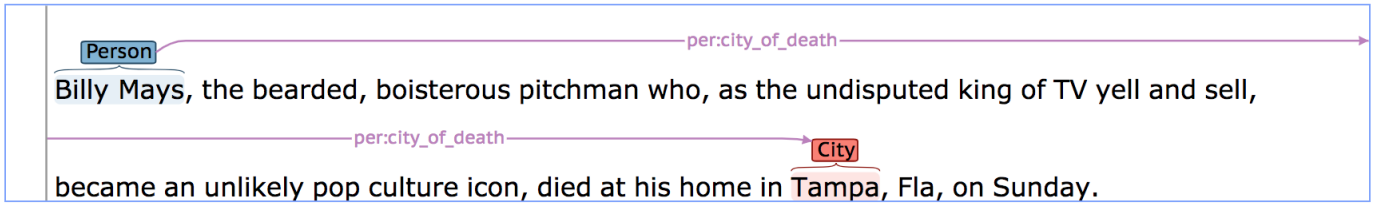
An Improved Baseline for Sentence-level Relation Extraction

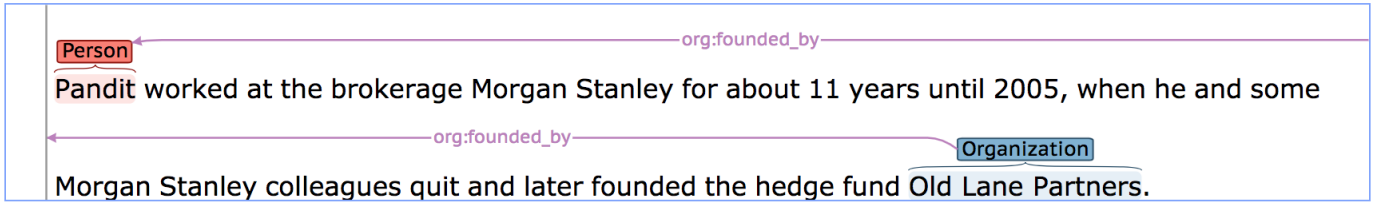
Wenxuan Zhou, Muhao Chen
University of Southern California

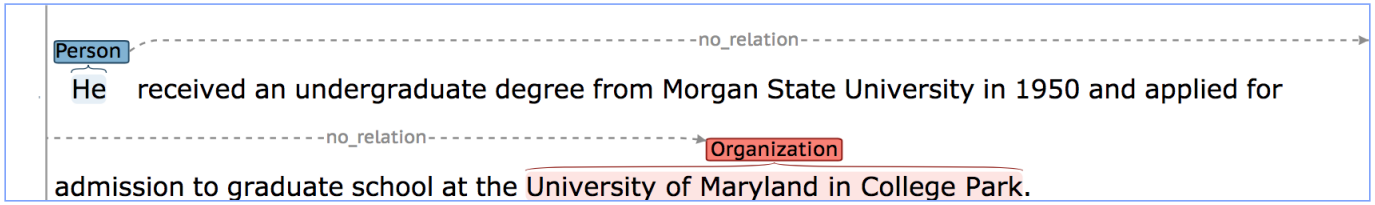
USC Viterbi
School of Engineering

Sentence-Level RE

Given a sentence and a pair of entities, predict the relationship(s) between the entities from a pre-defined set of relationships of interest.

