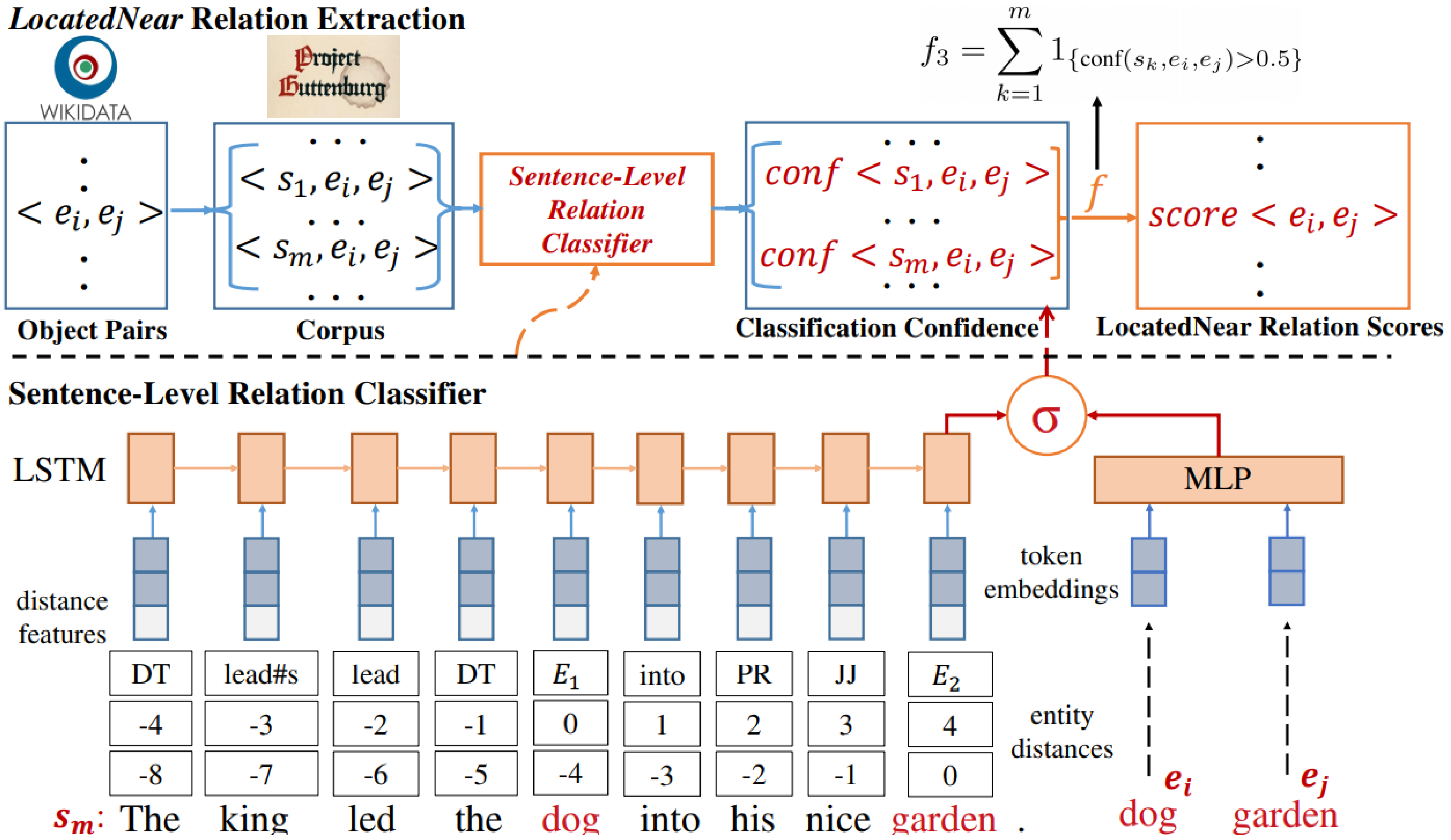
方位关系常识

- 描述现实生活中两个相邻物体的常识性知识
- 文本（句子）中获取方位常识



- 宾语对
- 与描述物理场景高度相关的动词、副词和介词—>原型
- 依赖关系:直接主语、宾语
- 语法分析符号：DT:determiner -限定词,CC-连词

Level	Examples
Objects	E_1, E_2
Lemma	open, lead, into, ...
Dependency Role	open#s, open#o, into#o, ...
POS Tag	DT, PR, CC, JJ, ...



