CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning.

Daojian Zeng*§, Haoran Zhang*†, Qianying Liu‡

Changsha University of Science & Technology, Changsha, 410114, China

University of Illinois at Urbana-Champaign, Illinois, 61820, USA

Kyoto University, Kyoto, 606-8501, Japan

zengdj916@163.com; haoranz6@illinois.edu; ying@nlp.ist.i.kyoto-u.ac.jp

Abstract

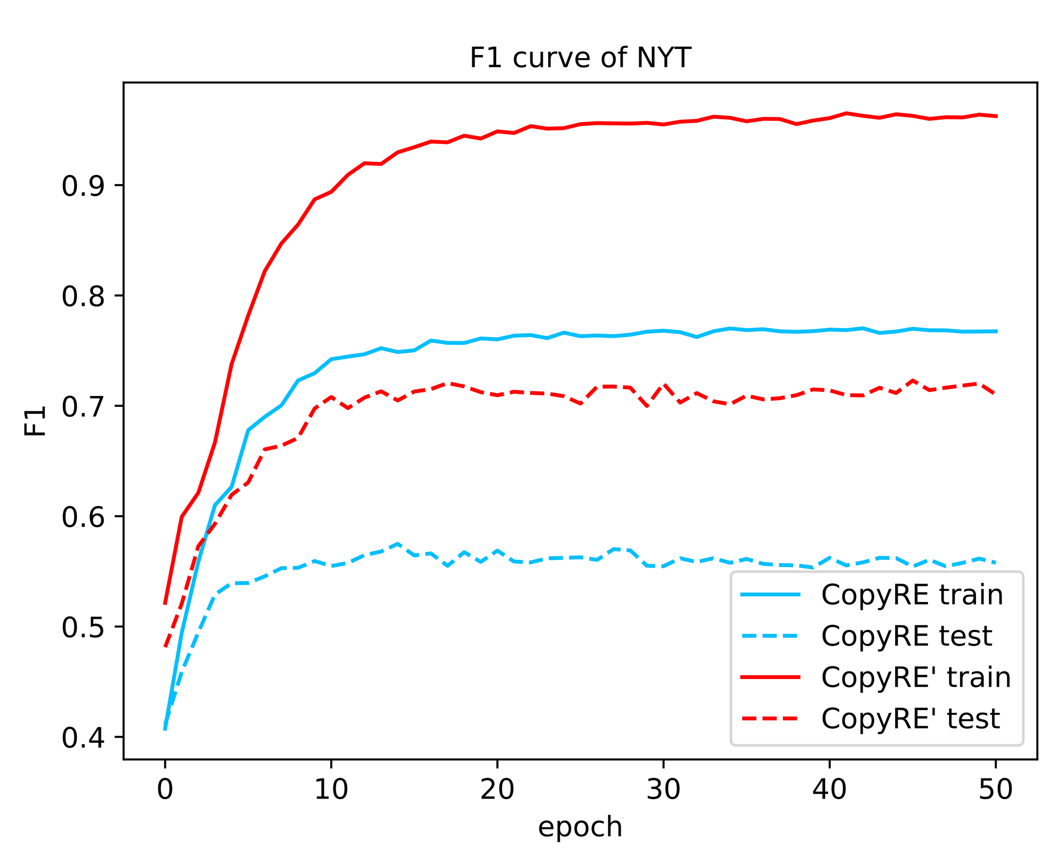
The existing Seq2Seq model for Joint Extraction of Entities and Relations, CopyRE [1], applies copy mode and generation mode in turn to decode the triplets from the plain text directly. Eg. SENTENCE $\rightarrow r_1, h_1, t_1, r_2, h_2, t_2$ ..., where the relation r_i is predicted by generation mode; the head entity h_i and tail entity t_i is the position of the last token of the entity, predicted by copy mode. However, CopyRE has two main problems:

1. The copy mode can only predict one token while an entity may have multiple tokens.
2. The copy mode for entity prediction predicts the same distribution all the time.

The 1st problem can be fixed by many ways, experiments show that the multi-task way (CopyMTL) is the best. The 2nd problem is caused by a linear algebra bug. We show detailed analysis from the perspective of theory and phenomena. We improve the performance over 10% F1 score than CopyRE.

Phenomena

1. After predicting head entity, CopyRE uses a mask to record the position and forbid it to predict the same position as tail entity. If we remove the mask, F1 = 0.
2. CopyRE cannot fit the training set well.

