Connecting Language and Knowledge with Heterogeneous Representations for Neural Relation Extraction

UNIVERSITY OF ALBERTA
Peng Xu, Denilson Barbosa

Background and Problem

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 Ψ with relation set $\mathcal{R}_\Psi$, a set of relations of interest $\mathcal{R}$, $\mathcal{R} \subseteq \mathcal{R}_\Psi$, and an automatically labelled training dataset $\mathcal{D}$ obtained via distant supervision. Given a sentence mentioning entities h, t, the output is a relation $r \in \mathcal{R}$ 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 $\{\mathcal{S}_i\}_{i=1}^N$, such that each sentence $s \in \mathcal{S}_i$ 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 $\mathcal{R} \cup \{NA\}$. Formally, for each relation r, we want to predict $P(r \mid h, t, \mathcal{S})$.

Case Study

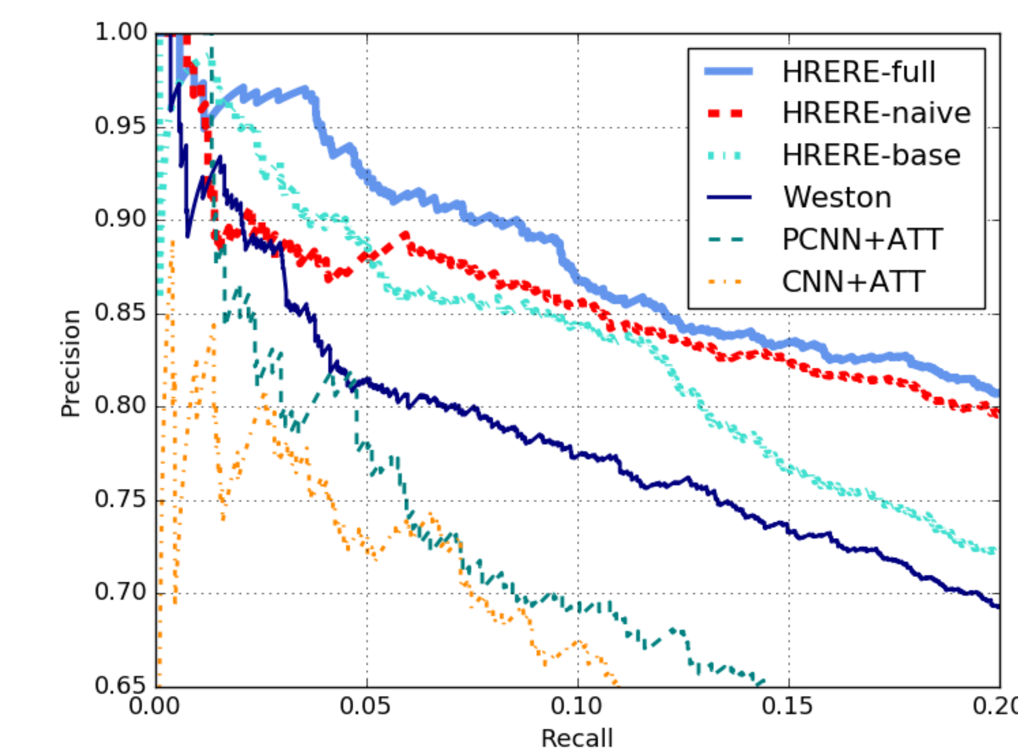
Relation	Textual Mention	base	naive	full
contains	Much of the **middle east** tension stems from the sense that shiite power is growing, led by **Iran**.	0.311	0.864	**0.884**
place_of_birth	Sometimes I rattle off the names of movie stars from **Omaha**: Fred Astaire, **Henry Fonda**, Nick Nolte ...	0.109	0.605	**0.646**
country	Spokesmen for **Germany** and Italy in Washington said yesterday that they would reserve comment until the report is formally released at a news conference in **Berlin** today.	0.237	0.200	**0.880**

Table 1: Some examples and the predicted probabilities of the true relations.

