Recurrent Interaction Network for Jointly Extracting Entities and Classifying Relations

Kai Sun,[1,2] Richong Zhang,[1,2]*Samuel Mensah,[1,2] Yongyi Mao,[3] Xudong Liu[1,2]
[1]SKLSDE, School of Computer Science and Engineering, Beihang University, Beijing, China
[2]Beijing Advanced Institution on Big Data and Brain Computing, Beihang University, Beijing, China
[3]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada
sunkai@buaa.edu.cn, zhangrc@act.buaa.edu.cn, samensah@buaa.edu.cn, ymao@uottawa.ca, liuxd@act.buaa.edu.cn

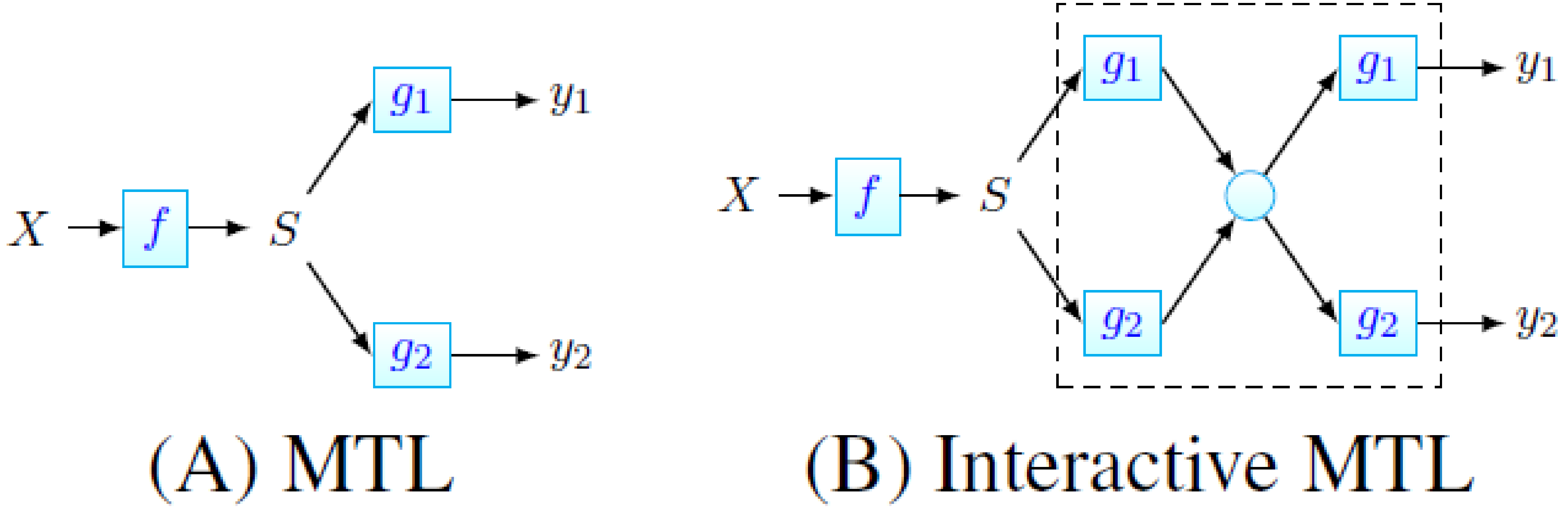


Figure 1: (A) A generalized multi-task model. (B) An interaction augmented multi-task model. X is the input, g_1 and g_2 are two models for different tasks. They correspond to NER and RE in this study.

Task and Motivation

- Sentence-level， joint NER and RE on NYT and WebNLG
- Explicit interaction between the NER model and the RE model will better guide the training of both models.

