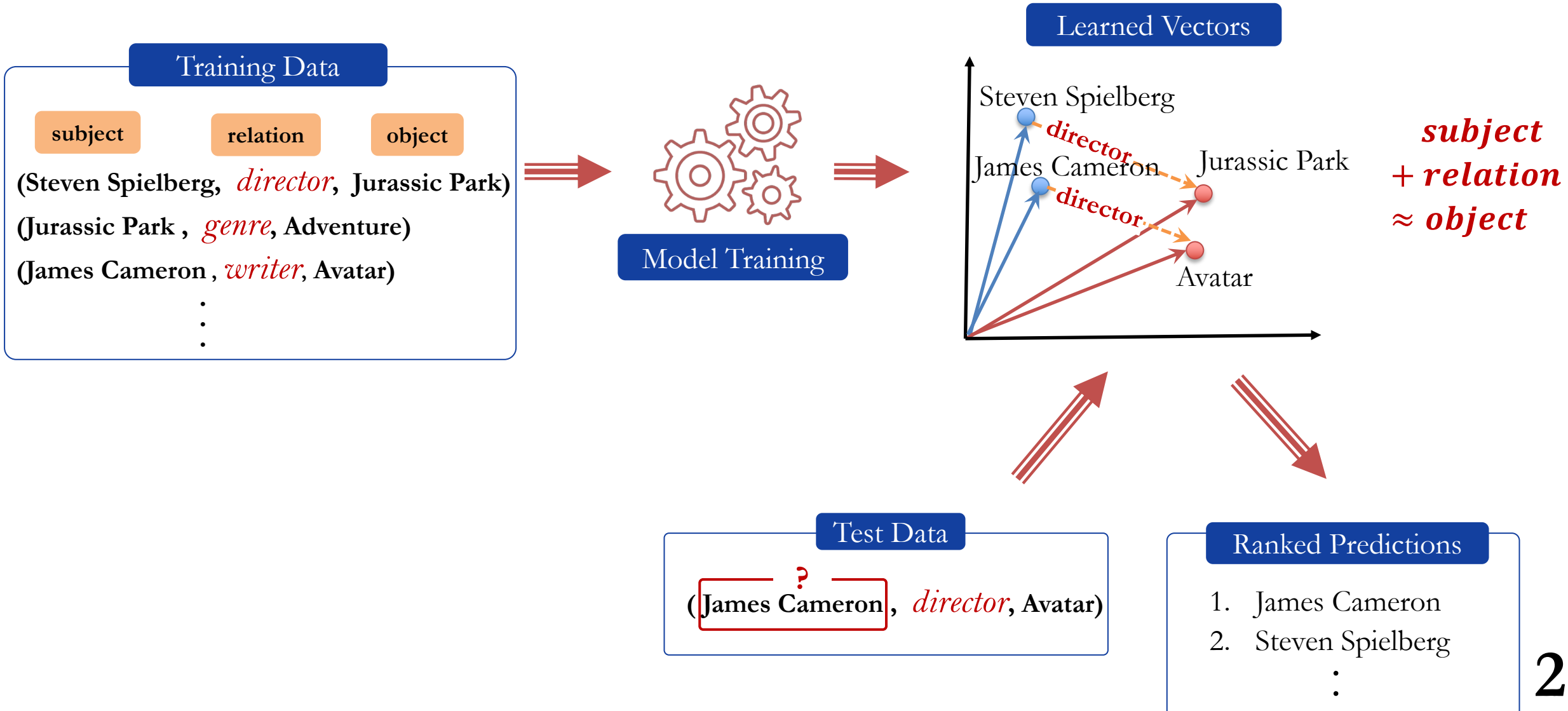
Realistic Re-evaluation of Knowledge Graph Completion Methods: An Experimental Study

Farahnaz Akrami[1], Mohammed Samiul Saeef[1], Qingheng Zhang[2], Wei Hu[2], Chengkai Li[1]

