Jointly Extracting Multiple Triplets with Multilayer Translation Constraints

Zhen Tan[1] Xiang Zhao[1,3,*] Wei Wang[2] and Weidong Xiao[1,3]

[1] National University of Defense Technology, China.

[2] The University of New SouthWales, Australia. [3] Collaborative Innovation Center of Geospatial Technology, China.

Introduction

Triplets extraction is an essential and pivotal step in automatic knowledge base construction, which captures structural information from unstructured text corpus. Conventional extraction models use a pipeline of named entity recognition and relation classification to extract entities and relations, respectively, which ignore the connection between the two tasks. Recently, several neural network-based models were proposed to tackle the problem, and achieved state-of-the-art performance. However, most of them are unable to extract multiple triplets from a single sentence, which are yet commonly seen in real-life scenarios. To close the gap, we propose in this paper a joint neural extraction model for multi-triplets, namely, **TME**, which is capable of adaptively discovering multiple triplets simultaneously in a sentence via ranking with translation mechanism. In experiment, TME exhibits superior performance and achieves an improvement of 37.6% on F1 score over state-of-the-art competitors.

Framework

A rationale that underlies many existing joint extraction models is that ***if a sentence contains more than two entities, it is possible that there is one relation existing between any pair of entities.*** Naturally, it suggests the following paradigm: given a sentence, we extract first entities (**Stage I**), and then relations between each candidate pairs of entities by classification (Stage **II**) However, the paradigm is flawed and performance deteriorates, when an entity has no relation, unwanted or multiple relations with others.

To resolve the issue, we present a revised framework (depicted in Figure 1 that generates candidate entity pairs without unpragmatic constraints in Stage I and refrains excessive irrelevant entities from going into Stage II.

It comprises: (1) a neural model using **BiLSTM+CRF** to obtain entity features, which are reused through feature sharing by a multi-layer module for capturing complex relation features via translation mechanism; (2) a tri-part tagging scheme for distinguishing whether an entity is involved with a wanted relation or not; and (3) a margin-based relation ranker, trained with negative samples, for discovering appropriate relations between entity pairs.

