组会

曾双

2020.11.14

- Learning from Context or Names? An Empirical Study on Neural Relation Extraction，EMNLP 2020

- Relation Extraction as Two-way Span-Prediction， ArXiv 2020，TACRED & SemEval SOTA

EMNLP 2020

Learning from Context or Names? An Empirical Study on Neural Relation Extraction

Hao Peng[1*], **Tianyu Gao**[2*], **Xu Han**[1], **Yankai Lin**[3], **Peng Li**[3], **Zhiyuan Liu**[1†], **Maosong Sun**[1], **Jie Zhou**[3]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]Princeton University, Princeton, NJ, USA
[3]Pattern Recognition Center, WeChat AI, Tencent Inc, China
{h-peng17,hanxu17}@mails.tsinghua.edu.cn, tianyug@princeton.edu

Motivation

- 目前RE领域对下面两个问题都没有清晰的理解：

（1）哪些信息影响着现有的RE模型去做决策？（为什么？）

文本上下文和实体mention（名字、实体类别、其他）

（2）如何让现有模型的效果得到更进一步的提升？（怎么做？）

提出预训练RE的新方法，说明预训练这个方向是可靠的

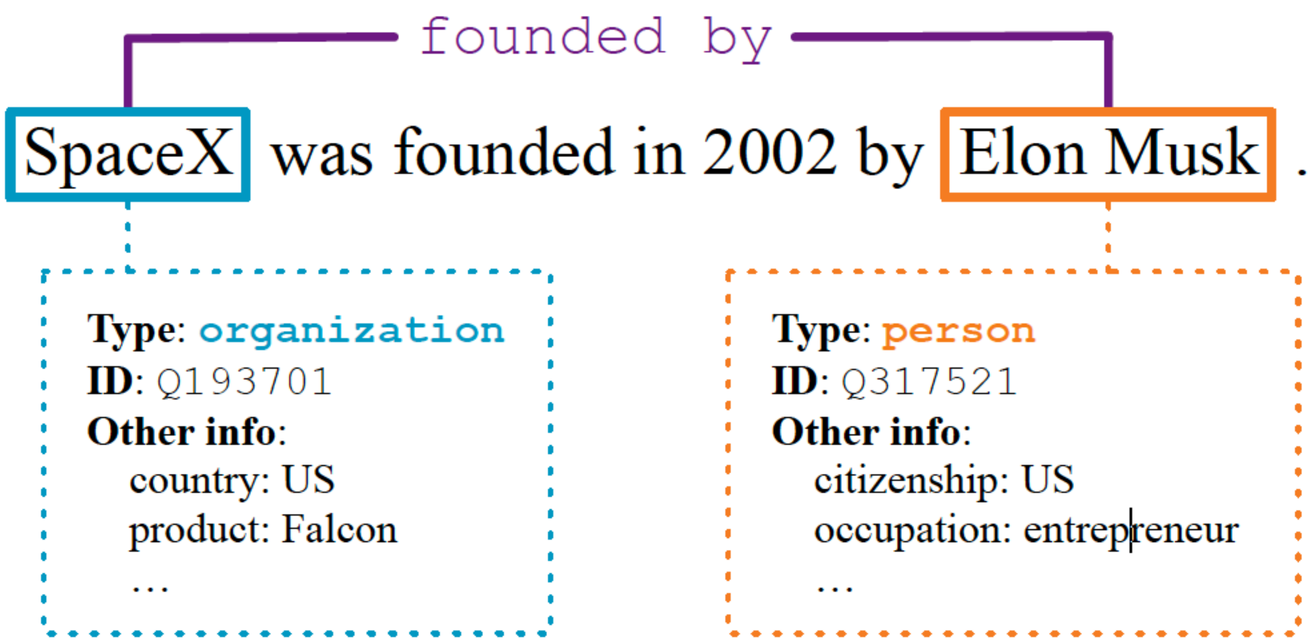

Figure 1: An example for the information provided by textual context and entity mentions in a typical RE scenario. From mentions, we can acquire type information and link entities to KGs, and access further knowledge about them. The IDs in the figure are from Wikidata.

Model	C+M	C+T	OnlyC	OnlyM	OnlyT
CNN	0.547	0.591	0.441	0.434	0.295
BERT	0.683	0.686	0.570	**0.466**	0.277
MTB	**0.691**	**0.696**	**0.581**	0.433	**0.304**

Table 1: TACRED results (micro F_1) with CNN, BERT and MTB on different settings.

