Improving Search Results Ranking Using a Knowledge Graph

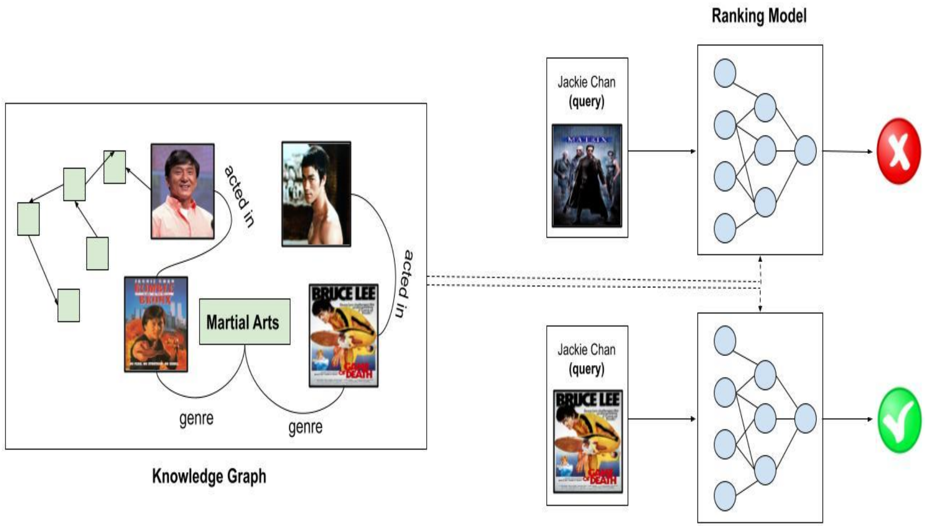
Hiba Ahsan (University of Massachusetts Amherst),
Sudarshan Lamkhede (Netflix Research)

Motivation

- Augmenting machine learned models with knowledge graphs (KGs) can help add domain-specific knowledge to models.
- Tasks such as question-answering, named entity recognition have benefitted from this augmentation.
- KGs have been successfully used in search as well for ranking.

Task

Improve the quality of search results using a KG



How can we use a KG to improve our model?

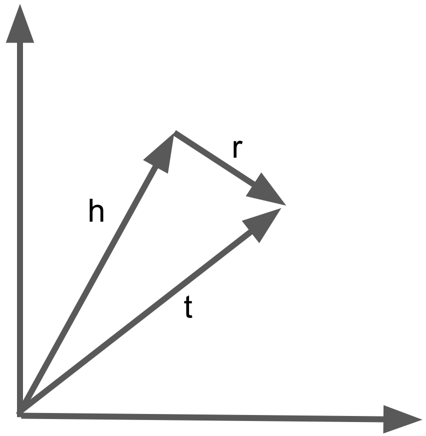
- KG-aware learning methods can be classified into two categories:
    - **Embedding-based methods**
    - **Regularization-based methods**
- Baseline model is a neural network that takes features derived from query, entity, and query-entity as inputs.
- Incorporated KG-based embeddings of the entity (and also query in case of KEWER) to make the model KG-aware

How can we use a KG to improve our model?

- We experimented with the following approaches to learn the embeddings:
  - TransE embeddings
  - Knowledge Graph Attention Network for Recommendation (KGAT)
  - Knowledge graph Entity and Word Embedding for Retrieval (KEWER)