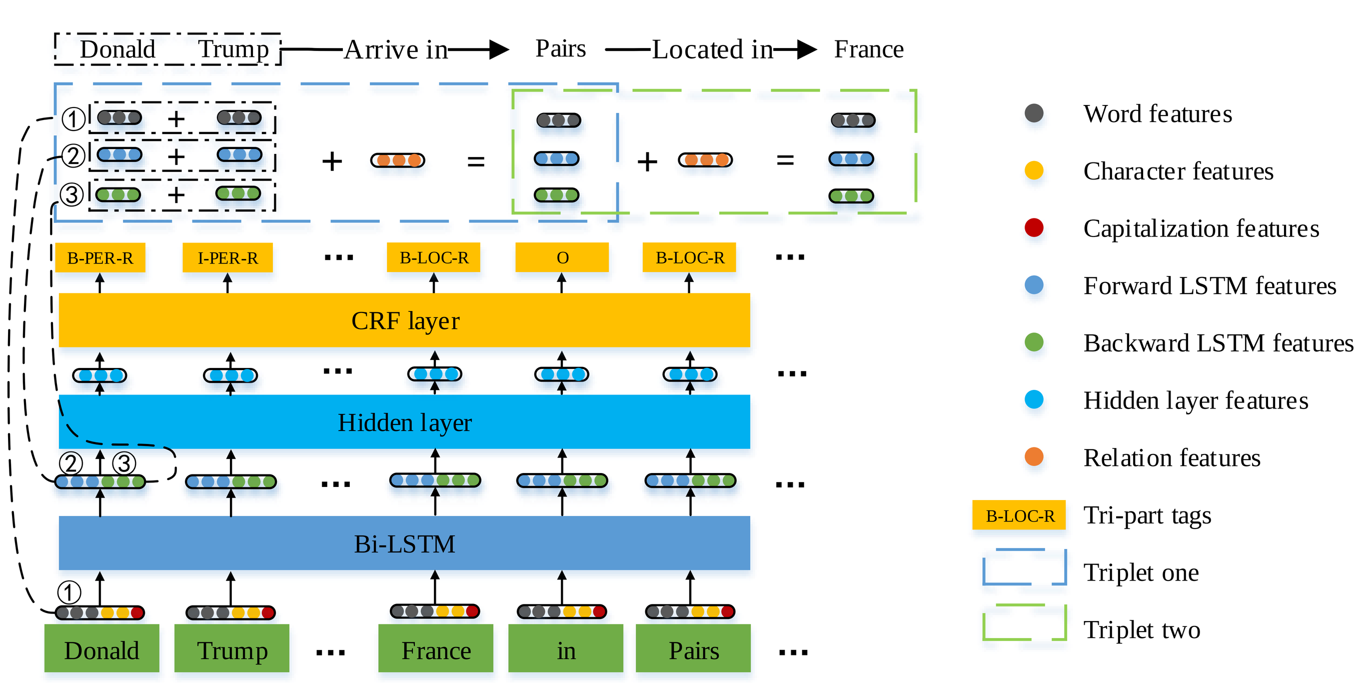

Figure 1: Framework of Joint Multi-triplets Extraction Model TME

Tri-part Tagging Scheme

We propose a tri-part tagging scheme **TTS** on the basis of **BiLSTM+CRF**, in order to give each word in a sentence a unique tag, which is used to extract entity features. It is constituted of three parts: (1) In position part (**PP**), we use ``BIO'' to encode the position information of the words regarding an entity; (2) In type part (**TP**), we associate words with type information of entities; and (3) In relation part (**RP**), we annotate whether an entity in the sentence is involved in any relation.

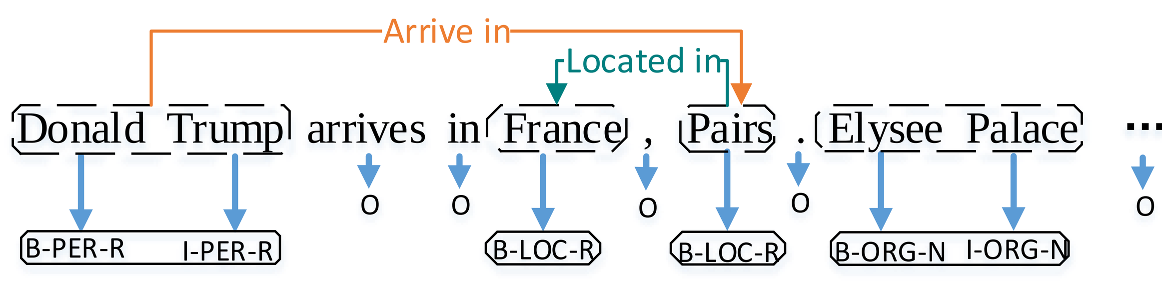

Figure 2: Sample Sentence with Tri-part Tagging

Multi-layer Translation Mechanism

For each triplet t = (e_1, e_2, r) ∈ T in the sequence, we obtain the head entity embedding $\mathbf{e_1}$ and tail entity embedding $\mathbf{e_2}$ in the embedding layer, and generate a corresponding relation embedding **r**. We require that $\mathbf{e_1}$ adding **r** is close to $\mathbf{e_2}$, i.e., $\mathbf{e_1} + \mathbf{r} \approx \mathbf{e_2}$mathematically. The score function is described as: $f(t) = -\|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_2\|_2^2$

Similarly, we obtain entity embeddings $\overrightarrow{\mathbf{e_1}}, \overrightarrow{\mathbf{e_2}}, \overleftarrow{\mathbf{e_1}}, \overleftarrow{\mathbf{e_2}}$ from the output of forward and reverse LSTM, respectively, and require that $\overrightarrow{\mathbf{e_1}} + \mathbf{r} \approx \overrightarrow{\mathbf{e_2}}$ and $\overleftarrow{\mathbf{e_1}} + \mathbf{r} \approx \overleftarrow{\mathbf{e_2}}$, Hence the score functions, respectively, are

$$\vec{f}(t) = -\|\overrightarrow{\mathbf{e_1}} + \mathbf{r} - \overrightarrow{\mathbf{e_2}}\|_2^2, \overleftarrow{f}(t) = -\|\overleftarrow{\mathbf{e_1}} + \mathbf{r} - \overleftarrow{\mathbf{e_2}}\|_2^2.$$

