A Hierarchical Framework for Relation Extraction with Reinforcement Learning

Ryuichi Takanobu*, Tianyang Zhang*, Jiexi Liu*, Minlie Huang

Tsinghua University

AAAI 2019

Relation Extraction

- **Relation Extraction**

Obama was **born in** the **United States**.

↓

Relation Triple: (**[Obama]**$_{es}$, ***BornIn,*** **[United States]**$_{et}$)

Source entity *Relation type* Target entity

- Joint extraction of **entity mentions** and **relation types**.

Motivation

- Deal with **overlapping relations.**
  - one entity participate in multiple relations in the same sentence

Parent--Children

Steve Belichick, the father of Bill Belichick, die in Annapolis.

Place of death

  - same entity pair in a sentence is associated with different relations

Company of

Bill Gates is the founder of Microsoft Corporation.

Founder of

Motivation

- Capture the **interaction** between **relation type** and **entities**.
  - Previous methods: pipeline or Cartesian product.
- Novel end-to-end **hierarchical** paradigm with HRL method, which identify entity mentions and relation types **jointly**.

Framework

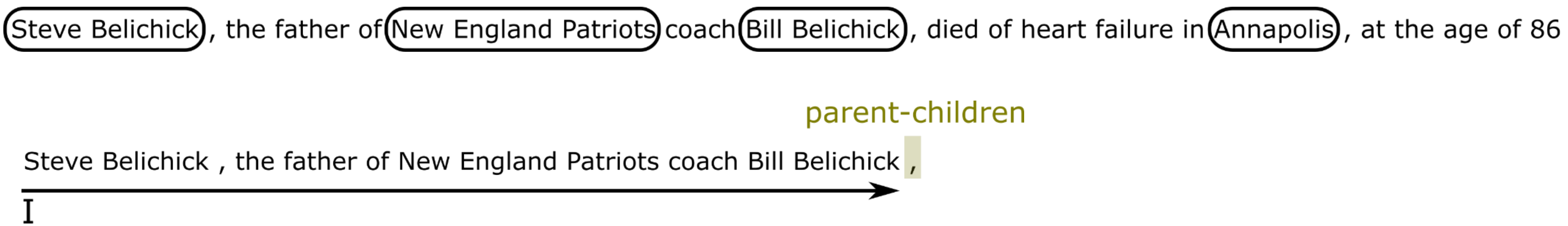


Figure 1: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes. This example shows how overlapping relations are extracted (*Steve Blichick* is included in both triples).

Framework

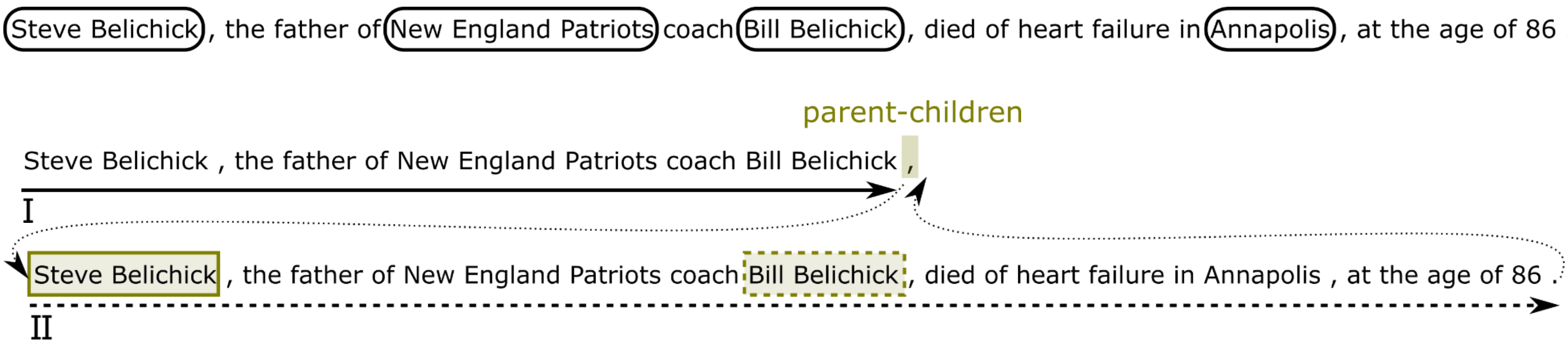


Figure 1: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes. This example shows how overlapping relations are extracted (*Steve Blichick* is included in both triples).