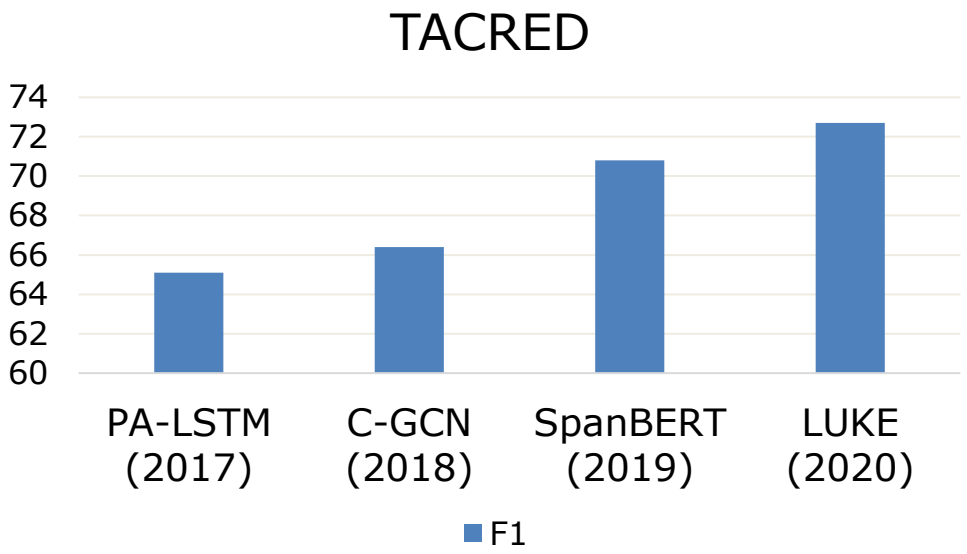






Sentences not containing any predefined relations are **NA**

Research Progress in RE



~8% increase, but still far from perfect

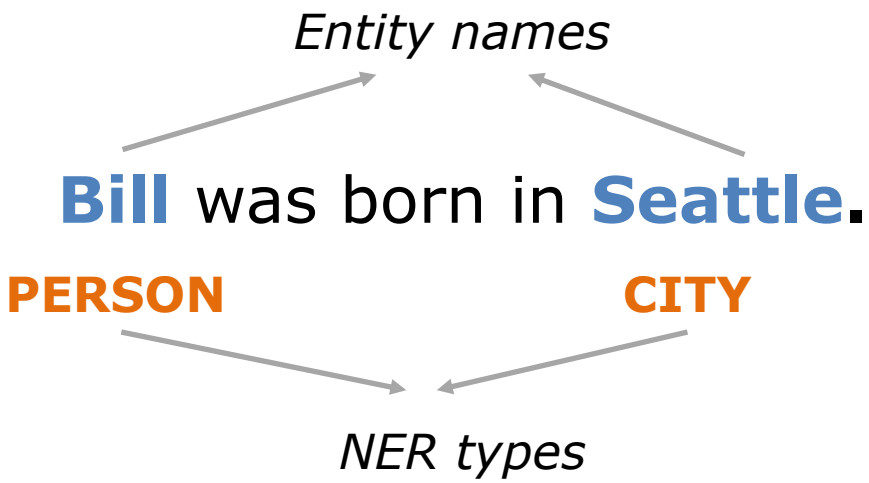
What part is missing from a promising RE system?

Overview

We discuss two obstacles that have hindered the performance of existing RE models:

a. RE task provides side information of entities, including entity types, names, and spans, while existing RE models fall short of representing all of them in input.
b. Existing datasets contain a large portion of noisy-labeled or ill-defined labels, causing model performance to be underestimated.

Typed Entity Markers



Entity mask (PA-LSTM, C-GCN):

[SUBJ-PERSON] was born in **[OBJ-CITY]**.

Entity marker (LUKE):

[E1] **Bill** [/E1] was born in [E2] **Seattle** [/E2].

Does not include both entity name and type.

Typed Entity Markers (Cont.)

Represent both entity names and types in the inputs.

Typed Entity Marker (Zhong and Chen, 2021):

<S:PERSON> Bill </S:PERSON> was born in **<O:CITY> Seattle </O:CITY>**.

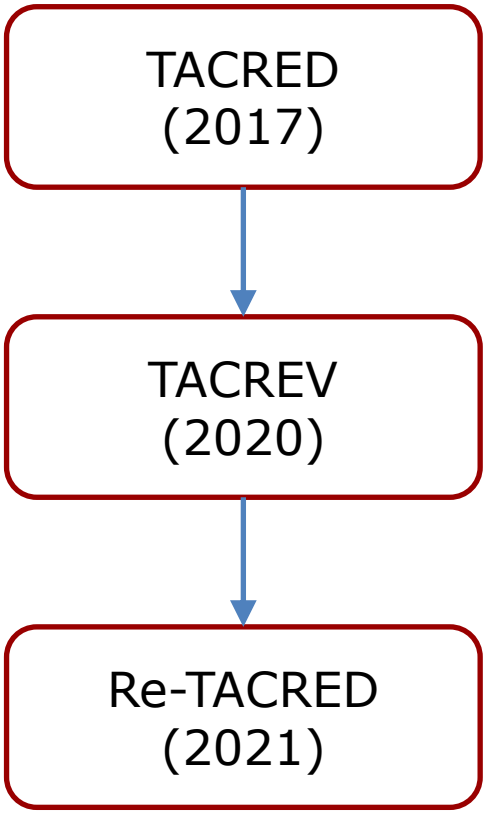
Typed Entity Marker (punct, ours):

@ * **person** * **Bill** @ was born in # ∧ **city** ∧ **Seattle** #.

