

Assignment 4 Report

GMLFA (AI60007) - Autumn,2024 - IIT Kharagpur

Group 14

Vinayak Mali
24AI60R13

Saurabh Jaiswal
24AI60R46

October 26, 2024

1 Model Training and Evaluation Report

This section summarizes the training and evaluation results for two models: **TransE** and **TransR**. Both models were trained over 50 epochs, and their performance was evaluated based on two metrics: **Mean Rank (MR)** and **Hits@10**.

1.1 1. TransE Model

Training Loss: The training loss for the TransE model decreased steadily over the epochs, indicating an improvement in the model’s ability to learn embeddings:

- **Epoch 1:** 45.4096
- **Epoch 50:** 16.8001

The loss showed significant improvements, especially in the initial epochs, where the model reduced the loss from 45.4 to around 30 by epoch 15. The training loss plateaued in the later epochs, suggesting that the model may have reached a local minimum.

Evaluation Results:

- **Mean Rank (MR):** 8.0373
- **Hits@10:** 72.64%

The evaluation metrics indicate that the TransE model performed well, achieving a mean rank of approximately 8 and a Hits@10 score of 72.64%. This suggests that, on average, the correct entity was ranked within the top 10 for about 72.64% of the test instances.

1.2 2. TransR Model

Training Loss: Similar to the TransE model, the TransR model also displayed a consistent decrease in training loss over the epochs:

- **Epoch 1:** 20.0288
- **Epoch 50:** 9.0756

The TransR model had a more gradual decrease in loss compared to TransE, particularly in the latter half of the training. The model started at a lower initial loss and showed minor fluctuations but maintained a downward trend throughout the training process.

Evaluation Results:

- **Mean Rank (MR):** 8.0622
- **Hits@10:** 70.40%

The TransR model’s evaluation metrics were very close to those of the TransE model, with a mean rank of approximately 8.06 and Hits@10 of 70.40%. The performance indicates that the TransR model was slightly less effective in ranking the correct entities within the top 10 compared to TransE, but the results are still competitive.

1.3 Comparative Analysis

- **Loss Reduction:** TransE exhibited a more significant reduction in loss throughout the training compared to TransR, especially in the early stages.
- **Mean Rank and Hits@10:** Both models achieved similar mean ranks, but TransE outperformed TransR in Hits@10 by a margin of approximately 2.24%.
- **Training Stability:** TransR showed more stability in loss reduction in later epochs, indicating that the model may benefit from additional training or hyperparameter tuning.

1.4 Evaluation Metrics

The models were evaluated using Mean Rank (MR) and Hits@10 metrics, as shown in Table 1.

Model	Mean Rank (MR)	Hits@10 (%)
TransE	8.04	72.64%
TransR	8.06	70.40%

Table 1: Link Prediction Performance

1.5 Conclusion

Both TransE and TransR models have shown promising results in terms of embedding quality for the given task. While TransE achieved slightly better performance metrics, TransR’s training stability suggests potential for further improvements. Future work could explore additional epochs, hyperparameter tuning, or the incorporation of different learning rate strategies to enhance the models’ effectiveness.

2 Similar Triple Retrieval Task

The models were assessed on their ability to find similar triples for selected validation triples.

2.1 Validation Triples

1. (brazil, commonbloc1, india)
2. (burma, intergovorgs3, indonesia)
3. (china, accusation, uk)
4. (cuba, reldiplomacy, china)
5. (egypt, embassy, uk)

2.2 Retrieval Results

2.2.1 TransE Retrieval Results

- (brazil, commonbloc1, india):
 - (ussr, commonbloc1, india), Similarity score: 0.8340
 - (brazil, commonbloc1, burma), Similarity score: 0.7740
- (burma, intergovorgs3, indonesia):
 - (india, intergovorgs3, indonesia), Similarity score: 0.7536

2.2.2 TransR Retrieval Results

- (brazil, commonbloc1, india):
 - (poland, commonbloc1, india), Similarity score: 0.9609
- (burma, intergovorgs3, indonesia):
 - (india, intergovorgs3, indonesia), Similarity score: 0.9319

3 All Top tuples for both TransE and TransR

We evaluated the ability of TransE and TransR to find similar triples for five selected validation triples based on cosine similarity.

3.1 Validation Triples

The validation triples are:

1. (brazil, commonbloc1, india)
2. (burma, intergovorgs3, indonesia)
3. (china, accusation, uk)
4. (cuba, reldiplomacy, china)
5. (egypt, embassy, uk)

3.2 Similar Triple Retrieval Results

The top 5 similar triples for each validation triple, along with their similarity scores, are presented below for both TransE and TransR models.