- Context+Mention (C+M)：RE任务最常见的设定，句子加上实体mention
- Context+Type (C+T)：将C+M 中的mention 使用它们的type 代替，类似于BERT 中的[MASK]，比如[person]
- Only Context (OnlyC)：把句子中的头尾实体分别用[SUBJ]和[OBJ]代替
- Only Mention (OnlyM)：只使用头尾实体的mention去做关系分类。
- Only Type (OnlyT)：和OnlyM类似，但是只提供实体类别。

Case分析

C+M
Although her family was from Arkansas, *she* was born in *Washington* state, where ... **Label**: **per:state_of_birth** **Prediction**: **per:state_of_residence**
Dozens of lightly regulated subprime lenders, including New Century Financial Corp., have failed and troubled *Countrywide Financial Corp.* was acquired by *Bank of America Corp.* **Label**: **org:parents** **Prediction**: **no_relation**

C+T
First, *Natalie Hagemo* says, *she* fought the Church of Scientology just to give birth to her daughter. **Label**: **no_relation** **Prediction**: **per:children**
Earlier this week Jakarta hosted the *general assembly* of the *Organisation of Asia-Pacific News Agencies*, ... **Label**: **no_relation** **Prediction**: **org:members**
The boy, identified by the Dutch foreign ministry as *Ruben* but more fully by Dutch media as *Ruben van Assouw*, ... **Label**: **per:alternate_names** **Prediction**: **no_relation**

Table 2: Wrong predictions made only by C+M and only by C+T, where red and blue represent subject and object entities respectively. As the examples suggest, C+M is more easily biased by the entity distribution in the training set and C+T loses some information from mentions that helps to understand the text.

Case分析

- C + M 与 C + T 共享了95.7%的预测正确的样本，C + M 中68.1%的错误预测和 C + T 一样，这说明了很多模型利用实体mention主要利用的是类别信息。

- 除了mention的type以外，entity还提供了其他的信息。

Case分析

Type	Example
Wrong 42%	..., *Jacinto Suarez*, Nicaraguan deputy to the *Central American Parliament* (PARLACEN) said Monday. Label: `org:top_members/employees` Prediction: `no_relation` US life insurance giant MetLife said on Monday it will acquire *American International Group* unit American Life Insurance company (*ALICO*) in a deal worth 155 billion dollars. Label: `org:subsidiaries` Prediction: `no_relation`
No pattern 31%	On Monday, the judge questioned the leader of the *Baptist* group, *Laura Silsby*, who ... Label: `per:religion` Prediction: `no_relation`
Confusing 27%	About a year later, *she* was transferred to Camp Hope, *Iraq*. Label: `per:countries_of_residence` Prediction: `per:stateorprovinces_of_residence`

Table 3: Case study on unique wrong predictions made by OnlyC (compared to C+M). We sample 10% of the wrong predictions, filter the wrong-labeled instances and manually annotate the wrong types to get the proportions. We use red and blue to highlight the subject and object entities.

Case分析

- Wrong：存在明显的pattern，但是模型没有理解到
- No pattern：mask掉entity pair后，对于人来说都很难判断的关系
- Confusing：mask掉entity pair后，有歧义的关系

- 这个case由于是Only C相对于C + M特有的，对于42%的Wrong，说明C + M模型确实是依赖了Mention中的特有模式（shallow heuristics），而没有利用上context的信息

