The goal in the task of Relation Extraction is to predict a KB relation that holds for a pair of entities given a set of sentences mentioning them (or NA if no such relation exists). The input is a KB $\Psi$ with relation set $R_{\Psi}$, a set of relations of interest $R_\subseteq R_{\Psi}$, and an automatically labelled training dataset $\mathcal{D}$ obtained via distant supervision. Given a sentence mentioning entities $h,t$, the output is a relation $r$ in $R$ that holds for $h,t$ or the catch-all relation $NA$ if no such $r$ exists. Formally, a labeled dataset for relation extraction consists of fact triples $\{(h_i, r_i, t_i)\}_{i=1}^{N}$ and a multi-set of extracted sentences for each triple $\{S_j\}_{j=1}^{N}$, such that each sentence $s \in S_i$ mentions both the head entity $h_i$ and the tail entity $t_i$.

**Problem Statement.** Given an entity pair $(h,t)$ and a set of sentences $\mathcal{S}$ mentioning them, the RE task is to estimate the probability of each relation in $R \cup \{NA\}$. Formally, for each relation $r$, we want to predict $P(r \mid h,t,\mathcal{S})$.

**HRERE (Heterogeneous REpresentations for neural Relation Extraction)**

HRERE is a neural relation extraction framework which learns language and knowledge jointly. HRERE’s backbone is a bi-directional LSTM network with multiple levels of attention to learn representations of text expressing relations. The knowledge representation machinery, borrowed from ComplEx (Trouillon et al., 2016), judges the language model to agree with facts in the knowledge base. Joint learning is guided by three loss functions: one for the language representation, another for the knowledge representation, and a third one to ensure these representations do not diverge.

$$J_L = -\frac{1}{N} \sum_{i=1}^{N} \log p(r_i \mid S_i; \Theta^{(L)})$$

$$J_K = -\frac{1}{N} \sum_{i=1}^{N} \log p(r_i \mid h_i, t_i; \Theta^{(K)})$$

$$J_D = -\frac{1}{N} \sum_{i=1}^{N} \log p(r^* \mid S_i; \Theta^{(B)})$$

$$\min \mathcal{J} = J_L + J_K + J_D + \lambda \Vert \Theta \Vert^2$$

where $r^*_i = \arg \max_{r \in \mathcal{R}_\Psi} p(r \mid h_i, t_i; \Theta^{(B)})$ and $\Theta = \Theta^{(L)} \cup \Theta^{(K)}$

**Experiment**

We study three variants of our framework:
- HRERE-base: basic neural model with local loss $J_L$ only;
- HRERE-naive: neural model with both local loss $J_L$ and global loss $J_D$ but without the dissimilarities $J_K$;
- HRERE-full: neural model with both local and global loss along with their dissimilarities.

<table>
<thead>
<tr>
<th>Method</th>
<th>P@10%</th>
<th>P@30%</th>
<th>P@50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weston</td>
<td>79.3</td>
<td>68.6</td>
<td>60.9</td>
</tr>
<tr>
<td>HRERE-base</td>
<td>81.8</td>
<td>70.1</td>
<td>60.7</td>
</tr>
<tr>
<td>HRERE-naive</td>
<td>83.6</td>
<td>74.4</td>
<td>65.7</td>
</tr>
<tr>
<td>HRERE-full</td>
<td>86.1</td>
<td>76.6</td>
<td>68.1</td>
</tr>
</tbody>
</table>

**References**


**code:** [github.com/billy-inn/HRERE](https://github.com/billy-inn/HRERE)