3.2.1 TransE Similar Triple Results

- **Validation Triple: (brazil, commonbloc1, india)**
 - Similar triple: (ussr, commonbloc1, india), Similarity score: 0.8340
 - Similar triple: (brazil, relintergovorgs, india), Similarity score: 0.7764
 - Similar triple: (brazil, commonbloc1, burma), Similarity score: 0.7740
 - Similar triple: (brazil, ngoorgs3, india), Similarity score: 0.7702
 - Similar triple: (brazil, commonbloc1, egypt), Similarity score: 0.7494
- **Validation Triple: (burma, intergovorgs3, indonesia)**
 - Similar triple: (india, intergovorgs3, indonesia), Similarity score: 0.7536
 - Similar triple: (burma, intergovorgs3, usa), Similarity score: 0.7249
 - Similar triple: (burma, exports3, indonesia), Similarity score: 0.7104
 - Similar triple: (burma, intergovorgs3, uk), Similarity score: 0.6988
 - Similar triple: (burma, relngo, indonesia), Similarity score: 0.6881
- **Validation Triple: (china, accusation, uk)**
 - Similar triple: (china, blockpositionindex, uk), Similarity score: 0.7470
 - Similar triple: (china, accusation, indonesia), Similarity score: 0.7130
 - Similar triple: (china, accusation, ussr), Similarity score: 0.7038
 - Similar triple: (china, accusation, usa), Similarity score: 0.6921
 - Similar triple: (china, relngo, uk), Similarity score: 0.6734
- **Validation Triple: (cuba, reldiplomacy, china)**
 - Similar triple: (cuba, timesinceally, china), Similarity score: 0.8051
 - Similar triple: (cuba, reldiplomacy, poland), Similarity score: 0.7666
 - Similar triple: (cuba, reldiplomacy, uk), Similarity score: 0.7535
 - Similar triple: (cuba, reldiplomacy, indonesia), Similarity score: 0.7265
 - Similar triple: (poland, reldiplomacy, china), Similarity score: 0.6966
- **Validation Triple: (egypt, embassy, uk)**
 - Similar triple: (egypt, dependent, uk), Similarity score: 0.7750
 - Similar triple: (egypt, booktranslations, uk), Similarity score: 0.7588
 - Similar triple: (egypt, unweightedunvote, uk), Similarity score: 0.7583
 - Similar triple: (egypt, timesincewar, uk), Similarity score: 0.7548
 - Similar triple: (egypt, intergovorgs, uk), Similarity score: 0.7484

3.2.2 TransR Similar Triple Results

- **Validation Triple: (brazil, commonbloc1, india)**
 - Similar triple: (poland, commonbloc1, india), Similarity score: 0.9609
 - Similar triple: (brazil, commonbloc1, burma), Similarity score: 0.9560
 - Similar triple: (brazil, commonbloc1, jordan), Similarity score: 0.9515
 - Similar triple: (usa, commonbloc1, india), Similarity score: 0.9479
 - Similar triple: (ussr, commonbloc1, india), Similarity score: 0.9450
- **Validation Triple: (burma, intergovorgs3, indonesia)**
 - Similar triple: (india, intergovorgs3, indonesia), Similarity score: 0.9319
 - Similar triple: (burma, intergovorgs3, india), Similarity score: 0.9312
 - Similar triple: (burma, intergovorgs3, brazil), Similarity score: 0.9286
 - Similar triple: (burma, intergovorgs3, uk), Similarity score: 0.9222
 - Similar triple: (burma, intergovorgs3, usa), Similarity score: 0.9130
- **Validation Triple: (china, accusation, uk)**
 - Similar triple: (china, accusation, usa), Similarity score: 0.9678
 - Similar triple: (china, accusation, india), Similarity score: 0.9620
 - Similar triple: (china, accusation, indonesia), Similarity score: 0.9608
 - Similar triple: (india, accusation, uk), Similarity score: 0.9568
 - Similar triple: (china, accusation, ussr), Similarity score: 0.9546
- **Validation Triple: (cuba, reldiplomacy, china)**
 - Similar triple: (cuba, reldiplomacy, indonesia), Similarity score: 0.9731
 - Similar triple: (cuba, reldiplomacy, netherlands), Similarity score: 0.9683
 - Similar triple: (ussr, reldiplomacy, china), Similarity score: 0.9648
 - Similar triple: (cuba, reldiplomacy, uk), Similarity score: 0.9640
 - Similar triple: (egypt, reldiplomacy, china), Similarity score: 0.9614
- **Validation Triple: (egypt, embassy, uk)**
 - Similar triple: (poland, embassy, uk), Similarity score: 0.9573
 - Similar triple: (egypt, embassy, china), Similarity score: 0.9572
 - Similar triple: (egypt, embassy, usa), Similarity score: 0.9499
 - Similar triple: (israel, embassy, uk), Similarity score: 0.9492
 - Similar triple: (egypt, embassy, india), Similarity score: 0.9432

3.3 Observations

- **Similarity consistency:** TransR consistently found triples with a higher similarity score compared to TransE, likely due to its ability to better capture nuanced relation-specific embeddings.
- **Contextual similarity:** Both models provided contextually relevant triples for each validation example, demonstrating the utility of embedding-based similarity in knowledge graph completion.

3.4 Observations

- **Similarity Consistency:** TransR consistently found triples with higher similarity scores compared to TransE.
- **Contextual Relevance:** Both models provided relevant triples for each validation example, showcasing the effectiveness of embedding-based similarity.