Deep Learning Assignment 2 Report

Kng Yew Chian

Introduction

In this experiment, we trained a TransE model and a STransE model on the PharmKG dataset. After obtaining the entity and relation embeddings, we visualized them on the tensorboard embeddings projector. Each model was trained for 50 epochs. Embedding dimensions of 50 and 100 were tested for each model. Two different negative samplers, basic and bernoulli, were also tested for each model.

Background

The TransE model learns the embeddings of the entities and relationships by modeling relationships as translations in the embedding space. Hence, the embeddings of both entities and relationships are represented as vectors of size d.

On the other hand, the STransE model supplements the approach of the TransE model with two new matrices representing the relationship with respect to the head and tail entity respectively. The approach of using these matrices was used in the SE model, one of the earlier models proposed by the same writer of the TransE and STransE model. Theoretically, by involving these two relationship matrices of dimensions (d,d) as well, the STransE model can also figure out how different relationships affect the embeddings of similar relationships.

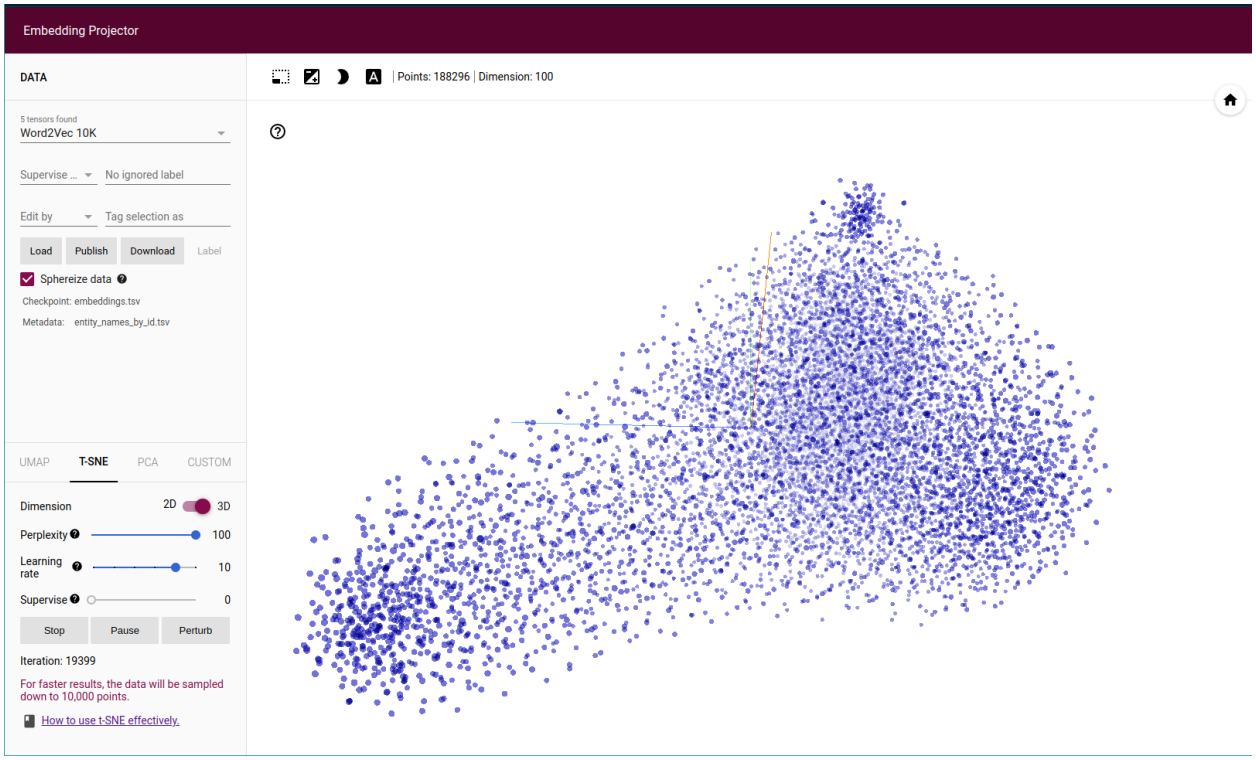
Results

Evaluated with inverse harmonic mean rank

- Rank: realistic, Side: both

	TransE			
NegativeSampler	Basic	Basic	Bernoulli	Bernoulli
Embedding Dimensions (d)	50	100	50	100
Inverse harmonic mean rank	0.026424	0.028052	0.028176	**0.028180**

	STransE			
NegativeSampler	Basic	Basic	Bernoulli	Bernoulli
Embedding Dimensions (d)	50	100	50	100
Inverse harmonic mean rank	0.011382	0.011600	0.012399	0.011530



Discussion

The evaluation metric, inverse harmonic mean rank, is the same as mean reciprocal rank. The best loss of 0.028180 can be interpreted as: the correct head/tail prediction lies around the 1/0.028180th = 35.5th ranked answer suggested by the model.

Comparing the minimum loss of TransE and the minimum loss of STransE, TransE would rank the correct answer around the 35.5th place, while STransE would rank it around the 1/0.012399 = 80.7th place.

TransE performed better in general compared to STransE. One explanation could be that, since STransE has two additional [d,d] sized relationship matrices to train, STransE was unable to generalize and optimize its weights as quickly as TransE did.

For each model, the changes to the embedding dimensions did not seem to affect the loss much. The use of a bernoulli negative sampler instead of a basic negative sampler also did not seem to affect the loss much.
In the embeddings projection, the clusters are not obvious, but entities of similar types such as diseases seem to be clustered together in the t-SNE visualization.

Conclusion

The results show a big difference in the performances of both models. Although STransE supplements the TransE model with the two extra relationship matrices to learn higher order relationships better, in this experiment, the STransE seems to have embeddings that are not as good as those in TransE in showing the relationships between the head and tail entities.

The models may have to be trained on more epochs for more conclusions on the performance of the two models.