Avoid adding new special tokens.

Reference:
Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for joint entity and relation extraction. NAACL 2021.

Data

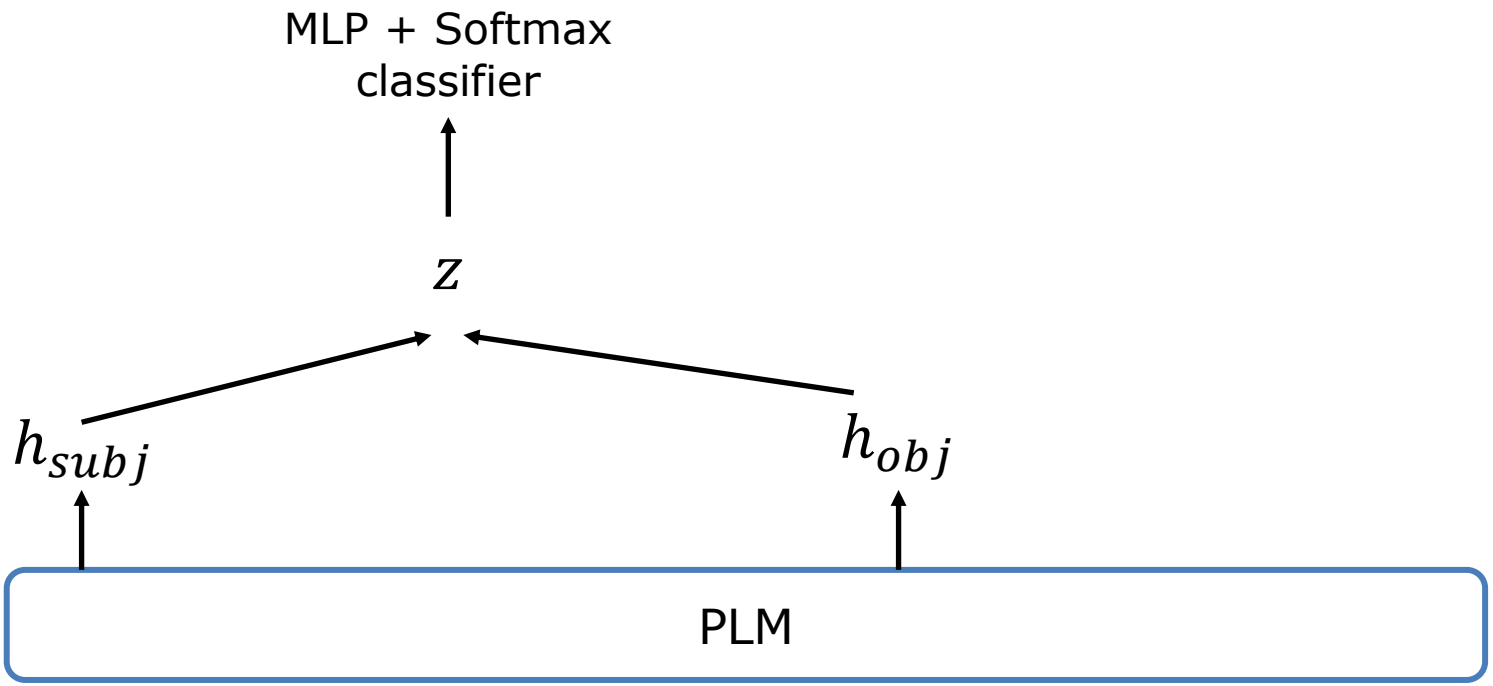


Crowdsourced, contains ~6.6% noisy labels.

Relabeled dev and test split.

Relabeled the whole dataset and revised ill-defined relations.