LocatedNear Relation Extraction
WIKIDATA
Project Guttenburg
$< e_i, e_j >$
$< s_1, e_i, e_j >$
$< s_m, e_i, e_j >$
Sentence-Level Relation Classifier
$conf < s_1, e_i, e_j >$
$conf < s_m, e_i, e_j >$
f
$score < e_i, e_j >$
$f_3 = \sum_{k=1}^{m} 1_{\{\mathrm{conf}(s_k, e_i, e_j) > 0.5\}}$
Object Pairs
Corpus
Classification Confidence
LocatedNear Relation Scores
Sentence-Level Relation Classifier
LSTM
σ
MLP
distance features
token embeddings
DT
lead#s
lead
DT
E_1
into
PR
JJ
E_2
-4 -3 -2 -1 0 1 2 3 4
-8 -7 -6 -5 -4 -3 -2 -1 0
entity distances
s_m: The king led the dog into his nice garden .
e_i
dog
e_j
garden

评分函数

$$f_0 = m \tag{1}$$

$$f_1 = \sum_{k=1}^{m} \mathrm{conf}(s_k, e_i, e_j) \tag{2}$$

$$f_2 = \frac{1}{m} \sum_{k=1}^{m} \mathrm{conf}(s_k, e_i, e_j) \tag{3}$$

$$f_3 = \sum_{k=1}^{m} 1_{\{\mathrm{conf}(s_k, e_i, e_j) > 0.5\}} \tag{4}$$

$$f_4 = \frac{1}{m} \sum_{k=1}^{m} 1_{\{\mathrm{conf}(s_k, e_i, e_j) > 0.5\}} \tag{5}$$

数据集

- Wikidata →"物理对象"
- Gutenberg语料库：3036本英语书籍
  - 假设：小说中的句子更有可能描述现实生活场景
  - 佐证：提到至少两个实体的句子中，维基百科：32.4%肯定句，纽约时报：25.1%，Gutenberg语料库：55.1%
- 15193对出现在至少10个句子中的实体对
- 人工标注（500对，5000句，4000训练，1000测试）

实验结果

	Random	Majority	SVM	SVM(-BW)	SVM(-BPW)	SVM(-BAP)	SVM(-GF)
Acc.	0.500	0.551	0.584	0.577	0.556	0.563	**0.605**
P	0.551	0.551	0.606	0.579	0.567	0.573	**0.616**
R	0.500	1.000	0.702	0.675	0.681	**0.811**	0.751
F1	0.524	0.710	0.650	0.623	0.619	0.672	**0.677**

	SVM(-SDP)	SVM(-SS)	DRNN	LSTM+Word	LSTM+POS	LSTM+Norm	
Acc.	0.579	0.584	0.635	0.637	0.641	**0.653**	
P	0.597	0.605	**0.658**	0.635	0.650	0.654	
R	0.728	0.708	0.702	**0.800**	0.751	0.784	
F1	0.656	0.652	0.679	0.708	0.697	**0.713**	

f	MAP	P@50	P@100	P@200	P@300
f_0	0.42	0.40	0.44	0.42	0.38
f_1	0.58	**0.70**	0.60	0.53	**0.44**
f_2	0.48	0.56	0.52	0.49	0.42
f_3	**0.59**	0.68	**0.63**	**0.55**	**0.44**
f_4	0.56	0.40	0.48	0.50	0.42

Temporal Common Sense Acquisition with Minimal Supervision

Ben Zhou,[1] Qiang Ning,[2] Daniel Khashabi,[2] Dan Roth[1]
[1]University of Pennsylvania, [2]Allen Institute for AI
{xyzhou,danroth}@cis.upenn.edu {qiangn,danielk}@allenai.org

- ACL 2020

例子

Example 1: choosing from "*will*" or "*will not*"

Dr. Porter is now (*e1:taking*) a vacation and _________ be able to see you soon.

Dr. Porter is now (*e2:taking*) a walk outside and _________ be able to see you soon.