Task formulation

- pre-defined l relation types: $T = \{t_1, t_2, \dots t_l\}$
- given sentence of n words: $S = \{w_1, w_2, \dots w_n\}$
- Relation tripe: $\langle w, t, w' \rangle$, $t \in T, w, w' \in S, w \neq w'$
- 用第一个词代替多词的实体

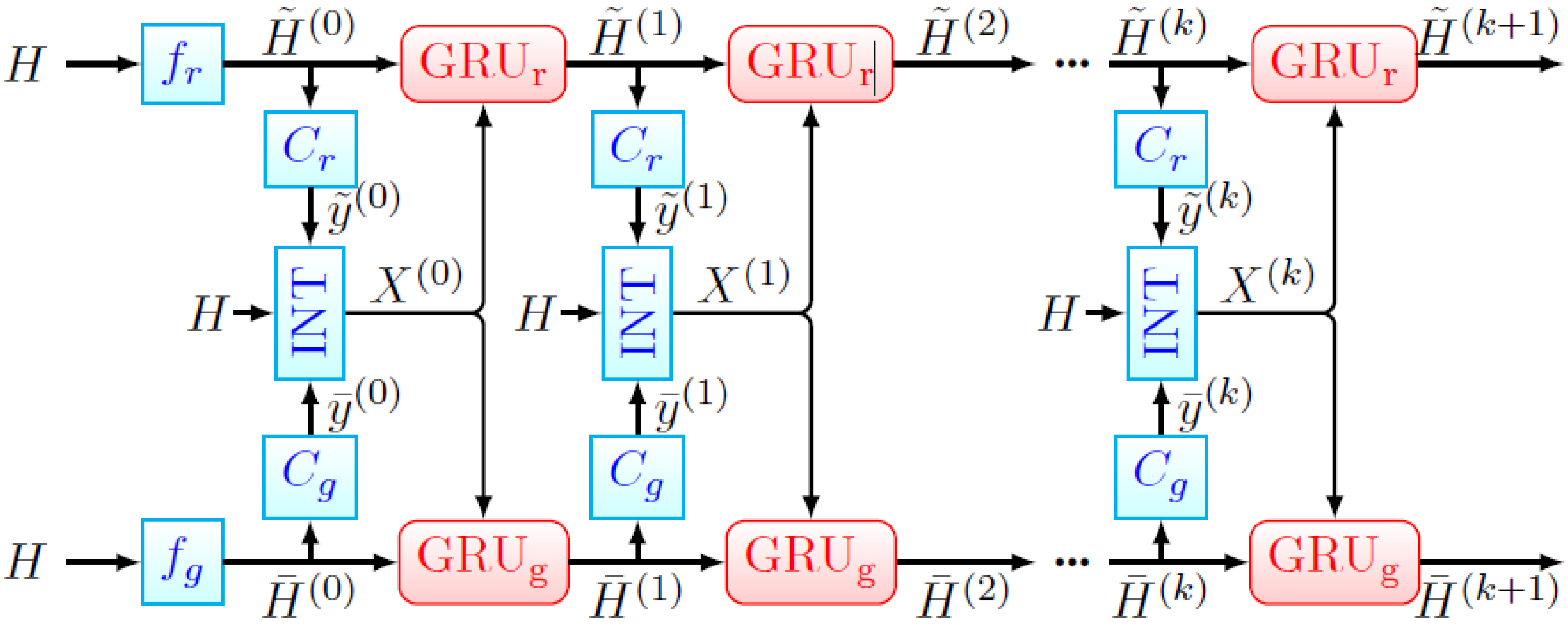


Figure 2: Overview of RIN. The f_r extracts relation-specific feature $\tilde{H}$ and f_g extracts entity-specific feature $\bar{H}$ from the sentence embedding H. The C_r is relation extraction model and the C_g is the entity recognition model. INT encodes the interaction information between two sub-tasks.

Input Model

- $h_i = Encoder(w_i)$
- $H = [h_1 \quad h_2 \quad \ldots \quad h_n]$
- $\tilde{h}_i = ReLU(W_r h_i + b_r), \quad \tilde{H} = [\tilde{h}_1 \;\; \tilde{h}_2 \;\; \ldots \;\; \tilde{h}_n]$
- $\bar{h}_i = ReLU(W_e h_i + b_e), \quad \bar{H} = [\bar{h}_1 \;\; \bar{h}_2 \;\; \ldots \bar{h}_n]$

