A Hierarchical Framework for Relation Extraction with Reinforcement Learning

Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, Minlie Huang

Department of Computer Science & Technology, Tsinghua University, Beijing, China

Introduction

Task definition

- **Relation extraction**: identify relations (e_s, r, e_t), a triple consisting of a relation type r, a source entity e_s and a target entity e_t from unstructured texts.

Motivation

- Capture the **interaction** between entity mentions and relation types : most existing methods determine relation types only after all the entities have been recognized.
- Deal with **overlapping** relations : one entity may participate in multiple relations in the same sentence, and same entity pair in a sentence may be associated with different relations.

Figure 1: Examples for two types of overlapping relations.

- A novel paradigm with **hierarchical reinforcement learning** to deal with relation extraction by regarding the related entities as the arguments of a relation.

Framework

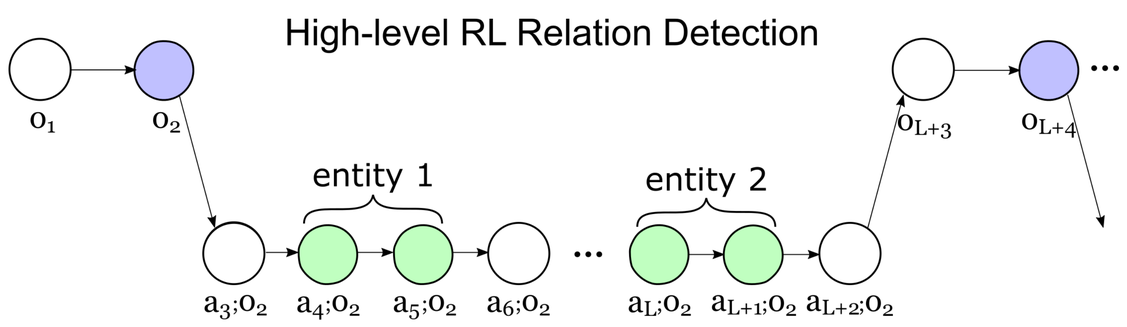

Figure 2: Overview of a hierarchical agent in relation extraction.

- **Relation Indicator**: the position in a sentence when sufficient information has been mentioned to identify a semantic relation.

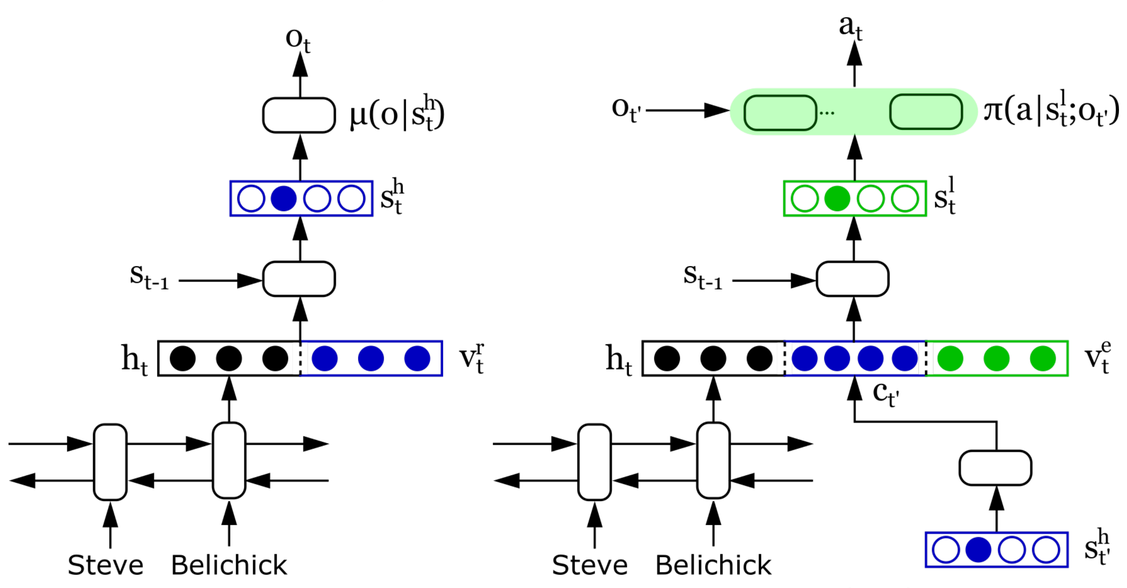

Figure 3: Illustration of a two-level hierarchical policy structure. (Left) High-level policy for relation detection. (Right) Low-level policy for entity extraction.

