Re-evaluating Embedding-Based Knowledge Graph Completion Methods

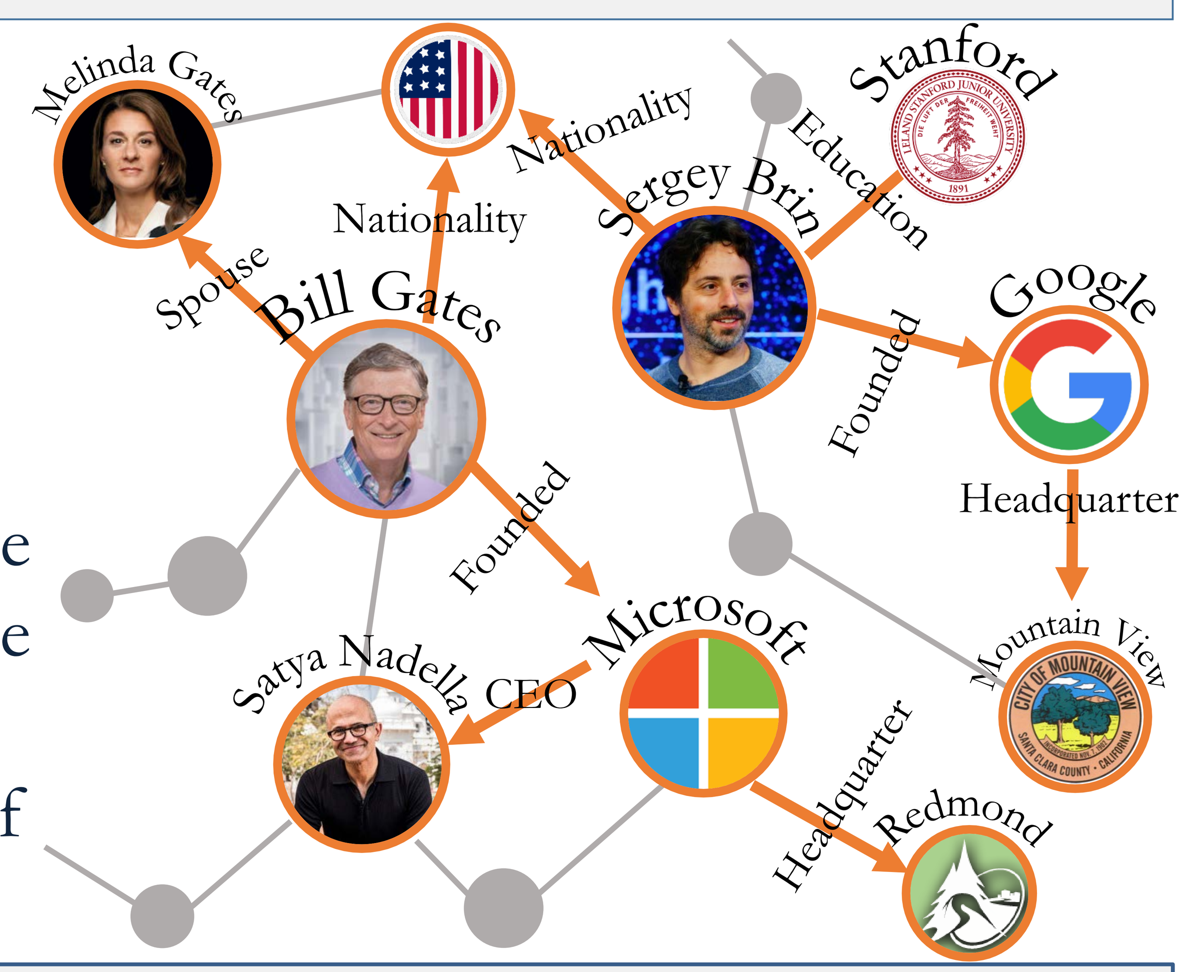
UNIVERSITY OF TEXAS ARLINGTON

Farahnaz Akrami[1], Lingbing Guo[2], Wei Hu[2], Chengkai Li[1]

University of Texas at Arlington[1], Nanjing University[2]

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 :

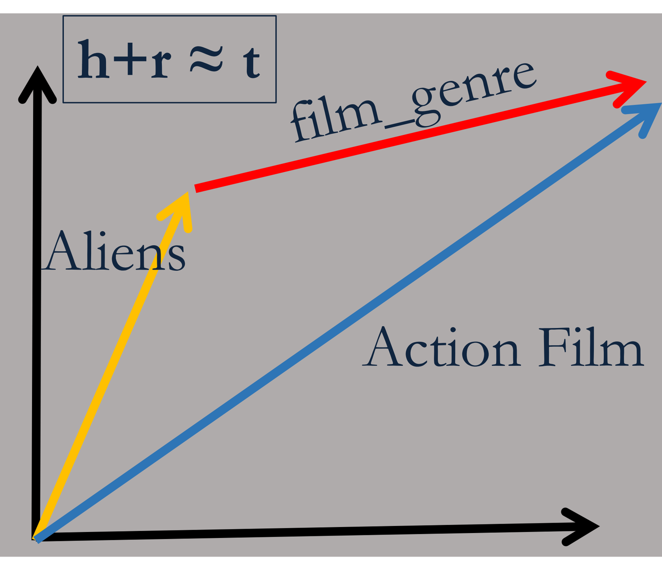
1) defining a scoring function to measure the plausibility of triples (h,r,t).
2) Learning the representations of h, r, and t by solving an optimization problem of maximizing the scores of correct triples while minimizing the scores of incorrect ones.

