Re-evaluating Embedding-Based Knowledge Graph Completion Methods

Farahnaz Akrami¹, Lingbing Guo², Wei Hu², Chengkai Li¹
University of Texas at Arlington¹, Nanjing University²

Motivation

• Importance of Knowledge Graphs (KG) for many AI-related applications such as question answering, web search, and fact checking.
• Incompleteness of KGs despite their large sizes.
• Popularity of embedding models among various KG completion methods.
• Prevalence use of the benchmark dataset FB15k to evaluate embedding methods.
• Existence of a bias in FB15k. It contains many pairs of \((h, r, t)\) and \((t, r^{-1}, h)\) where \(r^{-1}\) is inverse of \(r\). Therefore, the inverse of numerous test triples occurs in the training set.
• No previous investigation of the effect of the aforementioned bias in the results of embedding-based knowledge graph completion methods.

Embedding-Based Models

Steps employed by embedding-based methods:
1) defining a scoring function to measure the plausibility of triples \((h, r, t)\).
2) Learning the representations of \(h\), \(r\), and \(t\) by solving an optimization problem of maximizing the scores of correct triples while minimizing the scores of incorrect ones.

Benchmark Datasets

• FB15k [Bordes+NIPS13]: A subset of Freebase extensively employed for evaluating KG embedding approaches. Inverse triples of 81% of the test triples exist in the training set [Toutanova+CVSC15].
• FB15-237 [Toutanova+CVSC15]: A subset of FB15k created by removing inverse and near-duplicate relations from FB15k.

Results

Replacing \(h/t\) entity with all available entities in the dataset to form corrupted triples
Calculating score of each corrupted triple using \(f_r\)
Sorting scores by ascending order
Storing rank of the correct entity
Reporting mean of predicted ranks (MR), mean reciprocal rank (MRR), percentage of test triples that are ranked within top 10 (Hits@10)

• Performance reduction of all methods on FB15k-237:
  FMR of ConvE 68.9 (on FB15k) to 31 (on FB15k-237)

• Comparability of TransE on FB15k-237 to Many of its superior successors which outperformed TransE on FB15k:
  Fhits@10 of ANALOGY vs TransE 84.3 vs 61.8 (on FB15k) to 37.4 vs 42.5 (on FB15k-237)

• Superiority of ConvE results under many metrics.
• Promising results of observed feature models NLFeat and NeuralP.