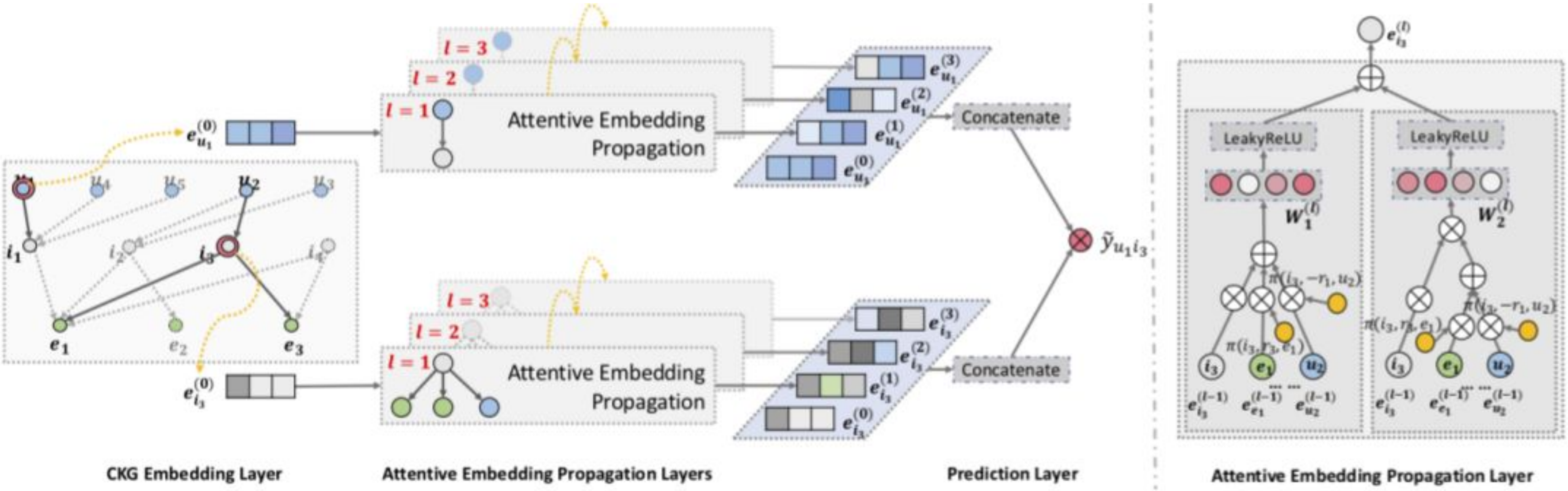
TransE

View relations as translations performed on entity embeddings.



[Translating Embeddings for Modeling Multi-relational Data](), Antoine Bordes et. al., NeuRIPS 2013

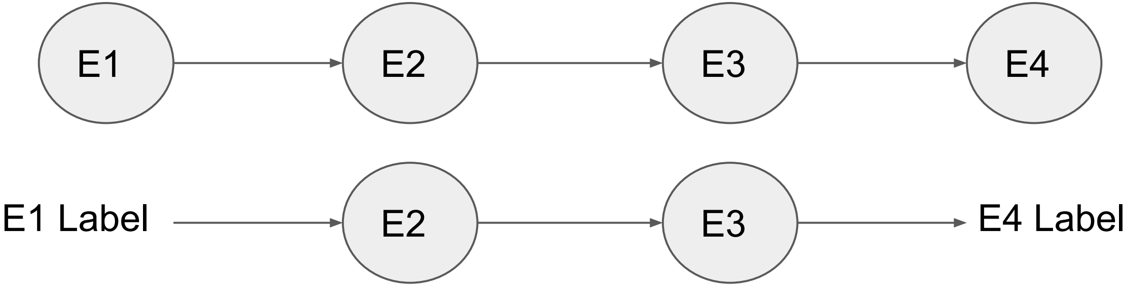
KGAT



[Knowledge graph attention network for recommendation](). Xiang Wang, et. al., KDD 2019.

KEWER

Entity search model which embeds words and entities in the same space.



Joint Word and Entity Embeddings for Entity Retrieval from a Knowledge Graph, F. Nikolaev, A. Kotov, ECIR 2020

Experiments

- Training data: 996,332 examples of query-entity features and binary relevance label tuples.
- Knowledge Graph:
    - 508,725 entities
    - 7,887,096 head-relation-tail triples
    - 20 entity types including Person, Video (i.e. movie or TV series), Genre (e.g. thriller).
    - 18 relations including acted_in, directed_by.
- Trained 64 dimensional embeddings of entities separately and added them as features into the baseline model.
- Evaluation based on Normalized Mean Reciprocal Rank (NMRR) metric.

Results

Model	# Model Parameters	NMRR Gain
Baseline	870,401	-
Baseline + KGAT entity embeddings	878,593	0%
Baseline + KEWER entity embeddings	878,593	1.36%
Baseline + KEWER entity and query embeddings	886,785	1.36%
Baseline + TransE entity embeddings	878,593	2.56%

Observations

- Incorporating transE embeddings of entities resulted in the largest NMRR improvement.
- KGAT did not scale well to large datasets.
- In case of KEWER, entity embeddings improved the NMRR compared to the baseline by 1.36%.
- Additionally incorporating query embeddings did not result in further improvement of the NMRR.

Future Work

- Scale up our experiments in terms of data and model training.
- Evaluate the improvements across different query classes.
- Explore jointly training the embedding model and the ranking model.
- Explore how to represent queries in KGs.