Jack rested for 2 hours before the speech.
(*Jack rested before the speech*, hours, **duration**)

We went to a bar last Friday.
(*We went to a bar*, Friday, **typical**)

Jane makes breakfast for herself everyday.
(*Jane makes breakfast for herself*, days, **frequency**)

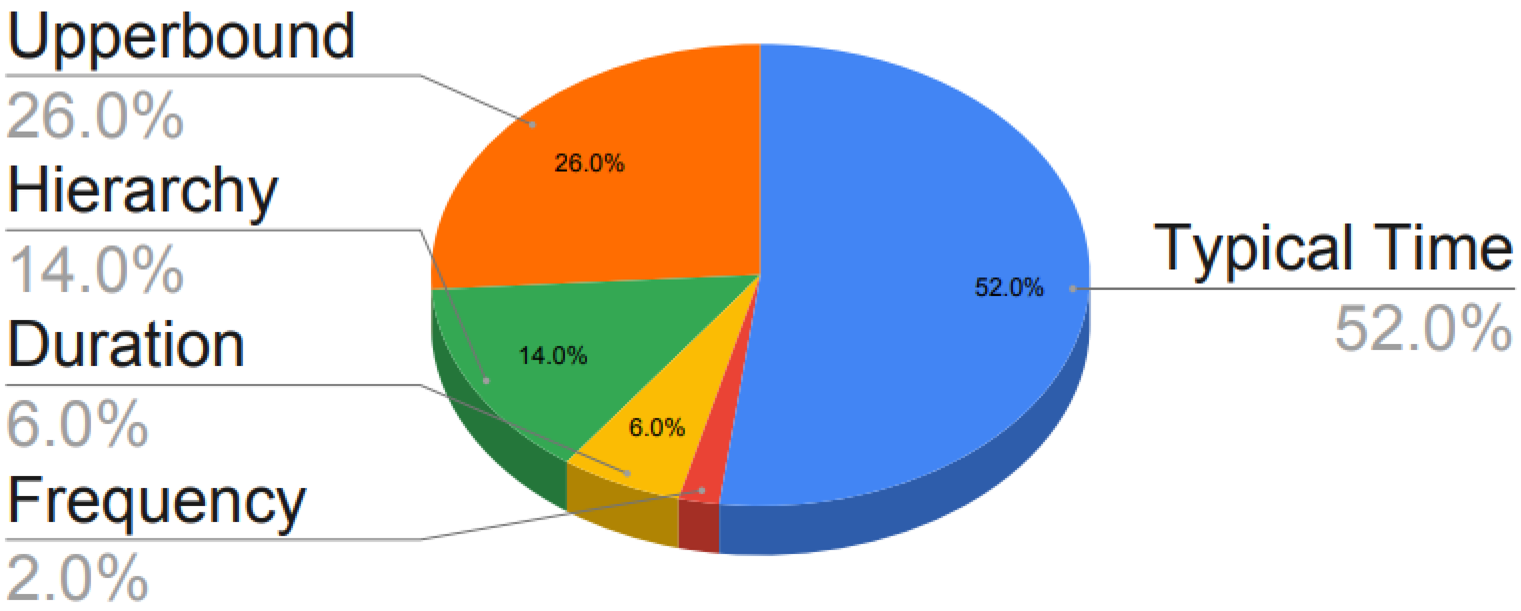
The city was surrounded by police yesterday.
(*The city was surrounded by police*, days, **upper-bound**)

The phone rang while I was in the bathroom.
(*The phone rang*, while I was in the bathroom, **hierarchy**)

- (event,value,dimension).

- Duration:for (“second,” “minute,” “hour,” “day,” “week,” “month,” “year,” “decade,” “century.”)
- Frequency: “every,” “per,” “once,” . . . “times.”+ duration (e.g., “week”).
- Typical Time: time of day (e.g., “morning” etc.), time of week (e.g., “Monday” etc.), month (e.g., “January” etc.) and season (e.g., “winter” etc.) + “until,” “since,” “following,”
- Duration Upper-bound: did [activity] yesterday； “in [temporal expression]” “next” (e.g., “the next day”), “last” (e.g., “last week”), “previous” (e.g., “previous month”), or “recent” (e.g., “recent years”)
- Event Relative Hierarchy： “before,” “after,” “during,” “while” and “when.”

