Using Knowledge Graphs to Represent Knowledge and Reason

Yanjian Zhang
Facultad de Informática de Barcelona
Universitat Politècnica De Catalunya
yanjian.zhang@outlook.com

Chun Han Li
Facultad de Informática de Barcelona
Universitat Politècnica De Catalunya
spencer850922@gmail.com

Abstract

Knowledge Graph has provided a very flexible way and explicit way to represent real-world knowledge through building up graphs connecting entities with relations. As a result of its enriching representation ability, Knowledge Graph has been frequently used for reasoning task. In this paper, we would go through knowledge representation methods and knowledge reasoning strategies that exist in recent years. We also provide the comparison between the performance of the most trendy models on the most used reasoning dataset.

1 Introduction

The introduction of human knowledge is one of the research directions of AI. Knowledge representation and reasoning is to provide artificial intelligence systems with knowledge that they can process, making them human-like capable of solving complex tasks.(Newell et al. [1959]) Knowledge graphs, which can represent human knowledge in a structured form, have attracted extensive attention in both academia and industry in recent years.

A knowledge graph is a structured representation of facts, entities, relationships, and semantic descriptions. Entities can be real-world objects or abstract concepts. Relationships represent relationships between entities. Semantic descriptions of entities and relationships between entities contain predefined types and attributes. A property graph is a widely used type of graph in which nodes and relationships have their own properties.

To facilitate the future research, there are reviews and surveys focusing on knowledge representation(Ji et al. [2021]) and knowledge reasoning(Chen et al. [2020]) respectively. Another survey focusing on Knowledge Graph-Based Recommender Systems(Guo et al. [2020]), which is one of sections of our work. Guo et al. [2020] compare the the Knowledge graph usage by knowledge graph type, knowledge embedding and issue resolved, while our approach is to analyse the Knowledge Graph usage in two dimensions, representation and reasoning. Since knowledge embedding is only one of knowledge representation methods mentioned in our review, we could say that our reviews are more inclusive and over-rounded.

While some of them has covered parts of the existing method of either part, a.k.a. knowledge representation or knowledge reasoning, none of they have present them jointly and provided the performance over existing task and dataset of existing models.

Our contribution in this paper is three-folds:

- We present knowledge representation and reasoning jointly. Instead of treating them separately like previous works did, we think that this two dimension is tightly related.

- We compare the existing knowledge representation and reasoning models over tasks and datasets. We think the direct comparison over datasets would be more straight-forward and would facilitate the future research even more.
- We extensively cover the most trendy knowledge representation and reasoning method in recent years, including transformer-based knowledge encoding(Wang et al. [2019c]), GNN-based knowledge dissemination(Schlichtkrull et al. [2017]), and meta-relation learning(Ali et al. [2021]).

A limitation of this paper is that we have not explored knowledge acquisition methods, which are covered by previous review(Ji et al. [2021]), including knowledge graph completion, link prediction , triple classification and etc. Instead, we focus more on downstream representation and reasoning after we completed knowledge acquisition.

2 Related Work

Figure 1: Knowledge Representation Trends

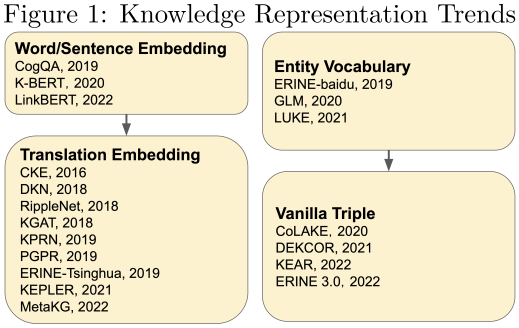
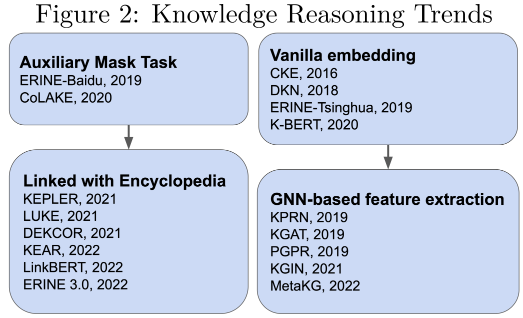


Figure 2: Knowledge Reasoning Trends



Knowledge Representation Methods

There are two trends of knowledge representation methods, as shown in Figure 1. The first trend is to embed the triple, the most straight-forward way is to use word embedding or sentence embedding. Examplary works are CogQA(Ding et al. [2019]), K-BERT(Liu et al. [2020]), LinkBERT(Yasunaga et al. [2022]).