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.

TransE [Bordes+NIPS13]

- The very first embedding model
- One of the simplest embedding methods
- Vector representations of h, r, and t (**h, r, t**) are learned so if (h, r, t) holds then **h**+**r** ≈ **t**.



(h, r, t)= (Aliens, film_genre, Action Film)

Scoring function

$$f_r(\mathbf{h}, \mathbf{t}) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|^2_{\ell 1/2}$$

Link Prediction Task for Triples (h, r, t)

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)

Results

	FB15k						FB15k-237					
	Raw			Filtered			Raw			Filtered		
Model	MR↓	Hits@10↑	MRR↑	FMR↓	FHits@10↑	FMRR↑	MR↓	Hits@10↑	MRR↑	FMR↓	FHits@10↑	FMRR↑
TransE [Bordes+NIPS13]	243.0 201.0	34.9 43.4	— 18.44	125.0 70.2	47.1 61.8	- 30.7	- 440.2	- 29.8	- 11.9	- 250.8	- 42.5	- 18.0
TransH [Wang+AAAI14]	211.0 213.8	42.5 47.3	- 28.3	84.0 69.3	58.5 70.1	- 16.3	- 511.8	- 29.0	- 10.5	- 309.8	- 42.9	- 16.3
TransR [Lin+AAAI15]	226.0 236.4	43.8 47.2	- 16.2	78.0 82.7	65.5 71.9	- 29.7	- 544.9	- 27.9	- 9.9	- 337.0	- 42.9	- 16.2
TransD [Ji+ACL15]	211.0 209.8	49.4 47.4	- 16.3	67.0 65.4	74.2 70.4	- 28.3	- 506.9	- 29.4	- 10.4	- 305.2	- 42.8	- 16.2
RESCAL [Nickel+ICML11]	828.0 374.7	28.4 31.1	- 15.2	683.0 220.4	44.1 47.2	- 28.3	- 850.6	- 19.8	- 10.0	- 640.8	- 31.6	- 18.0
DistMult [Yang+ICLR15]	- 315.0 269.6 279.0 -	- 45.3 50.6 50.0 -	- 20.4 24.6 25.5 -	- 161.6 112.3 120.4 89.9	57.7 70.9 83.3 84.2 81.3	35 41.8 65.4 70.5 64.8	- 993.7 708.8 708.4 -	- 12.4 18.0 22.1 -	- 5.5 7.9 11.7 -	- 783.1 494.0 495.4 391.7	- 25.3 35.2 37.6 46.1	- 13.2 17.5 21.5 29.6
ComplEx [Trouillon+ICML16]	- 347.6 266.2 292.7 -	- 44.3 48.5 49.2 -	24.2 20.4 23.0 24.9 -	- 189.5 106.0 132.9 97.5	84.0 73.0 82.6 82.5 79.2	69.2 51.3 67.5 72.4 62.3	- 1169.2 630.7 708.5 -	- 8.2 18.7 21.1 -	- 3.9 8.1 11.3 -	- 955.1 415.7 495.1 456.5	- 20.7 36.9 36.7 45.7	- 10.9 18.4 20.9 28.6
ANALOGY [Liu+ICML17]	- 279.4	- 50.5	25.3 26.0	- 120.9	85.4 84.3	72.5 72.2	- 715.9	- 21.9	- 11.5	- 502.7	- 37.4	- 21.3
ConvE [Dettmers+AAAI18]	- 190.8	- 52.5	- 27.2	64.0 51.2	87.3 85.1	74.5 68.9	- 489.3	- 28.4	- 15.4	246.0 277.0	49.1 48.5	31.6 31.0
NLFeat [Toutanova+CVSC15]	-	-	-	-	87.0	82.2	-	-	-	-	34.7	22.6
NeuralLp [Yang+NIPS17]	-	-	-	-	83.7	76.0	-	-	-	-	36.2	24.0

• Published results • OpenKE : https://github.com/thunlp/OpenKE • ComplEx : https://github.com/ttrouill/complex
• ANALOGY : https://github.com/quark0/ANALOGY • ConvE : https://github.com/TimDettmers/ConvE

- Performance reduction of all methods on FB15k-237:

  FMRR 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 NeuralLP.

NSF NSFC