Framework

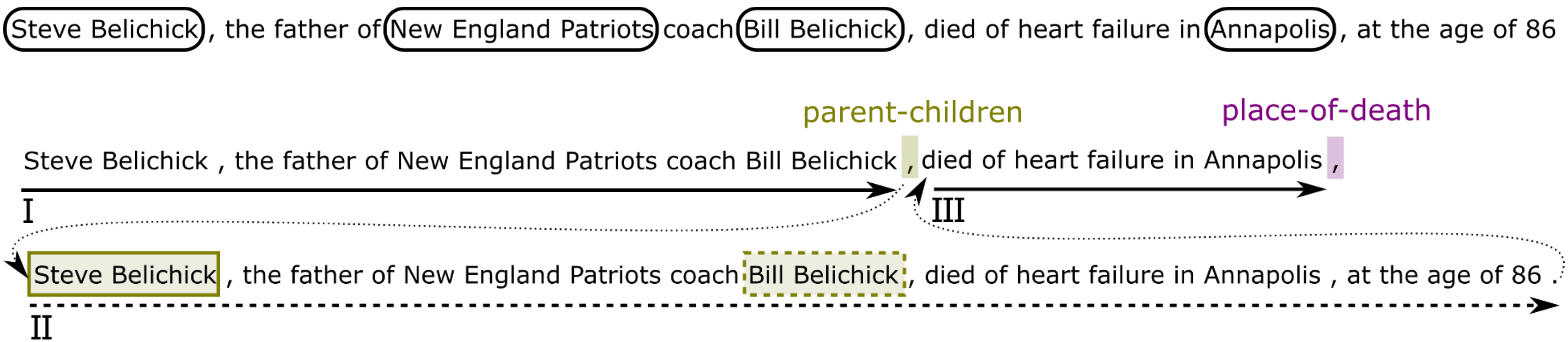


Figure 1: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes. This example shows how overlapping relations are extracted (*Steve Blichick* is included in both triples).