1-University of Texas at Arlington, 2-Nanjing University

ACM
SIGMOD
+PODS
2020

Embedding Models in Link Prediction



Evaluation of Embedding Models

- **Evaluation metrics**
  - MR, FMR
  - Hits@k, FHits@k
  - MRR, FMRR

- **Datasets**
  - FB15k, FB15k-237
  - WN18, WN18RR
  - YAGO3-10, YAGO3-10-DR

Model	Year	Performance	Citation
TransE	2013	47.1	2053
TransH	2014	58.5	1036
DistMult	2015	57.7	578
TransR	2015	65.5	1120
TransD	2015	74.2	452
DKRL	2016	57.6	245
HoLE	2016	73.9	431
TranSparse	2016	78.3	177
ComplEx	2016	84.0	468
R-GCN	2017	84.2	494
ANALOGY	2017	85.4	107
ProjE	2017	88.4	117
ConvE	2018	87.3	297
RotatE	2019	88.4	89
TuckER	2019	89.2	34

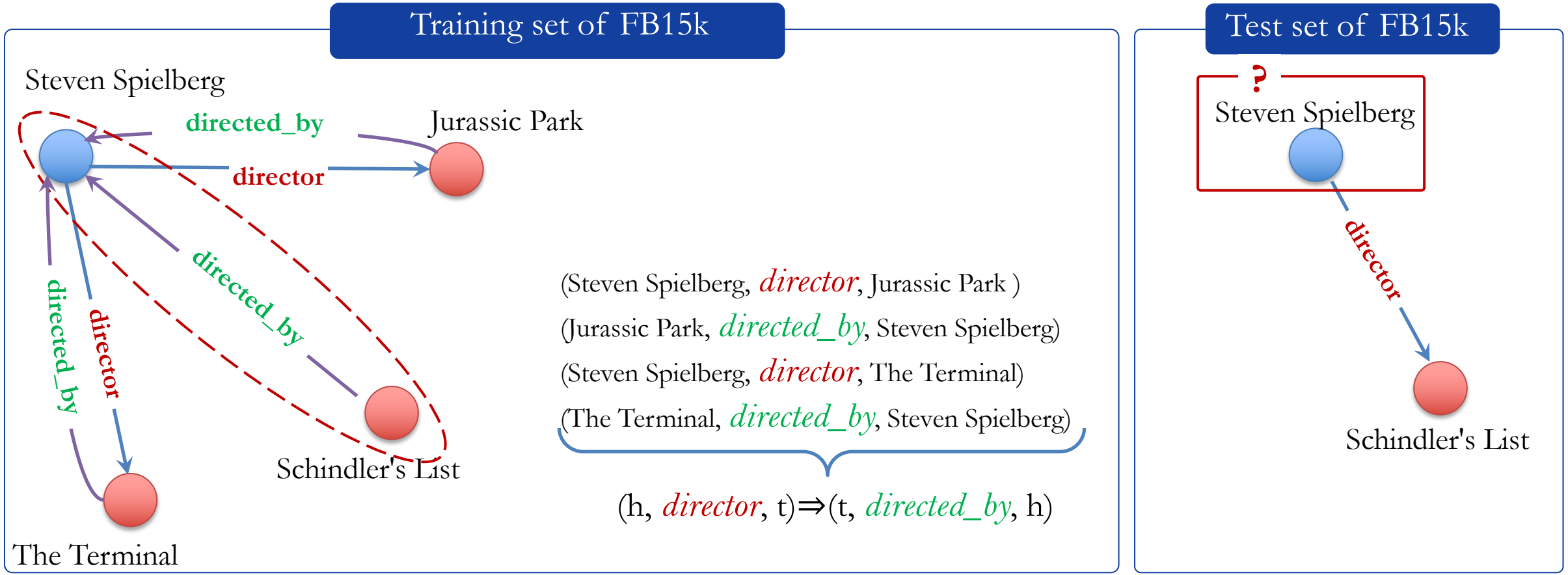
Findings from Our Re-Evaluation of the Models

Current link prediction methods do not work

1. Problems of the datasets substantially inflated the accuracy of these models.
   - Reverse relations
   - Duplicate relations
   - Cartesian product relations
2. Given the problematic data, very simple rules can challenge the accuracy of these complex machine learning models.
3. The true accuracy of these models is very poor on realistic data.
4. The link prediction task on the problematic datasets is nonexistent in the real-world.