With the exploitation of Translation Distance getting deeper, most of the works have turned to utilize the translation embedding which is learned from Translation Distance to embed the Knowledge Graph. Examplary works are CKE(Zhang et al. [2016]), DKN(Wang et al. [2018c]), RippleNet(Wang et al. [2018b]), KGAT(Wang et al. [2019d]), KPRN(Wang et al. [2019e]), PGPR(Xian et al. [2019]), ERINE-Tsinghua(Zhang et al. [2019]), KEPLER(Wang et al. [2021a]), MetaKG(Du et al. [2022]).

The second trend is to use literally the triple itself. With the development of large-scale pre-trained, methods like masking has been popular for several years, and therefore some works have directly treat the entities in Knowledge Graph as a vocabulary, which means when any combination of words are mentions in KG, then they are masked as a whole, Exemplary works are ERNIE-baidu(Sun et al. [2019]), GLM(Shen et al. [2020]), and LUKE(Yamada et al. [2020]).

In order to fully utilize the relations between entities, further works have use both entities and relation together. Exemplary works are CoLAKE(Sun et al. [2020]), DEKCOR(Xu et al. [2020]), KEAR(Xu et al. [2022]), and ERINE 3.0(Sun et al. [2021]).

Knowledge Reasoning Strategies

As shown in Figure 2, two trends of Knowledge Reasoning have been observed by us. To enhance the reasoning ability of large-scale pre-trained by KG-related tasks or resources. Some papers focusing on inserting the knowledge into language model through auxiliary masking task, like ERNIE-baidu(Sun et al. [2019]) and CoLAKE(Sun et al. [2020]).

The other treat Knowledge graph as a bridge, a.k.a. using the entities found in sentence to find entity-related information in encyclopedia. The Exemplary works of the latter one are KEPLER(Wang et al. [2021a]), LUKE(Yamada et al. [2020]), DEKCOR(Xu et al. [2020]), KEAR(Xu et al. [2022]), LinkBERT(Yasunaga et al. [2022]), ERINE 3.0(Sun et al. [2021]).

The second trend is to map Knowledge Graph information into high dimension feature. Some of works directly use the embedding itself, either word embedding or translation embedding, CKE(Zhang et al. [2016]), DKN(Wang et al. [2018c]), ERINE-Tsinghua(Zhang et al. [2019]), K-BERT(Liu et al. [2020]).

In the meanwhile, the most recent works have made use of Graph-based neural network to capture the Graph structure inside Knowledge Graph to achieve better result. The Exemplary works of the latter one are KGAT(Wang et al. [2019d]), KPRN(Wang et al. [2019e]), PGPR(Xian et al. [2019]), KGIN(Wang et al. [2021b]), MetaKG(Du et al. [2022]).

Model	Knowledge Representation Method	Reasoning Strategy
ERNIE-Tsinghua	Translational Distance(TransE)	Concatenation of word token and entity embedding
K-BERT	Word Embedding on triple	Query KG to feteh triple for explaining entity in Sentence
KEPLER	Translational Distance(TransE)	Train Language Encoder with entity description
ERNIE-Baidu	As a Entity Vocabulary	Use Entity Vocabulary to guide masking for pretrained model
CoLAKE	plain textual triple	Predict Entity mask and Relation mask
ERNIE 3.0	plain textual triple	Use embedded encyclopedia to predict the triple relation
KEAR	plain textual triple	Related knowledge is retrieved from external sources, using the input as the key to train the transformer architecture.
DEKCOR	plain textual triple	Model training, the input to ALBERT includes the question, choice, entity names, description text and triple.
GLM	As a Entity Vocabulary	Entity-level masked + negative sample entity pairs.
LUKE	As a Entity Vocabulary	Trained on a large number of entity-annotated corpora obtained from Wikipedia.
LinkBERT	Word Embedding on triple	Train the LM with two self-supervised objectives: masked language modeling, document relation prediction.
CogQA	Word Embedding on triple	Combines pre-trained language models and graph neural networks in terms of reasoning ability and interpretability.
CKE	Translational Distance(TransR)	Add TransR vector to textual embed directly
DKN	Translational Distance(TransD)	Use Knowledge Embedding to enhance CNN feature extraction
MKR	Semantic matching models	Multi-task for crossing feature of entities and items
RippleNet	Translational Distance	Assign Triple based on revelance of item, head and relation representation
KPRN	Translational Distance	Use LSTM to triple path embedding
PGPR	Translational Distance(TransE)	Using Reinforcement Learning to guide path reasoning
KGAT	Translational Distance(TransR)	Knowledge-aware Attention
KGIN	Translational Distance(TransR)	GNN-based Path-aware aggregation
MetaKG	Translational Distance(TransR)	GNN-based aggregation entity representation

Table 1: The Knowledge Representation method and Reasoning strategies of each model

3 Dataset and Metrics

As mentioned in previous reviews(Ji et al. [2021]), there are three major types of datasets for Knowledge Graph application: Language representation, Question Answering and Recommendation System.