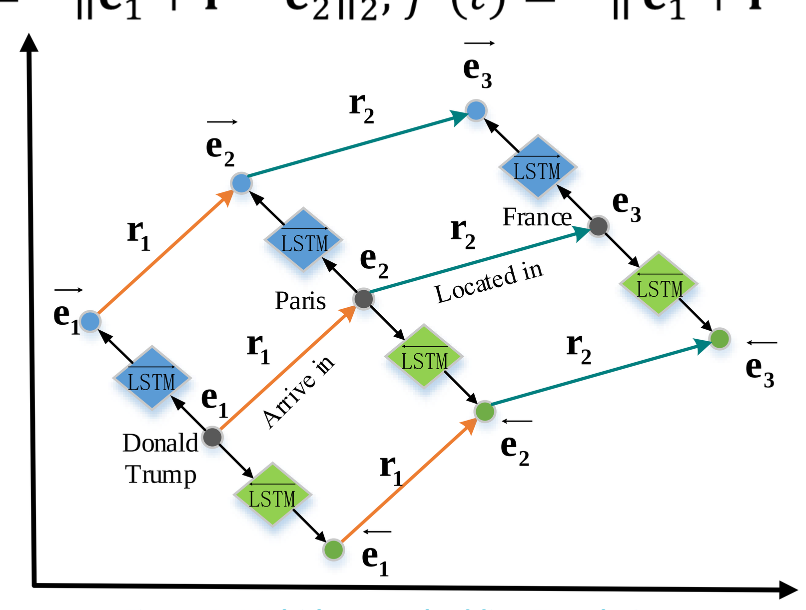

Figure 3: Multi-layer Embedding Translation

Experiments

Table 2: Experiment Results on NYT-single

Methods	Prec	Rec	F1
FCM	0.553	0.154	0.240
DS+logistic	0.258	0.393	0.311
LINE	0.335	0.329	0.332
MultiR	0.338	0.327	0.333
DS-Joint	0.574	0.256	0.354
CoType	0.423	0.511	0.463
NTS-Joint	**0.615**	0.414	0.495
TME (**top-1**)	0.583	0.485	**0.530**
TME (**top-2**)	0.515	0.508	0.511
TME (**top-3**)	0.458	**0.522**	0.489

Table 3: Experiment Results on NYT-multi

Methods	Prec	Rec	F1
CoType	0.385	0.340	0.361
NTS-Joint	0.533	0.336	0.412
TME-MR	0.638	0.421	0.507
TME-RR	0.423	0.452	0.437
TME-NS	0.558	0.496	0.525
TME (**top-1**)	**0.749**	0.436	0.551
TME (**top-2**)	0.696	0.478	**0.567**
TME (**top-3**)	0.631	**0.500**	0.558

Table 4: Ablation Study of TME on NYT-multi

Model	Top-1			Top-2			Top-3		
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
TME	**0.749**	**0.436**	**0.551**	**0.696**	0.478	**0.567**	**0.631**	0.500	**0.558**
−TTS (-**TP**)	0.741	0.436	0.549	0.680	0.478	0.561	0.610	0.498	0.548
−TTS (-**RP**)	0.610	0.376	0.465	0.488	**0.484**	0.486	0.400	**0.547**	0.462
-TTS (-**TP**-**RP**)	0.575	0.353	0.438	0.474	0.468	0.470	0.391	0.531	0.450
-Character	0.723	0.428	0.538	0.663	0.472	0.552	0.597	0.497	0.542
-CRF	0.690	0.414	0.517	0.608	0.470	0.530	0.522	0.495	0.509
-$\vec{f}$ - $\overleftarrow{f}$	0.552	0.310	0.398	0.521	0.368	0.431	0.468	0.399	0.431
-f	0.569	0.332	0.419	0.518	0.372	0.433	0.465	0.395	0.428
-Dropout	0.723	0.424	0.535	0.666	0.478	0.556	0.593	0.503	0.544
-Pretrain	0.686	0.411	0.514	0.613	0.466	0.530	0.539	0.495	0.516

Results Analysis

To prove the effectiveness of multi-triplets extraction, we use two dataset and compare with some baselines. The results are shown in Table 1,2. , and it reads that the F1 value in **TME**(**top-2**) is up to **0.567** and achieves a 36.7% improvement over **NTS**. Different from the results on **NYT-single**, the best results on **NYT-multi** are achieved by **top-2** rather than **top-1**, which can verify its abilities to process multi-triplets sentence.

To show the effectiveness of each component, we remove one particular component at a time to understand its impact on the performance. Compared with **TME**-*f*, multi-layer translation-based model gives the largest jump of 28.0% in F1 score, which verdicts the superiority of multi-layer model regarding triplet extraction. From the results of **TME**-**TTS**, we can conclude that **RP** and **TP** have positive effect on triplets extraction. Especially on **top-2**, the incorporation of **RP** brings a remarkable improvement (42.6%) in precision and negligible drop (-1.3%) in recall; this suggests that **RP** can effectively filter out entities irrelevant to target relations.

FOR FURTHER INFORMATION

Dr Zhen Tan

E: tanzhen08a@nudt.edu.cn

Acknowledgements

This work was partially supported by NSFC under grants Nos. 61872446, 71690233 and 71331008, ARC DPs 170103710 and 180103411, and D2DCRC DC25002 and DC25003.