Supplementary Material: On the Equivalence of Holographic and Complex Embeddings for Link Prediction

Katsuhiko Hayashi
NTT Communication Science Laboratories, Seika-cho, Kyoto 619 0237, Japan
hayashi.katsuhiko@lab.ntt.co.jp

Masashi Shimbo
Nara Institute of Science and Technology, Ikoma, Nara 630 0192, Japan
shimbo@is.naist.jp

1 Experiments

1.1 Embedding Methods for Knowledge Graph Completion

In this material, we conduct experiments to empirically verify the relationship between holographic and complex embeddings (HolE and ComplEx). Let (r, s, o) be a triple, $\mathbf{w}_r, \mathbf{e}_s, \mathbf{e}_o \in \mathbb{R}^n$ be n-dimensional real vectors, $\mathbf{z}_r \in \mathbb{R}^{n^2}$ be n^2-dimensinal real vector and $\boldsymbol{\omega}_r, \boldsymbol{\varepsilon}_s, \boldsymbol{\varepsilon}_o \in \mathbb{C}^n$ be n-dimensional complex vectors. Here, we compare the following embedding methods for knowledge graph completion:

$$
\begin{aligned}
f_{\text{RESCAL}}(r, s, o) &= \mathbf{z}_r \cdot (\text{vec}(\mathbf{e}_s \otimes \mathbf{e}_o)) \\
f_{\text{HolE}}(r, s, o) &= \mathbf{w}_r \cdot (\mathbf{e}_s \star \mathbf{e}_o) \\
f_{\text{ComplEx}}(r, s, o) &= \boldsymbol{\omega}_r \cdot (\overline{\boldsymbol{\varepsilon}_s} \odot \boldsymbol{\varepsilon}_o)
\end{aligned}
$$

where $\otimes$ denotes tensor product and $\text{vec}(\mathbf{Z})$ projects a tensor $\mathbf{Z}$ into a vector.

1.2 Datasets and Evaluation Protocol

We used a benchmark dataset FB15k[1], which is a large subgraph of Freebase where all entities are present in Wikilinks database. The FB15k dataset consists of 483,142 training, 50,000 development, and 59,071 testing triples, containing 14,951 entities and 1,345 relation types.

We adopt the same evaluation protocol used in previous studies. For each test triple (r, s, o), we corrupt it by replacing the object o (or the subject s) with every entity e in the knowledge graph and calculate a score of this corrupted triple (r, s, e) (or (r, e, o)). Then, we rank all these triples by their scores in decreasing order. To measure the quality of the ranking, we use the mean reciprocal rank (MRR). In this paper, we report its filtered version: for some testing sample (r, s, o), if we calculate the metric for the subject s (or the object o), the filtered metric removes all the other positive triples (r, e, o) $(e \neq s)$ (or (r, s, e) $(e \neq o)$) from the ranking that appear in either training, validation or test set.

[1] https://everest.hds.utc.fr/doku.php?id=en:transe

1.3 Implementation

We trained all models using the stochastic gradient descent algorithm by minimizing the negative log-likelihood of the logistic model with L2 regularization:

$$
\min_{\theta} \sum_{(r,s,o,y) \in \mathcal{D}} \log\{1 + \exp(-y f(r, s, o))\} + \lambda ||\boldsymbol{\Theta}||_F^2
$$

where $\boldsymbol{\Theta}$ corresponds to paramter matrix for entities and relations. Reported results in the next section are given for the best set of hyper-parameters evaluated on the development set for each model, after grid search on the following values: $n \in \{20, 40, 50, 100, 200, 400\}$, $\lambda \in \{0.0, 0.05, 0.1, 0.5\}$, $\eta \in \{0.01, 0.02, 0.05, 0.1\}$ where η is the initial learning rate. For all models, we generate 5 negative samples per one positive triple by heuristics such as the local closed world assumption (Dong et al., 2014).