结论

- 尽管上下文是支持预测结果的主要源头，RE模型也严重地依赖实体mention的信息，这些信息主要是实体的类别；

- 现有的数据可能通过实体mention泄露了浅层的启发，导致了RE模型效果都这么好

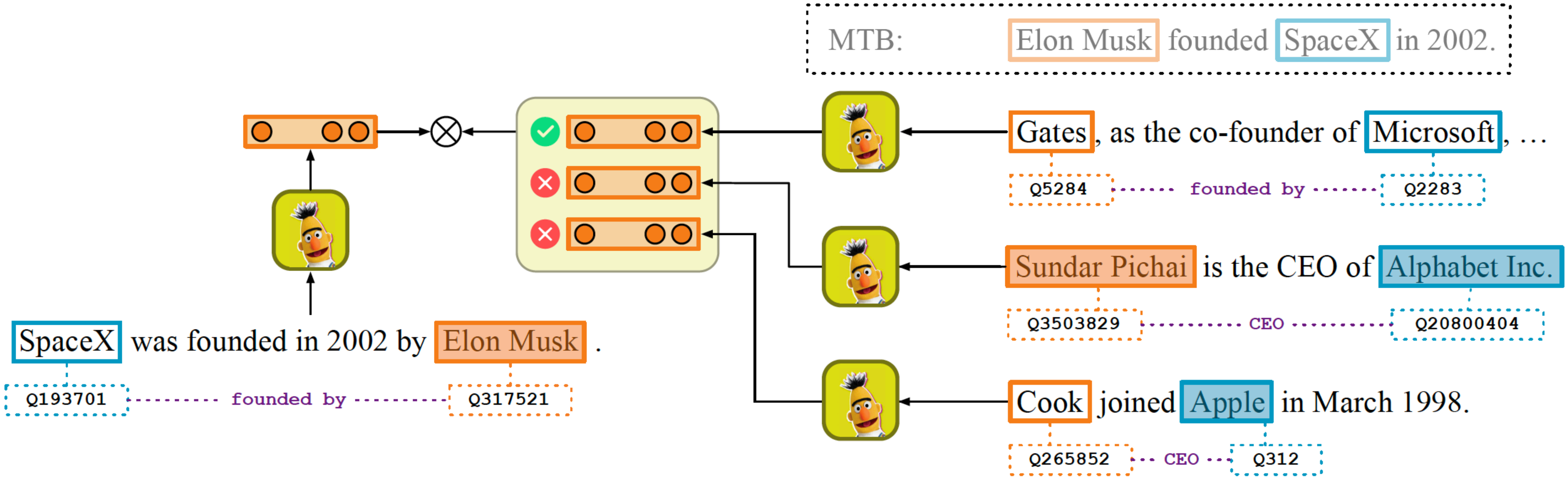

Figure 2: Our contrastive pre-training framework for RE. We assign relations to sentences by linking entity pairs in sentences to Wikidata and checking their relations in the KG. We assume that sentences with the same relation should have similar representations, and those with different relations should be pushed apart. Entity mentions are randomly masked (boxes with colored background) to avoid simple memorization. Compared to MTB (in the dotted box), our method samples data with better diversity, which can not only increase the coverage of entity types and diverse context but also reduce the possibility of memorizing entity names.

$$\mathbf{x} = \mathrm{ENC}_h(x) \oplus \mathrm{ENC}_t(x), \tag{1}$$

$$\mathcal{L}_{CP} = -\log \frac{e^{\mathbf{x}_A^T \mathbf{x}_B}}{e^{\mathbf{x}_A^T \mathbf{x}_B} + \sum_{i=1}^{i \leq N} e^{\mathbf{x}_A^T \mathbf{x}_B^i}}. \tag{2}$$

$$\mathcal{L} = \mathcal{L}_{CP} + \mathcal{L}_{MLM}. \tag{3}$$

Dataset	# Rel.	# Inst.	% N/A
TACRED	42	106,264	79.5%
SemEval-2010 Task 8	19	10,717	17.4%
Wiki80	80	56,000	-
ChemProt	13	10,065	-
FewRel	100	70,000	-

Dataset	Model	1% C+M	1% OnlyC	1% OnlyM	10% C+M	10% OnlyC	10% OnlyM	100% C+M	100% OnlyC	100% OnlyM
TACRED	BERT	0.211	0.167	0.220	0.579	0.446	0.433	0.683	0.570	**0.466**
	MTB	0.304	0.231	0.308	0.608	0.496	0.441	0.691	0.581	0.433
	CP	**0.485**	**0.393**	**0.350**	**0.633**	**0.515**	**0.453**	**0.695**	**0.593**	0.450
SemEval	BERT	0.367	0.294	0.245	0.772	0.688	0.527	0.871	0.798	0.677
	MTB	0.362	0.330	**0.249**	0.806	0.744	0.543	0.873	0.807	**0.682**
	CP	**0.482**	**0.470**	0.221	**0.822**	**0.766**	**0.543**	**0.876**	**0.811**	0.679
Wiki80	BERT	0.559	0.413	0.463	0.829	0.413	0.655	0.913	0.810	0.781
	MTB	0.585	0.509	0.542	0.859	0.509	0.719	0.916	0.820	0.788
	CP	**0.827**	**0.734**	**0.653**	**0.893**	**0.734**	**0.745**	**0.922**	**0.834**	**0.799**
ChemProt	BERT	0.362	0.362	0.362	0.634	0.584	0.385	0.792	0.777	0.463
	MTB	**0.362**	0.362	**0.362**	0.682	0.685	0.403	0.796	0.798	0.463
	CP	0.361	**0.362**	0.360	**0.708**	**0.697**	**0.404**	**0.806**	**0.803**	**0.467**

