

A Implementation Details

Our models are implemented in *PyTorch* (Paszke et al., 2019). Hyper-parameters are found by grid search within search range listed in Table 4. The hyper-parameters of the best performing model are summarized in 5. All experiments are conducted on a 12-CPU machine running CentOS Linux 7 (Core) and NVIDIA RTX 2080 with CUDA 10.1. The average runtime for an epoch is 20 minutes.

During preprocessing, SciBERT tokenizer is used to tokenize each sentence in GE'11. Meanwhile, the mapping from each word's position to its head token position is recorded. We feed each sentence in GE'11 to MetaMap to identify all the biomedical concepts in each sentence. Each recognized concept is aligned with corresponding head token with the mapping from the previous step.

To pre-train KGE, we leverage the TransE implementation from OpenKE (Han et al., 2018). All tuples associated with the selected nodes described in Section 3.1 are used for pre-training with margin loss and negative sampling,

$$\mathcal{L} = \sum_{(h,\ell,t) \in S} \sum_{(h',\ell,t') \notin S} \max(0, d(h,\ell,t) - d(h',\ell,t') + \gamma)$$

where γ denotes margin, and $d(x, x')$ denotes the $\ell - 1$ distance between x and x' . h and t are embeddings of head and tail entities from the gold training sets S with relation ℓ . (h', ℓ, t') denotes a corrupted tuple with either the head or tail entity replaced by a random entity. TransE is optimized using Adam (Kingma and Ba, 2015) with hyper-parameters illustrated in Table 6. Every 50 epochs, the model checkpoint is saved if the mean reciprocal rank on the development set improve from the last checkpoint; otherwise, training will be stopped.

B Dataset

The statistics of GE'11 is shown in 7. The corpus contains 14496 events with 37.2% containing nested structure (Björne and Salakoski, 2011).⁶ We use the official dataset split for all the results reported.

⁶The dataset can be downloaded from <http://bionlp-st.dbcls.jp/GE/2011/downloads/>.

Hyper-parameter	Range
Relation MLP dim.	{300, 500, 700, 1000}
Trigger MLP dim.	{300, 500, 700, 1000}
Learning rate	$\{1 \times 10^{-5}, 3 \times 10^{-5}, 5 \times 10^{-5}\}$

Table 4: Hyper-parameter search range for fine-tuning SciBERT.

Hyper-parameter	Value
Relation MLP dim.	300
Trigger MLP dim.	300
Learning rate	3×10^{-5}
GEANet node dim.	300
GEANet edge dim.	300
GEANet layers	2
Dropout rate	0.2

Table 5: Hyper-parameters of the best performing GEANet-SciBERT model.

Hyper-parameter	Value
Learning rate	0.5
Margin	3
Batch size	128
# corrupted tuples / # gold tuples	25
# Epochs	500

Table 6: Hyper-parameters for pre-training KGE.

Metric	Number
events	14496
sentences	11581
nested events	37.2%
intersentence events	6.0%

Table 7: GE'11 dataset statistics