各类数据占比



Upperbound
26.0%
Hierarchy
14.0%
Duration
6.0%
Frequency
2.0%
Typical Time
52.0%
26.0%
14.0%
6.0%
52.0%

Loss

$$\ell = -\sum_{i \in D} \mathbf{y}_i^\top \log(\mathrm{softmax}(\mathbf{x}_i)),$$

- Duration, Frequency, and Upper-bound：

$$\mathbf{y}[i] = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\mathrm{logsec}(l)-\mu)^2/2\sigma^2}$$

  - Logsec：“minute” → 60 → 4.1， σ = 4
    - μ = logsec(g)

- Typical time： 1-2-⋯-7-1， 设距离=1， 标准差=0.5的正态分布

- Hierarchy： one-hot

效果对比

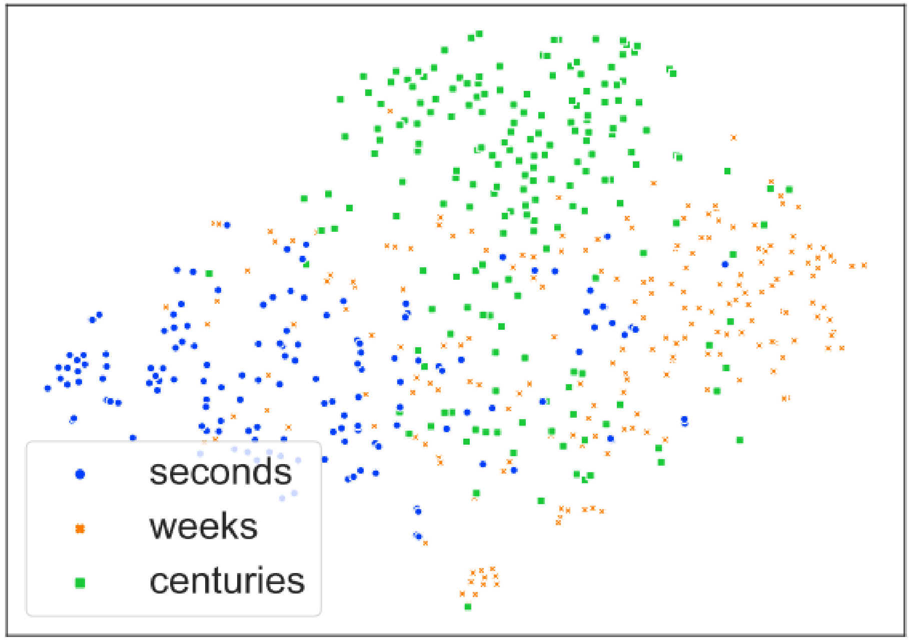


Figure 5: Representations of events (whose durations were labeled as seconds, weeks, or centuries) obtained from the original BERT base model.