Table 5: Results on supervised RE datasets TACRED (micro F_1), SemEval (micro F_1), Wiki80 (accuracy) and ChemProt (micro F_1). 1% / 10% indicate using 1% / 10% supervised training data respectively.

贡献

- 让模型去学习表达相同relation的context应该具有相同的表示（这个表示就是句子entity pair 的拼接），可以一定程度上从多样化的角度学习到关系的不同表达。
- Mask 掉entity 可以避免模型对entity 的死记硬背，鼓励模型去利用entity 的type 信息。

借鉴

- 从pilot experiment中分析模型是如何学习的

- 通过case来对比不同信息的重要性

ArXiv 2020

Relation Extraction as Two-way Span-Prediction

Amir DN Cohen
Bar Ilan University
amirdnc@gmail

Shachar Rosenman
Bar Ilan University
shacharosn@gmail

Yoav Goldberg
Bar Ilan University
yoav.goldberg@gmail.com

Motivation

- 将关系分类(RC)任务变成QA任务，作者认为传统的RE架构存在局限性（关系类别给定），将其替换为span prediction的QA架构，是更优的架构。
- 针对数据集对每个关系定义了问题模板
- 在TACRED和SemEval上取得了最新的SOTA。

QA v.s. RC

- 编码方式不同
- QA可以有更多的语义信息注入输入端
- QA模型学习难度更大，不光要预测关系，还要预测span

Reducing RC to Span-prediction

given the RC instance:

RC:$(c, Sam, 1991) \mapsto$ *date-of-birth*

we create the two QA instances:

QA1:$(c,$ *When was Sam born?*$) \mapsto$ *1991*

QA2:$(c,$ *Who was born in 1991?*$) \mapsto$ *Sam*

Results

Model	P	R	F_1
TACRED			
MTB(BERT)	-	-	70.1
token-TACRED (ours, BERT)	63.3	78.4	70
relation-TACRED (ours, BERT)	67	76	71.2
QA-TACRED (ours, BERT)	71.1	72.6	**71.8**
KEPLLER (RoBERTa + KG, sota)	72.8	72.2	72.5
token-TACRED (ours, ALBERT)	72.2	74.6	73.4
relation-TACRED (ours, ALBERT)	74.6	75.2	**74.8**
QA-TACRED (ours, ALBERT)	73.3	71.8	72.6
QA-TACRED (only head q, ALBERT)	75.8	65.4	70.2
SemEval			
MTB (BERT)	-	-	89.2
QA-SemEval (ours, BERT)	90.7	93.2	**91.9**
LiTian (sota)	94.2	88.0	91.0

Table 1: Supervised results on the TACRED and SemEval datasets.

Question Templates

Relation Name	Question
per:date_of_birth	Q1: When was e_1 born? Q2 Who was born in e_2?
per:title	Q1: What is e_1's title? Q2 Who has the title e_2
org:top_members/employees	Q1: Who are the top members of the organization e_1? Q2 What organization is e_2 a top member of?
org:country_of_headquarters	Q1: In what country the headquarters of e_1 is? Q2 What organization have it's headquarters in e_2?
per:parents	Q1: Who are the parents of e_1? Q2 Who are the children of e_2?
per:age	Q1: What is e_1's age? Q2 Whose age is e_2?
per:countries_of_residence	Q1: What country does e_1 resides in? Q2 Who resides in country e_2?
per:children	Q1: Who are the children of e_1? Q2 Who are the parents of e_2?

借鉴

- 将RE降解为QA的方式，可以作为一种zero shot的手段学习没有见过的关系。

Thanks!