- Option (high-level action): detect the *relation indicator*.

$$\mathcal{O} = \{\mathrm{NR}\} \cup \mathcal{R}$$

- High-level reward: intermediate reward to check out the relation detection process, and final reward to evaluate the global performance in the sentence.

$$r_t^h = \begin{cases} -1, \text{ if } o_t \text{ not in } S \\ 0, \quad \text{if } o_t = \mathrm{NR} \\ 1, \quad \text{if } o_t \text{ in } S. \end{cases}, r_{fin}^h = F_\beta(S) = \frac{(1+\beta^2)\mathrm{Prec} \cdot \mathrm{Rec}}{\beta^2 \mathrm{Prec} + \mathrm{Rec}}$$

- Primitive action (low-level action): according to the proposed entity annotation scheme.

$$\mathcal{A} = (\{S, T, O\} \times \{B, I\}) \cup \{N\}$$

- Low-level reward: the prediction error over gold-standard annotation.

$$r_t^l = \lambda(y_t) \cdot \mathrm{sgn}(a_t = y_t(o_{t'})), \lambda(y) = \begin{cases} 1, \text{ if } y \neq N \\ \alpha, \text{ if } y = N. \end{cases}$$

Experiment

Model	NYT10			NYT11			NYT10-sub			NYT11-plus		
	Prec	Rec	F_1	Prec	Rec	F_1	Prec	Rec	F_1	Prec	Rec	F_1
FCM	–	–	–	.432	.294	.350	–	–	–	.234	.199	.219
MultiR	–	–	–	.328	.306	.317	–	–	–	.241	.214	.227
CoType	–	–	–	.486	.386	.430	–	–	–	.291	.254	.271
SPTree	.492	.557	.522	.522	**.541**	.531	.272	.315	.292	**.466**	.229	.307
Tagging	.593	.381	.464	.469	.489	.479	.256	.237	.246	.292	.220	.250
CopyR	.569	.452	.504	.347	.534	.421	.392	.263	.315	.329	.224	.264
HRL	**.714**	**.586**	**.644**	**.538**	.538	**.538**	**.815**	**.475**	**.600**	.441	**.321**	**.372**

Table 1: (Left) Main results on relation extraction. (Right) Performance comparison on extracting overlapping relations.

Model	NYT11			NYT11-plus		
	Prec	Rec	F_1	Prec	Rec	F_1
FCM	.502	.479	.490	.447	.327	.378
MultiR	.465	.439	.451	.423	.336	.375
CoType	.558	.558	.558	.491	.413	.449
SPTree	.650	.614	.631	**.700**	.343	.460
CopyR	.480	**.714**	.574	.626	.426	.507
HRL-Ent	**.676**	.676	**.676**	.577	.321	.413
HRL	.654	.654	.654	.626	**.456**	**.527**

Table 2: Performance comparison on relation detection(classification) only.

The lawsuit contended that the chairman of the [[News Corporation]Et-Company]Es-Founder , [[[Rupert Murdoch]Es-Company]Et-Founder]Es-Nationality , promised certain rights to shareholders , including the vote on the poison pill , in return for their approval of the company 's plan to reincorporate in the United States from [Australia]Et-Nationality .

Both [Steven A. Ballmer]Es-Company , [[[Microsoft]Et-Company]Et-Company]Es-Founder 's chief executive , and [[Bill Gates]Es-Company]Et-Founder , the chairman , have been involved in that debate inside the company , according to that person .

Table 3: Extraction examples by our model. A predicted relation indicator is marked in background color.

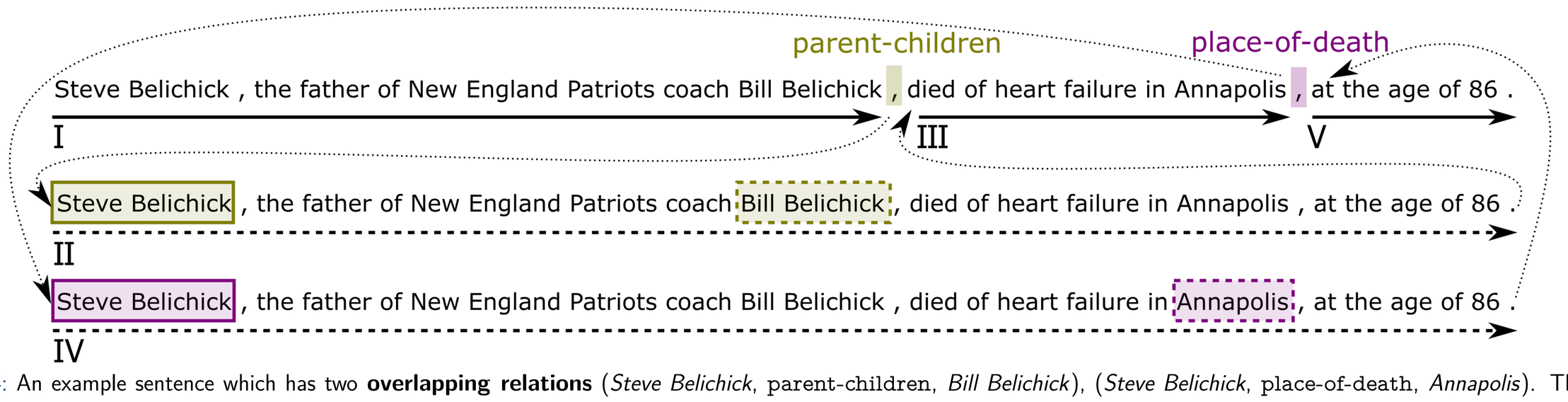

Figure 4: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes.

 Contact: {truthless11, keavilzhangzty}@gmail.com