Framework

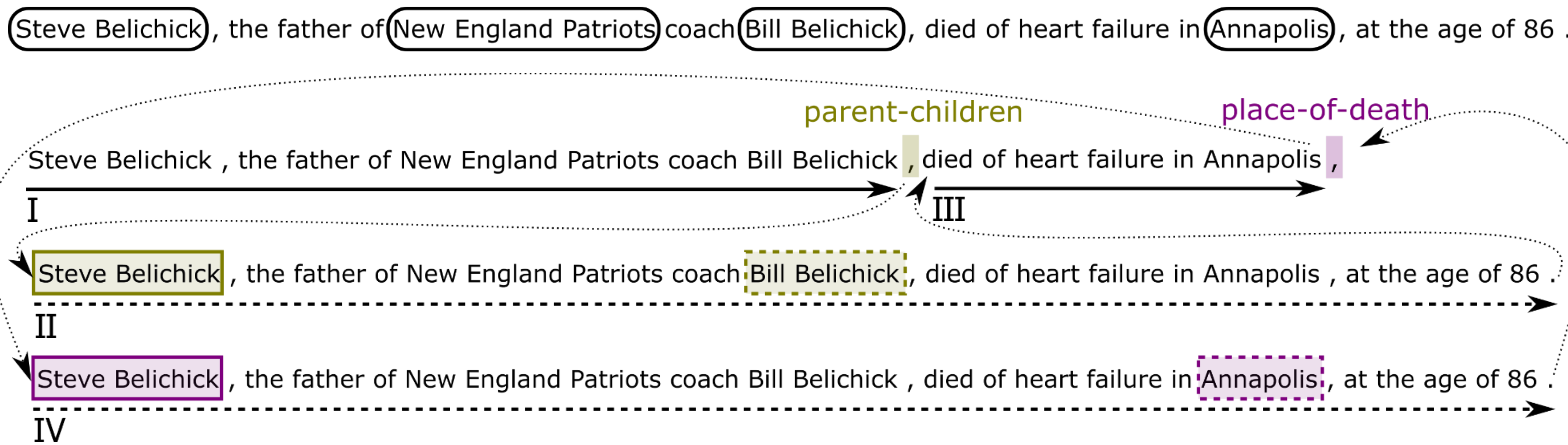


Figure 1: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes. This example shows how overlapping relations