NER & RE

$$\bar{y} = \mathrm{softmax}(W_{\mathrm{g}}\bar{h} + b_{\mathrm{g}})$$

$$m = \phi\left((W_{\mathrm{m}}\tilde{h}) \oplus (W_{\mathrm{r}}\tilde{h}')\right)$$

$$p(\langle w, t, w' \rangle) = \sigma\left(W_{\mathrm{p}} m + b_{i(t)}\right)$$

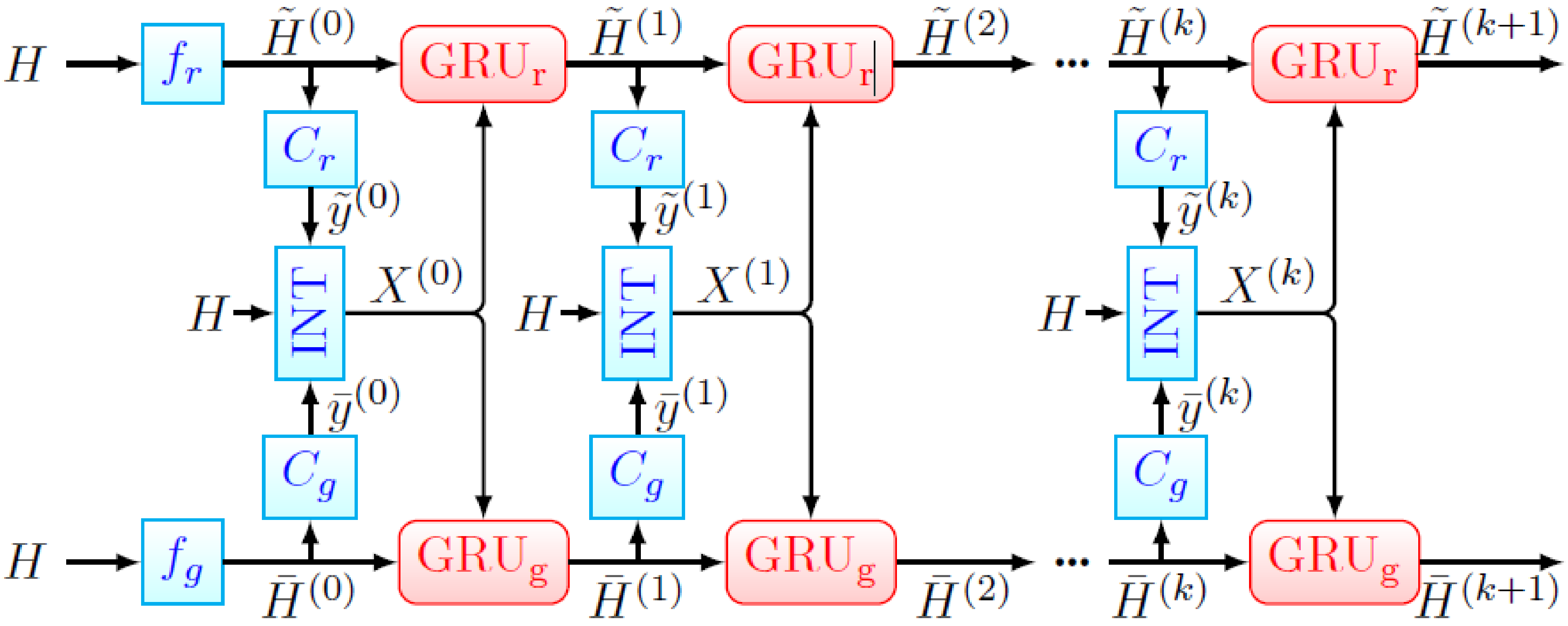


Figure 2: Overview of RIN. The f_r extracts relation-specific feature $\tilde{H}$ and f_g extracts entity-specific feature $\bar{H}$ from the sentence embedding H. The C_r is relation extraction model and the C_g is the entity recognition model. INT encodes the interaction information between two sub-tasks.