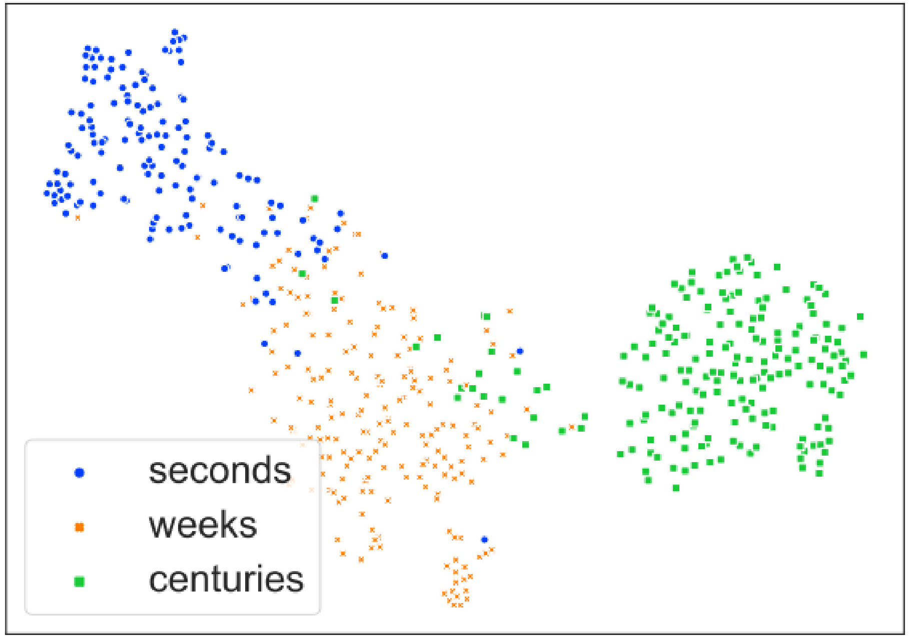
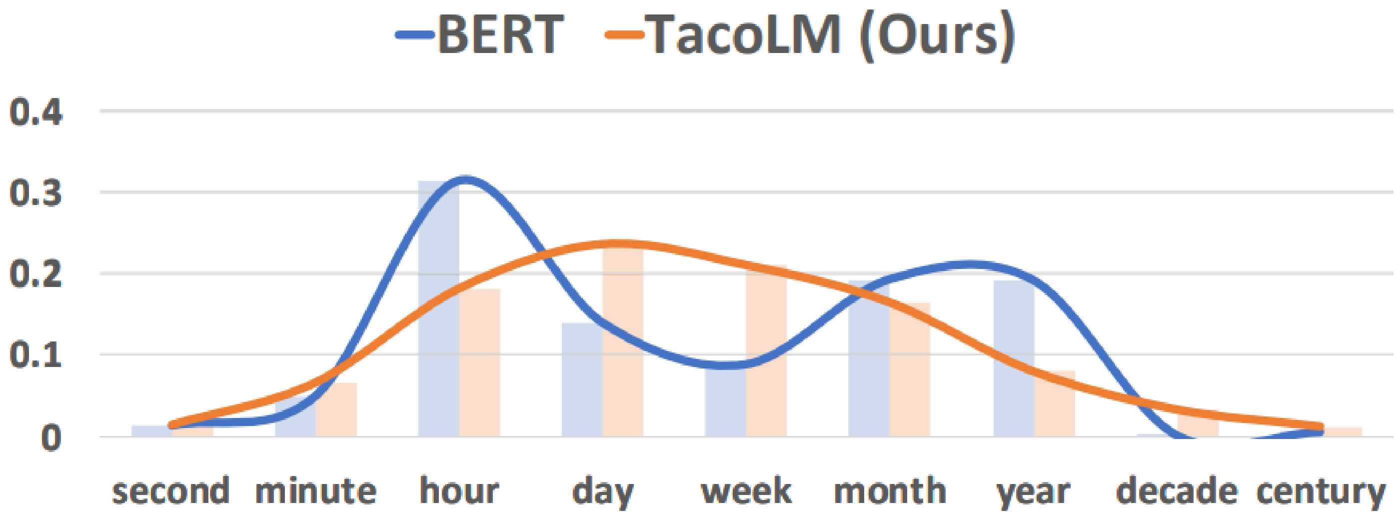


Figure 6: Representations of the same set of events as in Fig. 5 obtained from the proposed method.

效果对比


BERT
TacoLM (Ours)
0.4
0.3
0.2
0.1
0
second
minute
hour
day
week
month
year
decade
century

实验结果

- RealNews corpus： 1742-300 UDST dataset： 1047-142

Systems	RealNews						UDS-T
			Typical Time				
	Duration	Freq	Day	Week	Month	Season	Duration
BERT	1.33	1.68	1.75	1.53	3.78	0.87	1.77
BERT + naive finetune	1.21	1.45	1.47	1.28	3.28	1.08	2.06
TACOLM (ours)	**0.75**	**1.17**	**1.72**	**1.19**	**3.42**	**0.63**	**1.49**
TACOLM (ours), normalized	8%	13%	22%	17%	29%	16%	17%
TACOLM -ADJ	0.84	1.20	1.82	1.08	2.47	0.74	1.68
TACOLM -SL	0.77	1.30	1.88	1.06	2.50	0.74	1.50
TACOLM -AUX	0.77	1.28	1.61	1.31	3.25	0.78	1.51
TACOLM -MS	0.84	1.12	1.82	1.5	3.17	0.61	1.69
TACOLM -AM	0.68	1.20	1.86	1.31	3.11	0.70	1.58

Commonsense Knowledge Mining from Pretrained Models

Joshua Feldman*, **Joe Davison***, **Alexander M. Rush**
School of Engineering and Applied Sciences
Harvard University

- EMNLP 2019

三元组→自然语言

- (musician, CapableOf, play musical instrument)