are extracted (*Steve Blichick* is included in both triples).

Framework

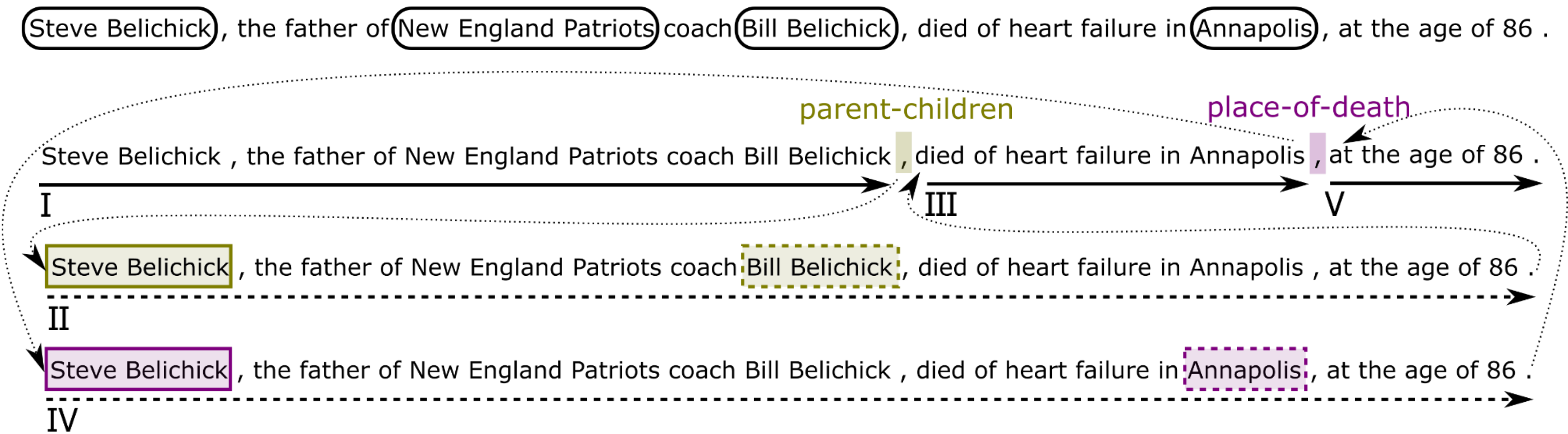
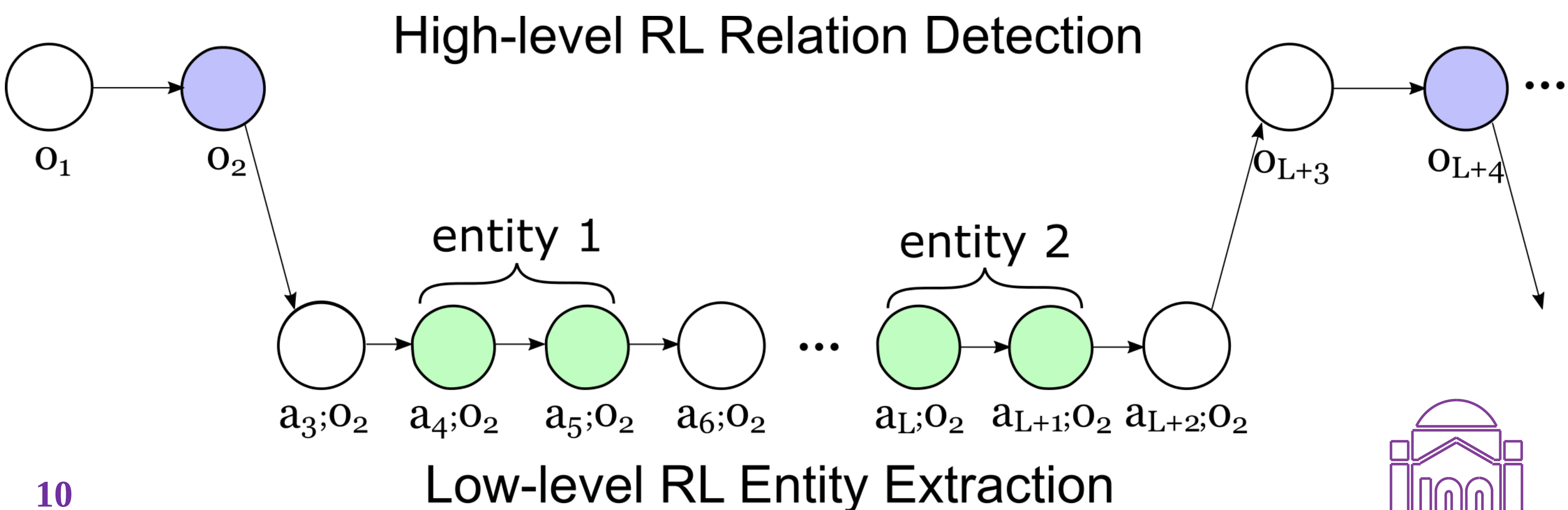