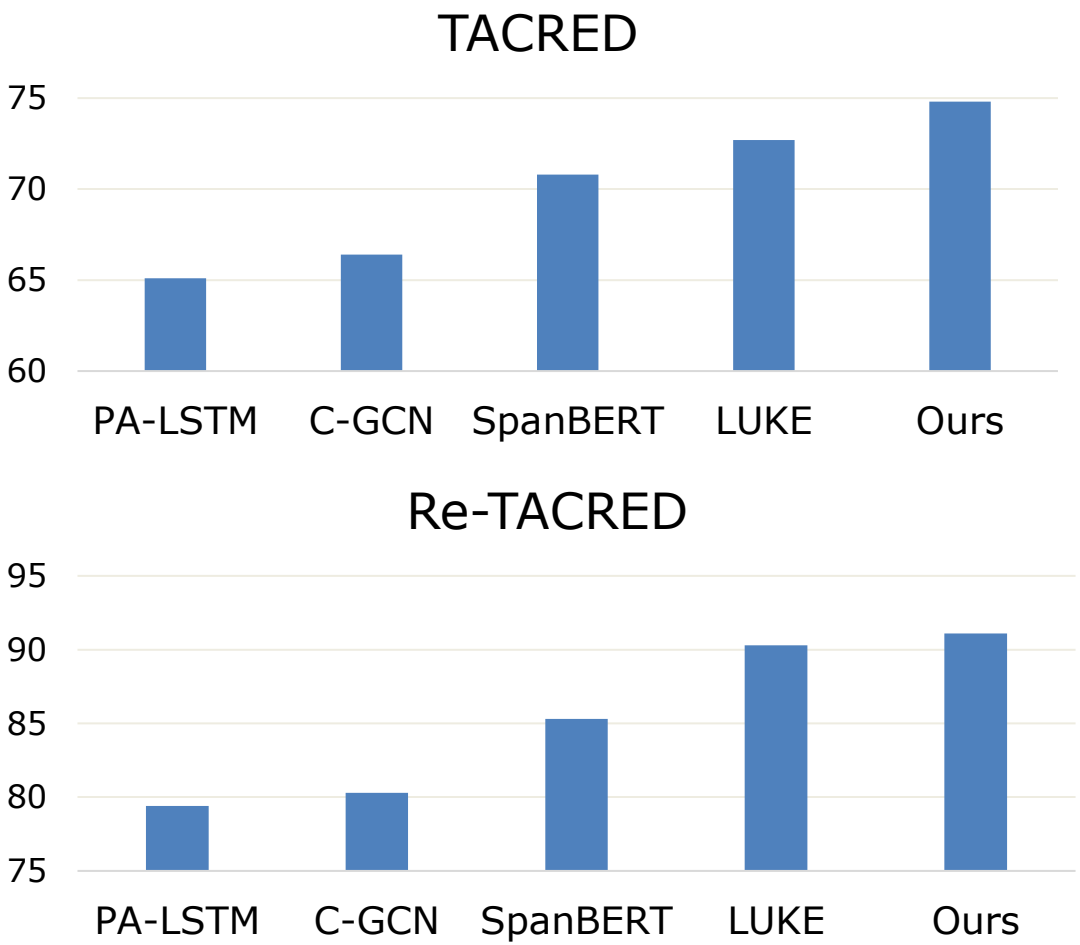
Model



@ * **person** * **Bill** @ was born in # ∧ **city** ∧ **Seattle** #.