⬇

- H（musician） can T（play musical instrument）

⬇ GPT-2 117M LM

Candidate Sentence S_i	$\log p(S_i)$
"musician can playing musical instrument"	-5.7
"musician can be play musical instrument"	-4.9
"musician often play musical instrument"	-5.5
"a musician can play a musical instrument"	$\mathbf{-2.9}$

计算置信度

$$\mathbf{PMI}(\mathbf{t}, \mathbf{h}|r) = \log p(\mathbf{t}|\mathbf{h}, r) - \log p(\mathbf{t}|r)$$

- A musician can play a musical instrument
- P(t|h,r) = A musician can play a k_1 k_2
- P(t|r) = A k_{t1} can play a k_{h1} k_{h2}

Task 1: Commonsense Knowledge Base Completion

- 数据集 2400正例、2400负例
  - 正例：ConceptNet 5数据集中的众包条目（OMCS）
  - 负例：随机替换三元组中的一个元素

Task 2: Mining Wikipedia

- 10关系 × 300三元组
- 两个标注者评分0-4
- 模型选出置信度最高的100个三元组，平均得分3.0
- 优于当前有监督模型

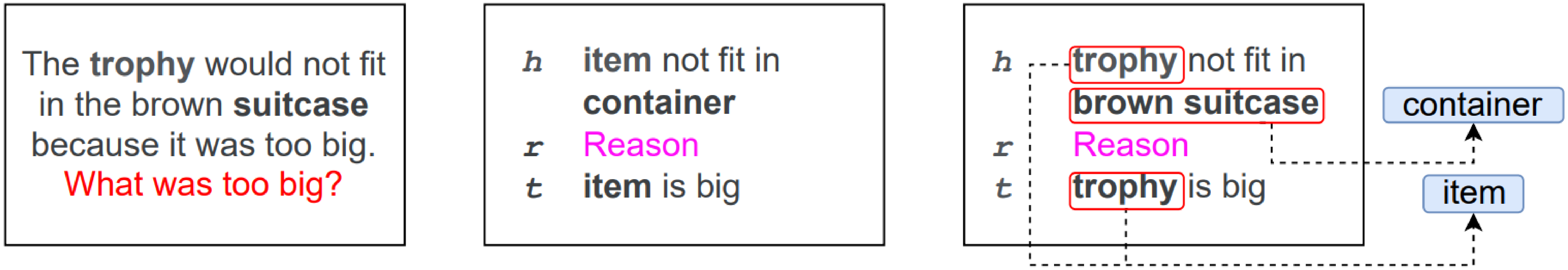
实验结果

Model	**Task 1**	**Task 2**
Unsupervised		
CONCATENATION	68.8	2.95 ± 0.11
TEMPLATE	72.2	2.98 ± 0.11
TEMPL.+GRAMMAR	74.4	2.56 ± 0.13
COHERENCY RANK	78.8	$\mathbf{3.00} \pm 0.12$
Supervised		
DNN	**89.2**	2.50
FACTORIZED	89.0	2.61
PROTOTYPICAL	79.4	2.55

On the Role of Conceptualization in Commonsense Knowledge Graph Construction

Mutian He[1], **Yangqiu Song**[1], **Kun Xu**[2], **Yu Dong** [2]
[1]Hong Kong University of Science and Technology
[2]Tencent
{mhear, yqsong}@cse.ust.hk
{kxkunxu,dyu}@tencent.com