Figure 1: An example sentence which has two **overlapping relations** (*Steve Belichick*, `parent-children`, *Bill Belichick*), (*Steve Belichick*, `place-of-death`, *Annapolis*). The solid arrow indicates the high-level relation detection process, and the dashed arrow for low-level entity extraction. The dotted arrow marks a transition between the two processes. This example shows how overlapping relations are extracted (*Steve Blichick* is included in both triples).

Framework

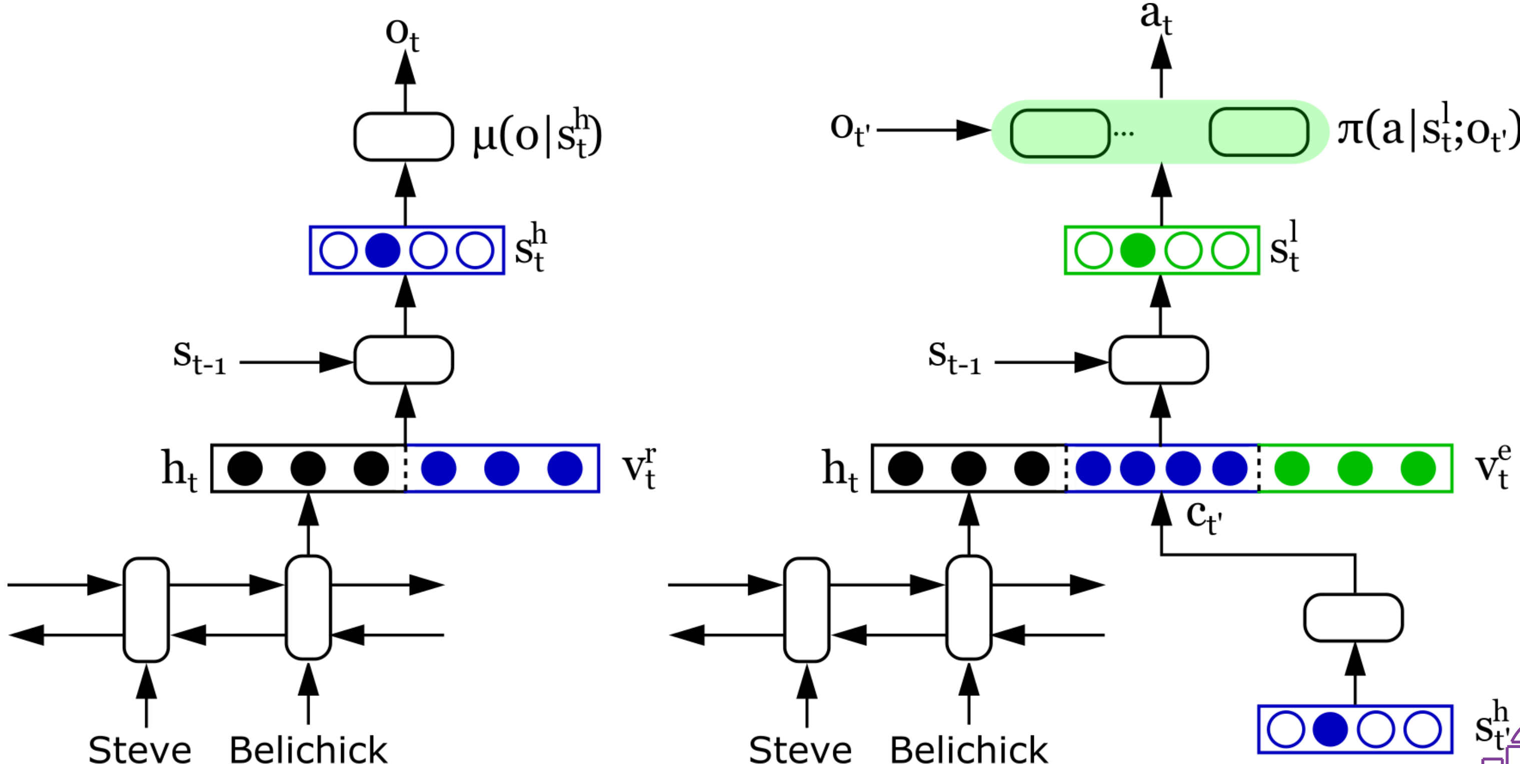
- **Relation Indicator**: the position in a sentence when sufficient information has been mentioned to identify a semantic relation.
- Treats entities as the **arguments** of a relation



Framework

Left: high-level policy for **relation detection** (as option)

Right: low-level policy for **entity extraction** (as primitive action)



Relation Detection

- **Option:** $\boldsymbol{O} = \{\boldsymbol{NR}\} \cup \boldsymbol{R}$, Relation type set.
- **State:** $\boldsymbol{s_t^h = f^h(W_s^h[h_t; v_t^r; s_{t-1}])}$
  - $\boldsymbol{v_t^r}$: The embedding of the ***latest*** option.
  - $\boldsymbol{h_t}$: LSTM output.
- **Policy:** $\boldsymbol{o_t \sim \mu(o_t | s_t^h) = softmax(W_\mu s_t^h)}$
- **Reward:**
  - $\boldsymbol{r_t^h \begin{cases} -1, if\ o_t\ not\ in\ S \\ 0, \quad if\ o_t = NR \\ 1, \quad if\ o_t\ in\ S \end{cases}}$ $\qquad \boldsymbol{r_{fin}^h = F_\beta(S)}$