Interaction Module

$$\hat{p}(\langle w, t, \cdot\rangle) = \max_{w' \in s} p(\langle w, t, w'\rangle)$$

$$\hat{y} = \hat{p}(\langle w, t_1, \cdot\rangle) \oplus \hat{p}(\langle w, t_2, \cdot\rangle) \oplus \cdots \oplus \hat{p}(\langle w, t_l, \cdot\rangle)$$

$$x = \phi\left(W_{\mathrm{a}}(\hat{y} \oplus \bar{y} \oplus h) + b_{\mathrm{a}}\right)$$

embedding. By combing these three kinds of information in the INT module, we aim to learn a feature that conveys information about the alignment of NER and RE on word w. The interaction features on all words of the sentence s make up the interaction feature matrix $X = \{x_1, ..., x_n\}$.

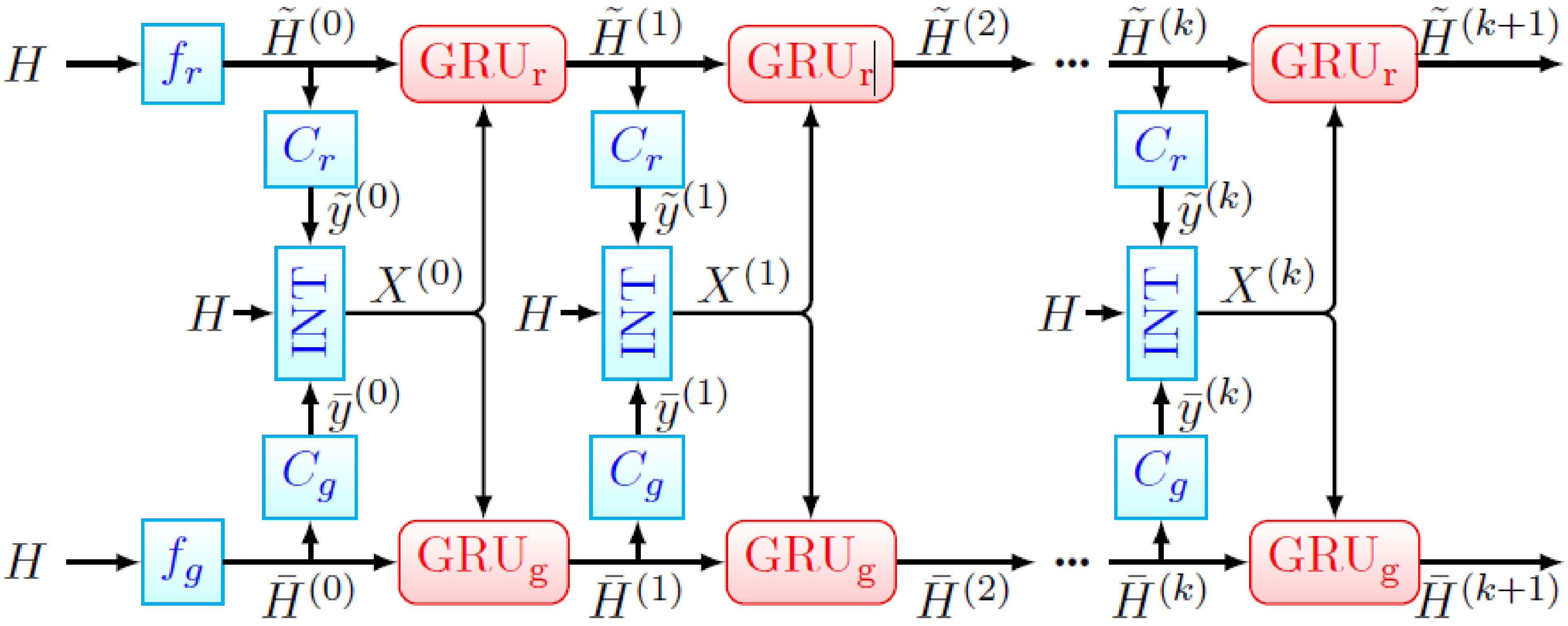


Figure 2: Overview of RIN. The f_r extracts relation-specific feature $\tilde{H}$ and f_g extracts entity-specific feature $\bar{H}$ from the sentence embedding H. The C_r is relation extraction model and the C_g is the entity recognition model. INT encodes the interaction information between two sub-tasks.