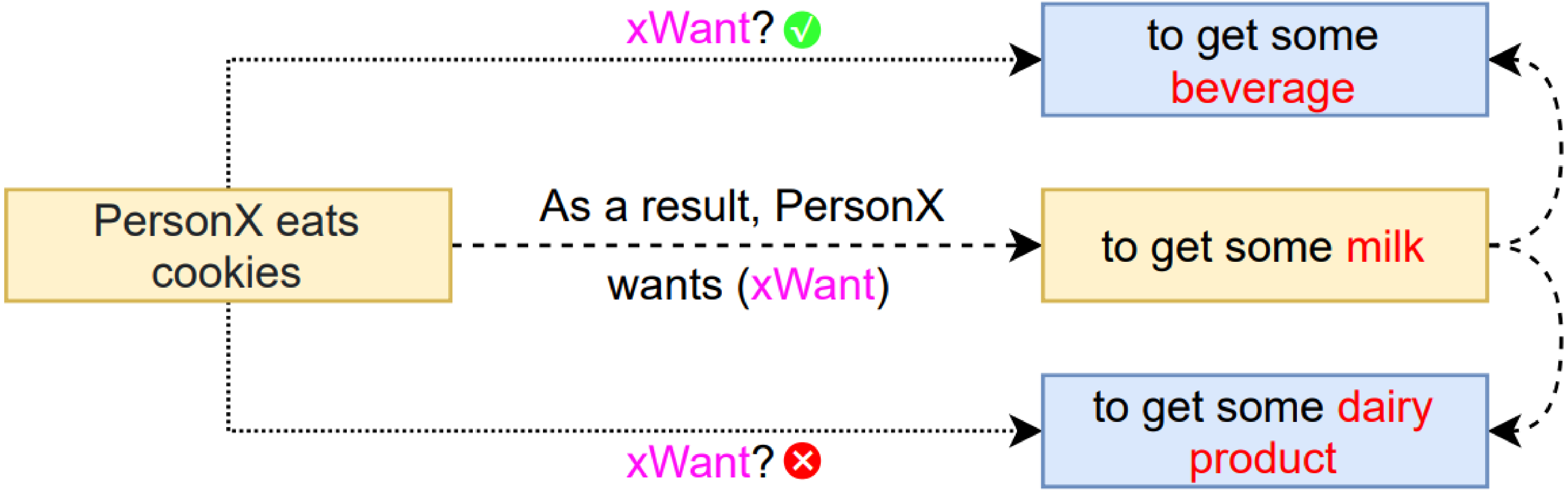
概念化



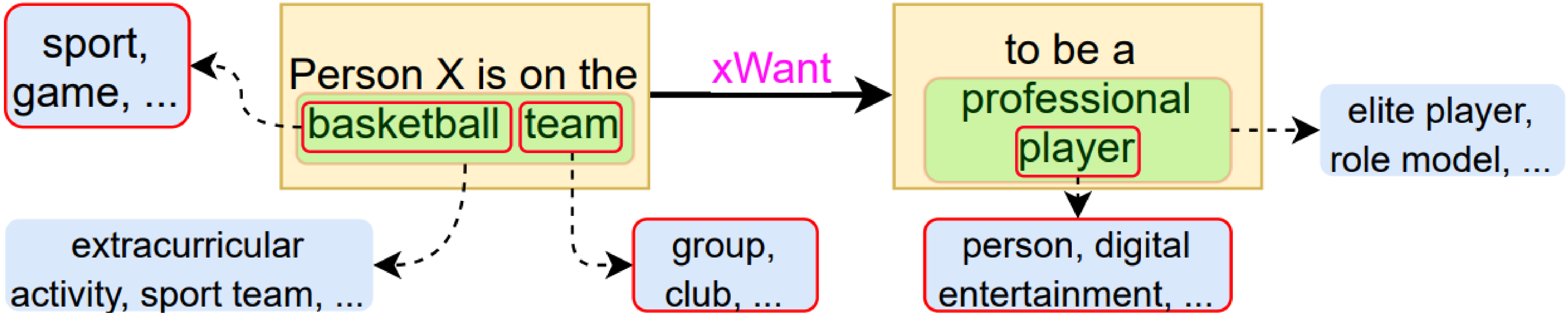
(a) A commonsense reasoning problem

(b) An existent triple in the CKG

(c) Conceptualization on the extracted triple


xWant?
to get some beverage
PersonX eats cookies
As a result, PersonX
wants (xWant)
to get some milk
to get some dairy product
xWant?