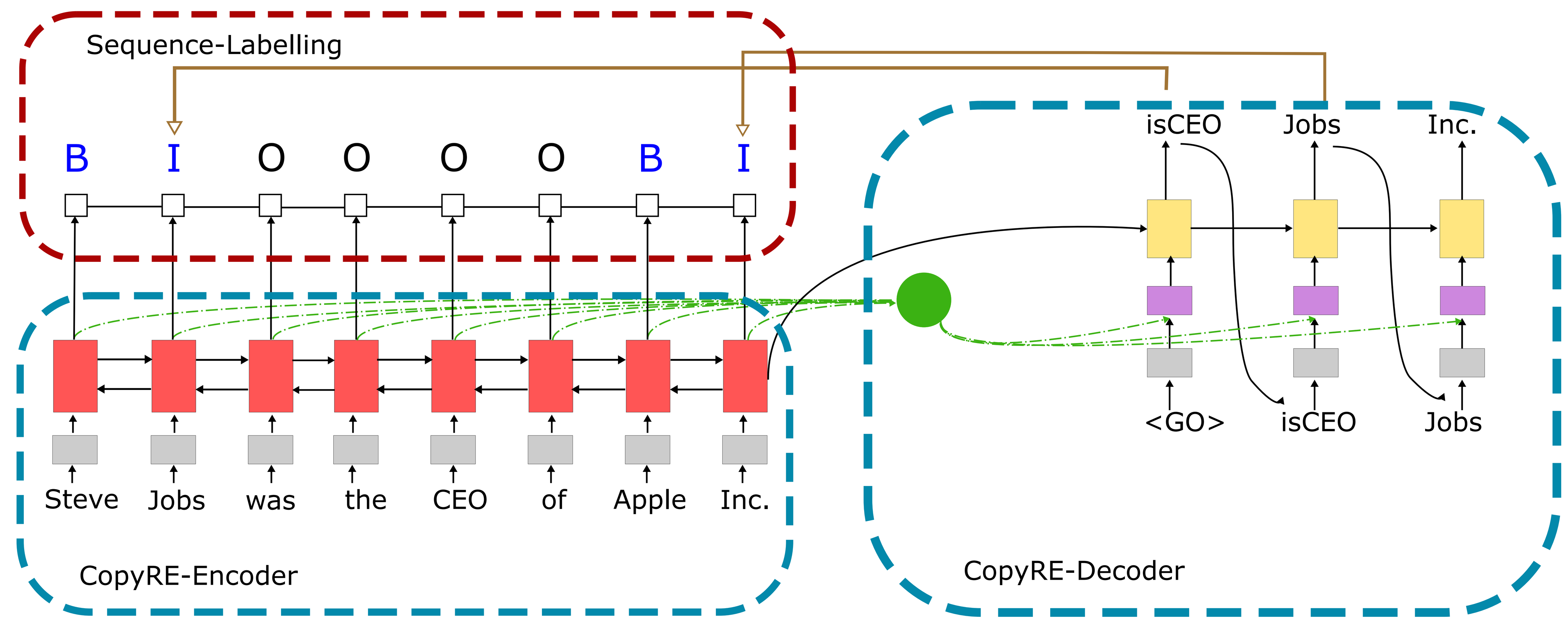


References

[1] Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. Extracting relational facts by an end-to-end neural model with copy mechanism. In *ACL Long*, pages 506–514, Melbourne, Australia, July 2018. Association for Computational Linguistics.

1. CopyMTL: How to predict entities with multiple token?

STEVE JOBS WAS THE CEO OF APPLE INC. $\rightarrow$ *isCEO, 1, 7*, where *1* and *7* are the position of JOBS and INC.
In addition to CopyRE, we add a CRF layer on the CopyRE encoder part. After decoding, the tagged boundary is used to find the beginning of the entities.



2. CopyRE': Why CopyRE predicts the same entity all the time?

CopyRE: In the entity copying steps, 2,3,5,6..., $\boldsymbol{h}_t^D$ fade away.

$$p(y_t|y_{<t}, s, t\%3=2,0) = \frac{e^{[\boldsymbol{h}_t^D;\boldsymbol{h}_i^E]\cdot \boldsymbol{W}^e}}{\sum_j e^{[\boldsymbol{h}_t^D;\boldsymbol{h}_j^E]\cdot \boldsymbol{W}^e}} = \frac{\cancel{e^{\boldsymbol{h}_t^D\cdot \boldsymbol{W}_1^e}}\cdot e^{\boldsymbol{h}_i^E\cdot \boldsymbol{W}_2^e}}{\cancel{e^{\boldsymbol{h}_t^D\cdot \boldsymbol{W}_1^e}}\cdot \sum_j e^{\boldsymbol{h}_j^E\cdot \boldsymbol{W}_2^e}} = \frac{e^{\boldsymbol{h}_i^E\cdot \boldsymbol{W}_2^e}}{\sum_j e^{\boldsymbol{h}_j^E\cdot \boldsymbol{W}_2^e}}$$

$\boldsymbol{W}^e \in \mathbb{R}^{d_{encoder}\times 1}$. The problematic CopyRE results in $p(t=2)=p(t=3)=p(t=5)=p(t=6)$...
CopyRE': Fix it with only one more non-linear layer!

$$p(y_t|y_{<t}, s, t\%3=2,0) = \frac{e^{\sigma([\boldsymbol{h}_t^D;\boldsymbol{h}_i^E]\cdot \boldsymbol{W}^f)\cdot \boldsymbol{W}^o}}{\sum_j e^{\sigma([\boldsymbol{h}_t^D;\boldsymbol{h}_j^E]\cdot \boldsymbol{W}^f)\cdot \boldsymbol{W}^o}}$$

$\boldsymbol{W}^f \in \mathbb{R}^{d_{encoder}\times 100}$, $\boldsymbol{W}^o \in \mathbb{R}^{100\times 1}$, σ is the activation function.

Result

Dataset	Model	Prec	Rec	F1
NYT	GraphRel-2p	.639	.600	.619
	CopyRE'5	.680	.663	.671
	CopyMTL	**.727**	**.692**	**.709**
WebNLG	GraphRel-2p	.447	.411	.429
	CopyRE'5	.572	.536	.553
	CopyMTL	**.578**	**.601**	**.589**

Dataset	Model	Prec	Rec	F1
NYT	CopyRE	.612	.530	.571
	CopyRE'	**.747**	**.700**	**.722**
WebNLG	CopyRE	.312	.272	.291
	CopyRE'	**.583**	**.629**	**.605**

CopyRE'5: SENTENCE $\rightarrow r_1, h_1, l_{h_1}, t_1, l_{t_1}, r_2, h_2, l_{h_2}, t_2, l_{t_2}$, ..., where l is the length of the entities.
In the left table, we compare the models with a stricter evaluation, considering the whole entity span. This required us to reprocess the dataset (with Stanford NER).
In the right table, we only consider the last token of the entity, which is the same with CopyRE.