Feature update using GRU

$$z = \sigma\left(W_{\mathrm{z}}(\tilde{h} \oplus x)\right)$$

$$u = \sigma\left(W_{\mathrm{u}}(\tilde{h} \oplus x)\right)$$

$$\check{h} = \tanh\left(W_{\mathrm{o}}((u * \tilde{h}) \oplus x)\right)$$

$$\tilde{h}^{\mathrm{new}} = (1 - z) * \tilde{h} + z * \check{h}$$

Loss

$$L_{\mathrm{g}}(w) = \mathrm{CrossEntropy}\left(\bar{t}, \bar{y}\right)$$

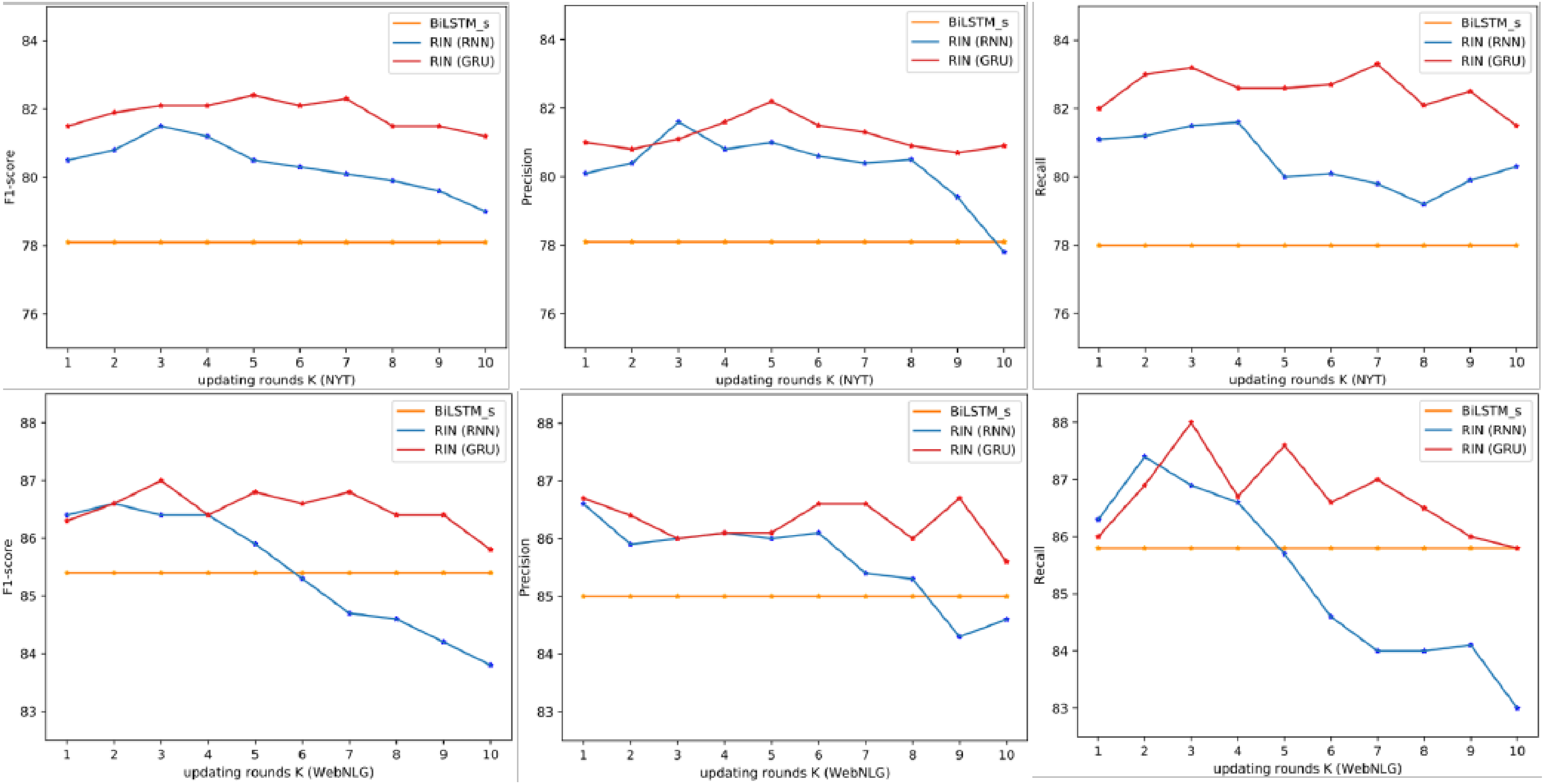
$$L_{\mathrm{r}}(\langle w, t, w' \rangle) = \mathrm{CrossEntropy}\left(\tilde{t}, \tilde{y}\right)$$