FB15k Contains Many Pairs of Reverse Triples

- 70% of training triples are reverse of each other
- 70% of test triples have their reverse available in training set

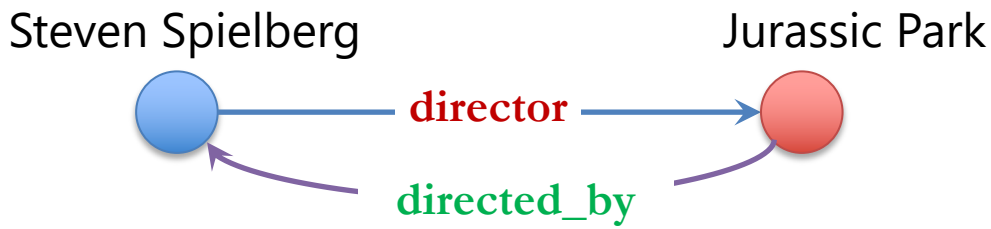


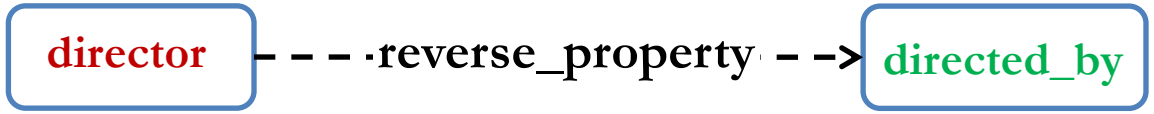
On Fb15k Very Simple Rules Can Challenge the Accuracy of Complex Machine Learning Models

$$(h, r_1, t) \Rightarrow (t, r_2, h)$$

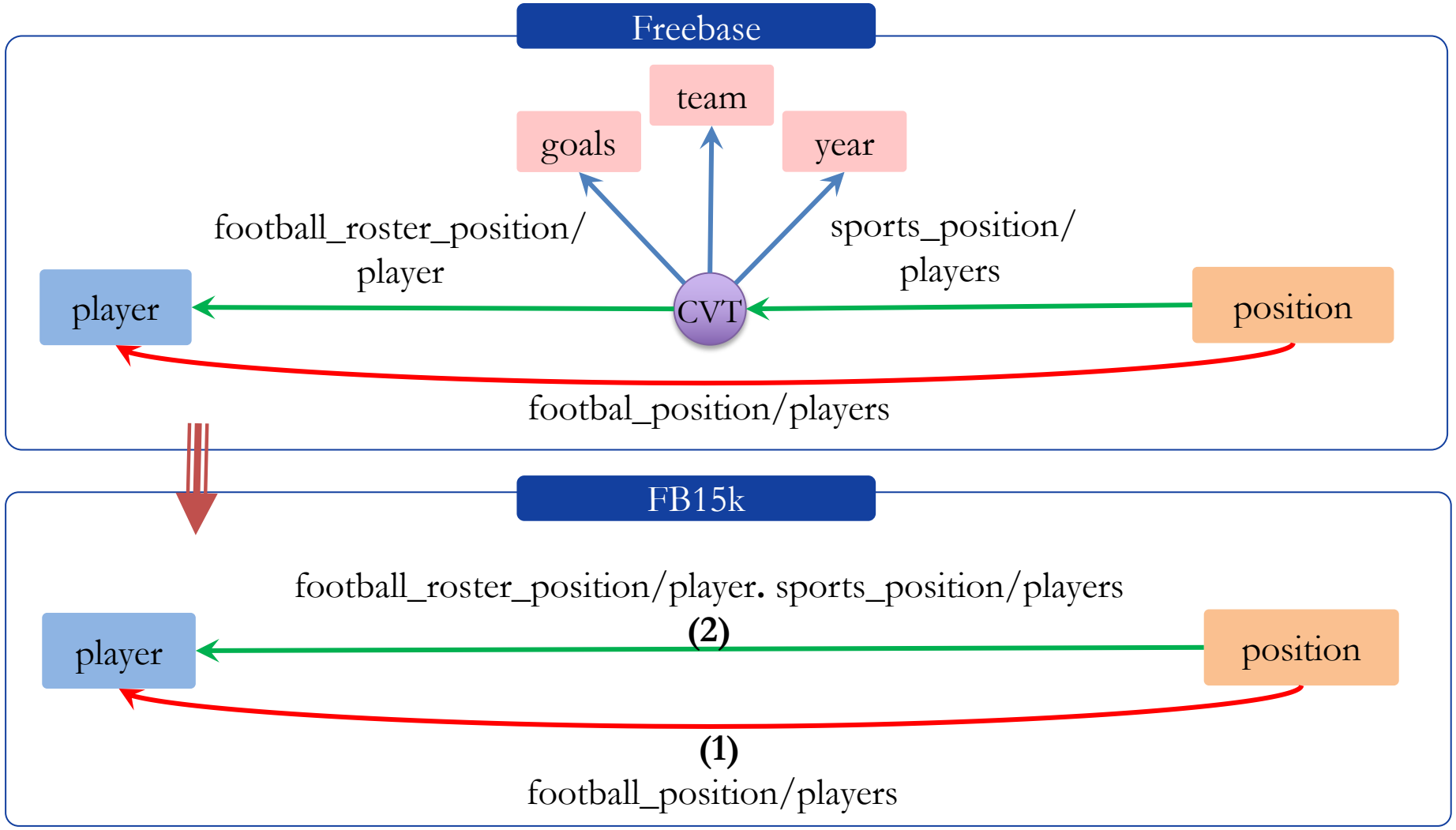
Model	Hits@1
TransE (2013)	26.9
ComplEx (2016)	59.5
ConvE (2018)	60.7
RotatE (2019)	73.8
TuckER (2109)	72.5
Simple Model	71.6