Experiment

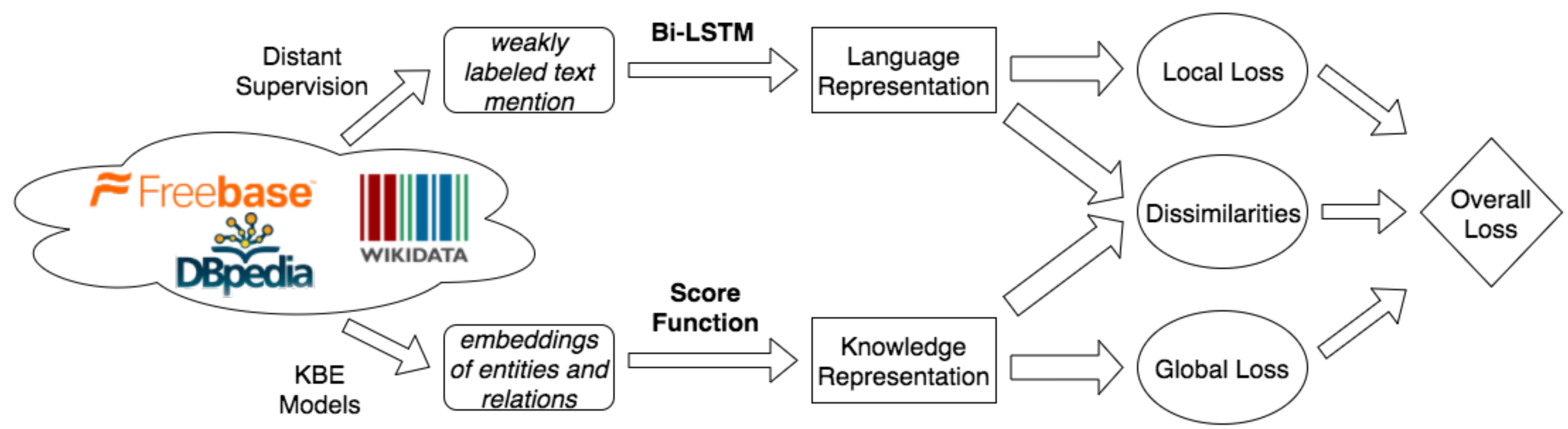
We study three variants of our framework:

- HRERE-**base**: basic neural model with local loss $\mathcal{J}_L$ only;
- HRERE-**naive**: neural model with both local loss $\mathcal{J}_L$ and global loss $\mathcal{J}_G$ but without the dissimilarities $\mathcal{J}_D$;
- HRERE-**full**: neural model with both local and global loss along with their dissimilarities.



P@N(%)	10%	30%	50%
Weston	79.3	68.6	60.9
HRERE-base	81.8	70.1	60.7
HRERE-naive	83.6	74.4	65.7
HRERE-full	**86.1**	**76.6**	**68.1**

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), nudges 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.

$$\mathcal{J}_L = -\frac{1}{N}\sum_{i=1}^{N} \log p(r_i|\mathcal{S}_i; \Theta^{(L)})$$

$$\mathcal{J}_G = -\frac{1}{N}\sum_{i=1}^{N} \log p(r_i|(h_i, t_i); \Theta^{(G)})$$

$$\mathcal{J}_D = -\frac{1}{N}\sum_{i=1}^{N} \log p(r_i^*|\mathcal{S}_i; \Theta^{(L)})$$

$$\min_{\Theta} \mathcal{J} = \mathcal{J}_L + \mathcal{J}_G + \mathcal{J}_D + \lambda\|\Theta\|_2^2$$

where $r_i^* = \arg\max_{r \in \mathcal{R} \cup \{NA\}} p(r|(h_i, t_i); \Theta^{(G)})$ and $\Theta = \Theta^{(L)} \cup \Theta^{(G)}$

Misc

References

1. Theo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. *Complex embeddings for simple link prediction*. In International Conference on Machine Learning, pages 2071-2080.

arXiv: arxiv.org/abs/1903.10126
code: github.com/billy-inn/HRERE