Main Results

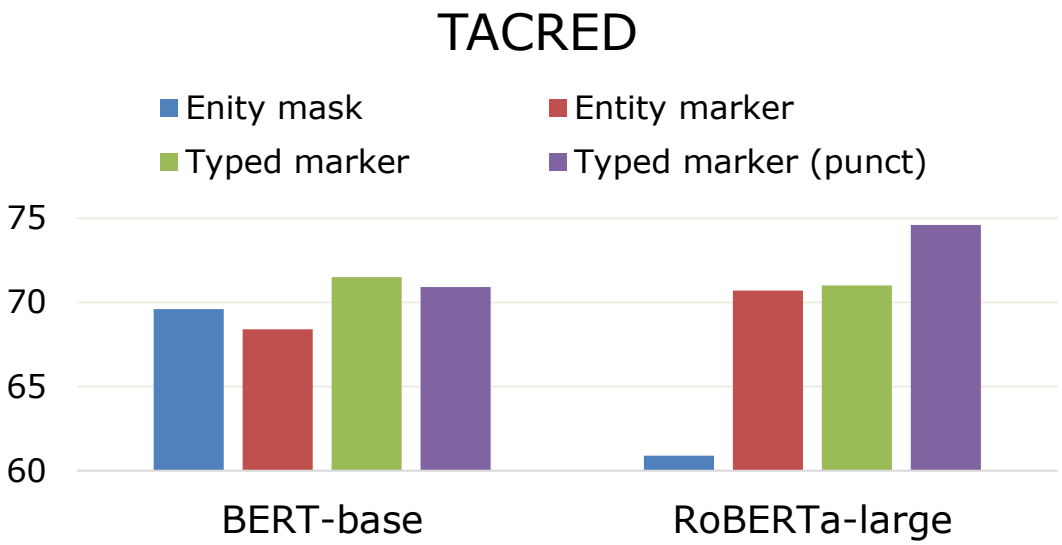
Datasets: TACRED, Re-TACRED



Observations:

1. Our model (RoBERTa + typed entity marker, punct) consistently outperforms existing methods.
2. Our model on Re-TACRED achieves an F1 of 91.1%, showing that PLMs already achieve high results on this task.

Results of Different Inputs



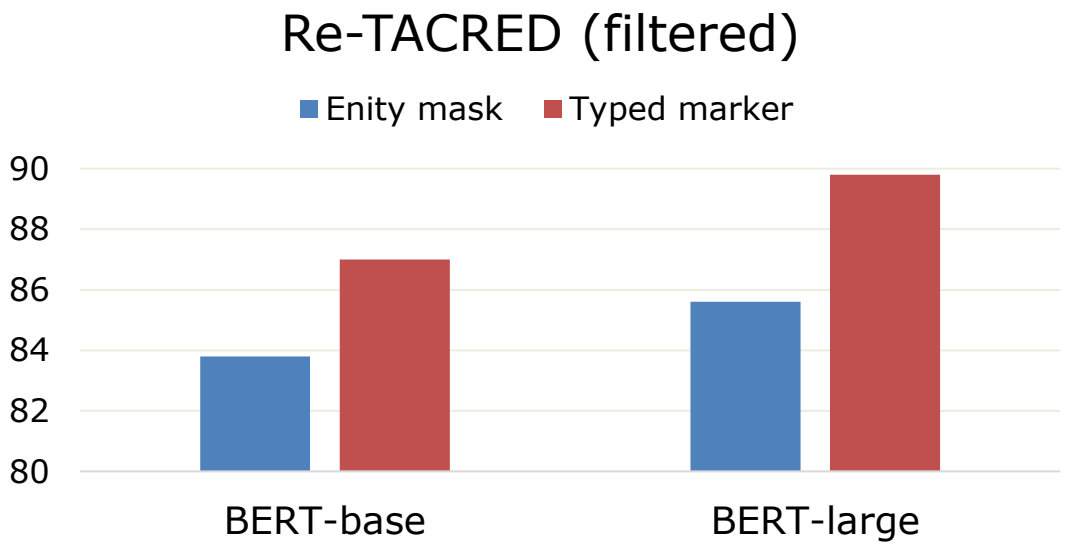
Observations:

1. Typed markers (original and punct) consistently outperform entity mask and entity marker.
2. The best representation technique depends on the backbone.

Results on Unseen Entities

Q: Does unrevealing entity names improves generalization?

Filtered evaluation setting: remove relation mentions from test set if their subject/object entities have appeared in training.



Observations:
Typed marker achieves better results even for unseen entities, may be due to that entity names contain more fine-grained information than entity types.

Conclusion

1. We present a simple yet strong **RE baseline** that offers new SOTA performance.
2. Specifically, we revisit two technical problems in sentence-level RE, namely **entity representation** and **noisy or ill-defined labels**.
3. Our improved RE baseline achieves an F1 score of 91.1% on the Re-TACRED dataset, showing that PLMs already achieve satisfactory performance on this task.

GitHub