The Link Prediction Scenario on such Data Is Non-Existent in the Real-World

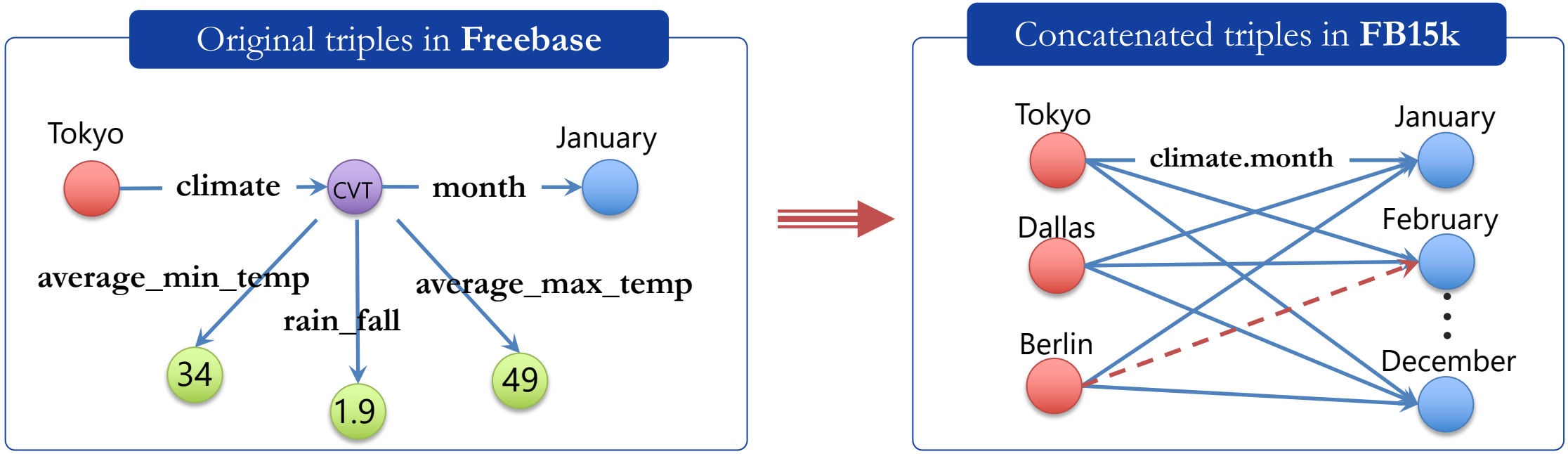




Out of 1345 Relations in FB15k, 168 Are Duplicate Relations



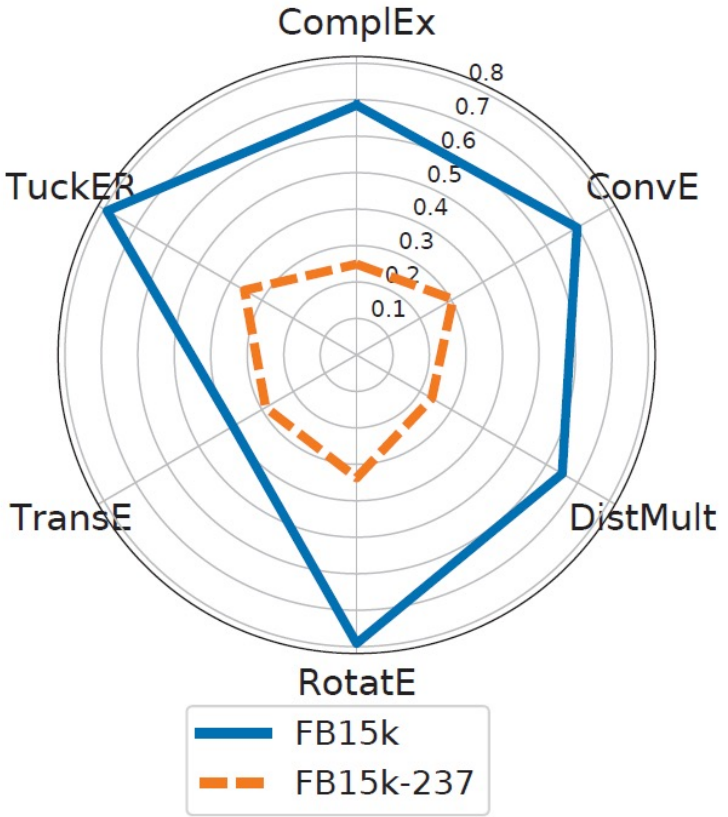
In FB15k, 142 out of the 1345 Relations Are Cartesian Product Relations



- The link prediction (whether a city has a climate in a month) is meaningless in the real-world.
- Simple method: If bipartite graph of a relation is almost complete, it is a Cartesian product relation; missing links in the complete bipartite graph are predicted to be true.
- On all 9 Cartesian product relations in FB15k , this simple model has an average FHits@10↑ of 98.3% which is higher than the 96.3% FHits@10↑ of TransE.

Findings: Current Link Prediction Methods Do Not Work

1. Problems of the datasets substantially inflated the accuracy of these models.
2. Given the problematic data, very simple rules can challenge the accuracy of these complex machine learning models.
3. The true accuracy of these models is very poor on realistic data.
4. The link prediction task on the problematic datasets is nonexistent in the real-world.



Performance of Embedding Models on Datasets without Data Redundancy

Similar or Even Worse Performance of Many Successors of TransE on Datasets without Data Redundancy

Model	TransE (2013)	ComplEx (2016)	ConvE (2018)	RotatE (2019)	TuckER (2109)