<h, IsA, t,f>



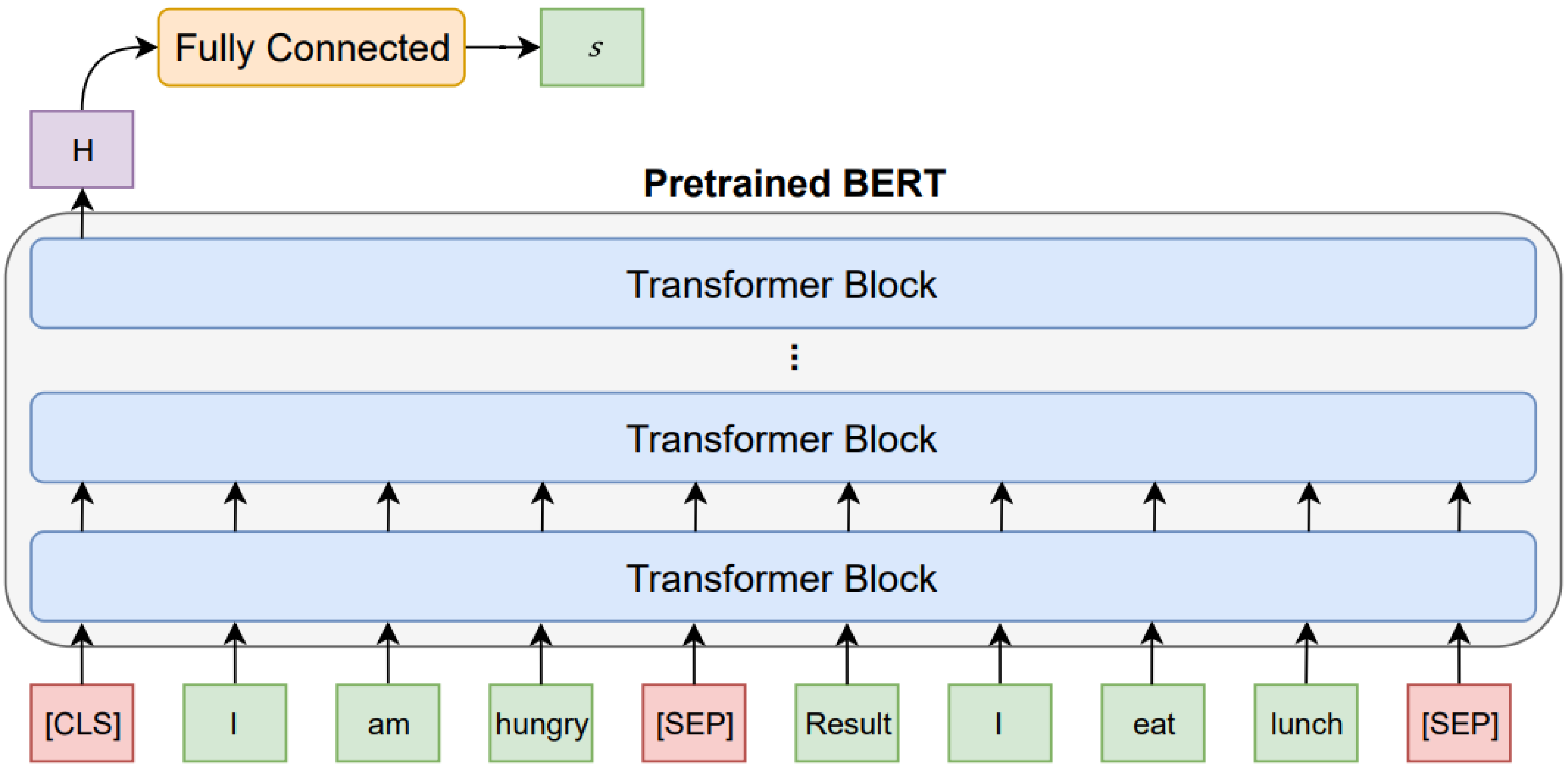
- Probase数据集：17.9M nodes and 87.6M edges

生成负例

- Node Substitution (NS)：随即替换h、t

- Entity Conceptualization (EC):根据f，将h或t概念化或具体化，再替换

$$L = - \sum_{(x,y)\in D_+ \cup D_-} (y \log s + (1-y)\log(1-s)).$$



实验结果

	ASER		ATOMIC	
	EC	NS	EC	NS
COMeT	0.6388	0.5869	0.6927	0.5730
KG-BERT	0.7091	**0.8018**	0.7669	0.6575
CCC-50	0.8716	0.7775	0.9016	**0.7840**
CCC-75	0.8995	0.7250	0.9221	0.7446
CCC-87.5	**0.9156**	0.6635	**0.9355**	0.6980
CCC-75-scratch	0.8284	0.5587	0.8579	0.5003
CCC-75-RoBERTa	0.8999	0.6938	0.9305	0.7350