For RESCAL and HolE, we randomly generate initial real vectors from the uniform distribution $\text{U}[-\frac{\sqrt{6}}{\sqrt{2n}}, \frac{\sqrt{6}}{\sqrt{2n}}]$ (Glorot and Bengio, 2010). For ComplEx, we directly generate initial complex vectors from the normal Gaussian distribution, following Trouillon's implementation of ComplEx, which is available at https://github.com/ttrouill/complex. Note that as shown in Fig. 1, a distribution of real and imaginary values in a complex vector generated by DFT of a real vector from $\text{U}[-\frac{\sqrt{6}}{\sqrt{2n}}, \frac{\sqrt{6}}{\sqrt{2n}}]$ also follows a Gaussian distribution which is in a similar value range to a complex vector from the normal Gaussian. It could be predicted from this fact and our theory that HolE and ComplEx will achieve similar performance.

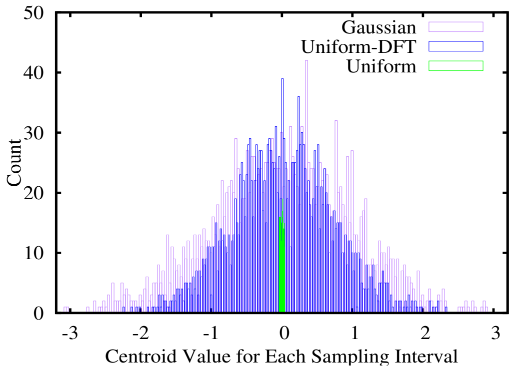

Figure 1: Histograms for two random vectors from the normal Gaussian distribution and the uniform $U[-\frac{\sqrt{6}}{\sqrt{2n}}, \frac{\sqrt{6}}{\sqrt{2n}}]$ distribution, and for one vector generated by DFT of the latter vector.

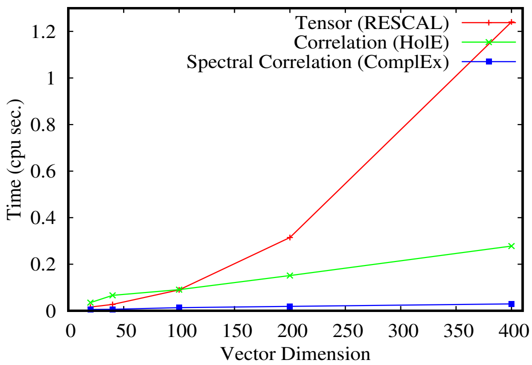

Figure 2: Speed comparison: time vs. vector dimension.

In the following experiments, all algorithms have been implemented in Java and were evaluated on a single 2.8GHz Intel Core i7 and 16GB RAM.

1.4 Results

The first interesting advantage of ComplEx over HolE is the faster training time. Actually, in our experiments, the training time per iteration was about 10 times faster. We also show in Fig. 2 that for test data, ComplEx can be computed in linear-time and is much faster than RESCAL and HolE.

Fig. 3 plots the filtered MRR accuracy curves on test data. The RESCAL model is pretty worse than the other models. As mentioned in (Nickel et al., 2016), the tensor product makes models prone to overfitting because it requires high (n^2-) dimensional parameters. ComplEx achieves almost the same accuracy as HolE when the vector size of the former is just half of that of the

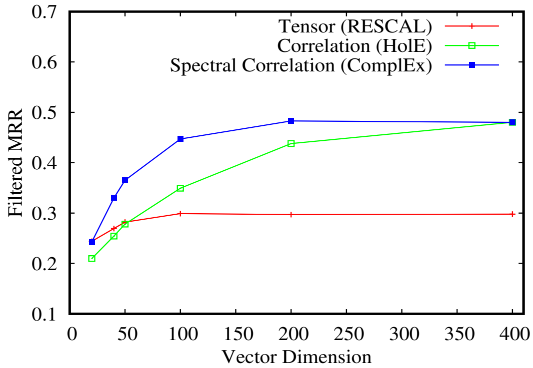

Figure 3: Performance comparison: filtered MRR accuracy vs. vector dimension.

latter (though ComplEx needs the same amount of memory as HolE since the former uses complex vectors). These results empirically support our theory.

References

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. pages 601–610.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*. pages 249–256.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI '16)*. pages 1955–1961.