FMRR(FB15k)	0.392	0.685	0.698	0.791	0.790
FMRR(FB15k-237)	0.288	0.249	0.305	0.337	0.355

- Percentages of test triples, on which various models outperformed TransE, that have reverse and duplicate triples in training set

Model	ComplEx (2016)	ConvE (2018)	RotatE (2019)	TuckER (2109)
FB15k	84.67 %	78.04 %	78.16%	78.87%

Related Work

- [Toutanova and Chen, 2015]
  - Detected the reverse relation problem in FB15k and generated FB15k-237
- [Dettmers et al., 2018]
  - Detected the reverse relation problem in WN18 and generated WN18RR
- [Meilicke et al., 2018]
  - Analysis of the types of rules that help knowledge graph completion
- [Akrami et al. 2018]
  - Reevaluation of embedding models
- [Wang et al., 2019]
  - Analysis of existing evaluation methods
- This paper for the first time
  - Conducts a systematic study with the main objective of assessing the true effectiveness of embedding models when the unrealistic triples are removed
  - Provides a thorough investigation of the data redundancy problem and identifies Cartesian product relations.

Take-Home Messages

1. Our community hasn't been solving real-world challenges in link prediction.
2. Current link prediction methods, especially the embedding models, have poor accuracy on realistic data.
3. We shouldn't use FB15k, WN18, and YAGO3-10 at all.

- GitHub repository of all source codes, datasets, and results
  - https://github.com/idirlab/kgcompletion