Entity Extraction

- **Action:** $A = (\{S, T, O\} \times \{B, I\}) \cup \{N\}$
- **State:** $s_t^l = f^l\left(W_s^l[h_t; v_t^e; s_{t-1}; c_{t'}]\right)$
  - v_t^e: a learnable embedding of a_{t-1}
  - $c_{t'}$: relational state representation assigned to the latest option $s_{t'}^h$, $c_{t'} = g(W_h^l s_{t'}^h)$
- **Policy:** $a_t \sim \pi\left(a_t \middle| s_t^l; o_{t'}\right) = softmax(W_\pi[o_{t'}]s_t^l)$
- **Reward:**
  - $r_t^l = \begin{cases} sgn\left(a_t = y_t(o_{t'})\right), & if\ y \neq N \\ \alpha \times sgn\left(a_t = y_t(o_{t'})\right), & if\ y = N \end{cases}$

Entity Tag

- **S:** the source entity
- **T:** the target entity
- **O:** the entities that are not associated with the predicted relation
- **B:** beginning of an entity
- **I:** inside of an entity
- **N:** non-entity words

Steve	Belichick	,	the	father	of	New	England	Patriots	coach	Bill	Belichick	,	died	of	heart	failure	in	Annapolis	,	···
S_B	*S_I*	*N*	*N*	*N*	*N*	*O_B*	*O_I*	*O_I*	*N*	*T_B*	*T_I*	*N*	*N*	*N*	*N*	*N*	*N*	*O_B*	*N*	

Experiment

Model	NYT10			NYT11		
	Prec	Rec	F_1	Prec	Rec	F_1
FCM	–	–	–	.432	.294	.350
MultiR	–	–	–	.328	.306	.317
CoType	–	–	–	.486	.386	.430
SPTree	.492	.557	.522	.522	**.541**	.531
Tagging	.593	.381	.464	.469	.489	.479
CopyR	.569	.452	.504	.347	.534	.421
HRL	**.714**	**.586**	**.644**	**.538**	.538	**.538**

Table 2: Main results on relation extraction.

Model	NYT10-sub			NYT11-plus		
	Prec	Rec	F_1	Prec	Rec	F_1
FCM	–	–	–	.234	.199	.219
MultiR	–	–	–	.241	.214	.227
CoType	–	–	–	.291	.254	.271
SPTree	.272	.315	.292	**.466**	.229	.307
Tagging	.256	.237	.246	.292	.220	.250
CopyR	.392	.263	.315	.329	.224	.264
HRL	**.815**	**.475**	**.600**	.441	**.321**	**.372**

Table 3: 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 5: Performance comparison on relation detection.

Case Study

The lawsuit contended that the chairman of the [[News Corporation]$_{\text{Et-Company}}$]$_{\text{Es-Founder}}$, [[[Rupert Murdoch]$_{\text{Es-Company}}$]$_{\text{Et-Founder}}$]$_{\text{Es-Nationality}}$, 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]$_{\text{Et-Nationality}}$.

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

Table 4: Extraction examples by our model. The words in a bracket represents an entity extracted by the model. *Es* stands for source entity and *Et* for target entity. A predicted relation indicator is marked in background color (e.g. "Murdoch" in the first instance). The entities which form a triple are bracketed in the same color.

- Two triples (Red & Brown) share two entities (both head and tail entities) within a sentence

Case Study

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 4: Extraction examples by our model. The words in a bracket represents an entity extracted by the model. *Es* stands for source entity and *Et* for target entity. A predicted relation indicator is marked in background color (e.g. "Murdoch" in the first instance). The entities which form a triple are bracketed in the same color.

- Two triples (Red & Blue) share only one entity within a sentence

Summary

- A **hierarchical extraction** paradigm to approach relation extraction via **hierarchical reinforcement learning**.
  - Treats entities as the arguments of a relation.
  - Decomposes the relation extraction task into a hierarchy of two subtasks
- Good at modeling the **interactions** between the two subtasks.
- Particularly excels at extracting **overlapping relations**.
- **May be generalized to other pairwise tasks. (e.g. aspect-opinion mining)**