$$L = \sum_{s}\left(\sum_{w \in s} L_{\mathrm{g}}(w) + \sum_{w, w' \in s, t \in T} L_{\mathrm{r}}(\langle w, t, w' \rangle)\right)$$

Dataset	Train	Dev	Test
NYT	56195	5000	5000
WebNLG	5019	500	703

Table 2: Distribution of splits on NYT and WebNLG

Model	NYT			WebNLG		
	Precision	Recall	F1	Precision	Recall	F1
NovelTagging (Zheng et al., 2017)	62.4	31.7	42.0	52.5	19.3	28.3
MultiDecoder (Zeng et al., 2018)	61.0	56.6	58.7	37.7	36.4	37.1
GraphRel (Fu et al., 2019)	63.9	60.0	61.9	44.7	41.1	42.9
SeqtoSeq+RL (Zeng et al., 2019)	77.9	67.2	72.1	63.3	59.9	61.6
HBT (Wei et al., 2019)	**89.7**	85.4	87.5	89.5	88.0	88.8
BiLSTM (100d)	79.0	77.4	78.2	84.9	86.3	85.6
BiLSTM_s (100d)	78.1	78.0	78.1	85.0	85.8	85.4
RIN (100d, $K = 1$)	81.0	82.0	81.5	86.7	86.0	86.3
RIN (100d, $K = 3$)	81.1	83.2	82.1	86.0	88.0	87.0
RIN (100d, $K = 5$)	82.2	82.6	82.4	86.1	87.6	86.8
RIN (BERT, $K = 1$)	88.5	86.5	87.5	89.1	90.3	89.7
RIN (BERT, $K = 2$)	88.4	**87.1**	**87.8**	**90.0**	**90.3**	**90.1**


BiLSTM_s
RIN (RNN)
RIN (GRU)
F1-score
Precision
Recall
84
82
80
78
76
1 2 3 4 5 6 7 8 9 10
updating rounds K (NYT)
88
87
86
85
84
83
updating rounds K (WebNLG)

Model	RE	NER
BiLSTM (100d)	78.2	87.3
BiLSTM_s (100d)	78.1	87.6
RIN (100d, $k = 1$)	81.5	90.1
RIN (100d, $k = 5$)	**82.4**	**90.9**

Table 4: F1 performance of NER and RE on the NYT dataset.

Case1: A cult of victimology arose and was happily exploited by clever radicals among Europes Muslims, especially certain religious leaders like Imam Ahmad Abu Laban in Denmark and Mullah Krekar in Norway.	**Golden**:*Europe, Denmark, Norway* *(Europe, /location/location/contains, Denmark)* *(Europe, /location/location/contains, Norway)* **BiLSTM_s**: *Europe, Denmark, Norway* *(Europe, /location/location/contains, Denmark)* **RIN**: *Europe, Denmark, Norway* *(Europe, /location/location/contains, Denmark)* ***(Europe, /location/location/contains, Norway)***
Case2: Scott (No rating , 75 minutes) Engulfed by nightmares, blackouts and the anxieties of the age, a Texas woman flees homeland insecurity for a New York vision quest in this acute, resourceful and bracingly ambitious debut film.	**Golden**: *Scott, New York* *(York, /location/location/contains, Scott)* **BiLSTM_s**: ***Texas***, *New York* *(York, /location/location/contains, Scott)* **RIN**: *Scott, New York* *(York, /location/location/contains, Scott)*

Table 3: Case study for RIN and BiLSTM_s. The entities and relational triples are marked by blue and orange.

Thanks!