3.1 Dataset for Language Representation

We introduced FewRel(Han et al. [2018]), Open Entity(Choi et al. [2018]), GLUE(Wang et al. [2018a]), XNLI(Conneau et al. [2018]), SuperGLUE(Wang et al. [2019a]) tasks for comparison the performance of existing model over language representation tasks.

- **FewRel**: FewRel dataset(Han et al. [2018]) is a large-scale few-shot relation extraction dataset, which contains more than one hundred relations and tens of thousands of annotated instances cross different domains.
- **Open Entity**: The Open Entity dataset(Choi et al. [2018]) is a collection of about 6,000 sentences with fine-grained entity types annotations. The entity types are free-form noun phrases that describe appropriate types for the role the target entity plays in the sentence. Sentences were sampled from Gigaword, OntoNotes and web articles. On average each sentence has 5 labels.
- **GLUE**: General Language Understanding Evaluation benchmark (Wang et al. [2018a]) is a collection of tools for evaluating the performance of models across a diverse set of existing NLU tasks. By including tasks with limited training data, GLUE is designed to favor and encourage models that share general linguistic knowledge across tasks.
- **XNLI**: The Cross-lingual Natural Language Inference corpus (Conneau et al. [2018]) is a crowd-sourced collection of 5,000 test and 2,500 dev pairs for the MultiNLI corpus. The pairs are annotated with textual entailment and translated into 14 languages: French, Spanish, German, Greek, Bulgarian, Russian, Turkish, Arabic, Vietnamese, Thai, Chinese, Hindi, Swahili and Urdu.
- **SuperGLUE**: SuperGLUE is a new benchmark (Wang et al. [2019a]) styled after GLUE with a new set of more difficult language understanding tasks, improved resources, and a new public leaderboard.

3.2 Dataset for Question Answering

Question Answering is the task of answering questions (typically reading comprehension questions), but abstaining when presented with a question that cannot be answered based on the provided context.

Question answering can be segmented into domain-specific tasks like community question answering and knowledge-base question answering. Popular benchmark datasets for evaluation question answering systems include SQuAD(Rajpurkar et al. [2016]), CommonsenseQS(Talmor et al. [2018]), HotpotQA(Yang et al. [2018]), and many others. Models for question answering are typically evaluated on metrics like EM(exact match) and F1. More details of the datasets are as follows:

- **CommonsenseQA**: The CommonsenseQA dataset(Talmor et al. [2018]) is a dataset for commonsense question answering task. The dataset consists of 12,247 questions with 5 choices each.
- **HotpotQA**: HotpotQA dataset(Yang et al. [2018]) is a question answering dataset featuring natural, multi-hop questions, with strong supervision for supporting facts to enable more explainable question answering systems.
- **SQuAD**: Stanford Question Answering Dataset(Rajpurkar et al. [2016]) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.

3.3 Dataset for Recommendation System

The following datasets are used to track progress in Recommendation System.

- **MovieLens-1M**: MovieLens-1M[1] is a widely used benchmark dataset in movie recommendations, which consists of approximately 1 million explicit ratings (ranging from 1 to 5) on the MovieLens website.
- **Bing-News**: Bing-News dataset contains 1,025,192 pieces of implicit feedback collected from the server logs of Bing News[2] from October 16, 2016 to August 11, 2017. Each piece of news has a title and a snippet.
- **Last-FM**: Last-FM[3] is a music listening dataset provided by Last.fm online music system. The Last.fm dataset consists of two kinds of data at the song level: tags and similar songs.
- **Amazon-Book**: Amazon-Book dataset[4] contains product reviews and metadata from Amazon(He and McAuley [2016]), including 142.8 million reviews spanning May 1996 - July 2014.

4 Existing Models and Performance

We have selected ERNIE-Tsinghua(Zhang et al. [2019]), K-BERT(Liu et al. [2020]), KEPLER(Wang et al. [2021a]), ERNIE-Baidu(Sun et al. [2019]) , GLM(Shen et al. [2020]), CoLAKE(Sun et al. [2020]), ERNIE 3.0(Sun et al. [2021]) for comparison on Language representation performance with Knowledge Graph. The selected models are consisted of the most recent models including the State-of-the-Art of given datasets. Some brief summary of those models are as follows:

- **ERNIE-Tsinghua**: ERNIE-Tsinghua model(Zhang et al. [2019]) utilized both large-scale textual corpora and KGs to train an enhanced language representation model, take full advantage of lexical, syntactic, and knowledge information simultaneously
- **K-BERT**: Knowledge-enabled language representation model (Liu et al. [2020]) with knowledge graphs (KGs), in which triples are injected into the sentences as domain knowledge
- **KEPLER**: A unified model for Knowledge Embedding and Pre-trained LanguagERepresentation (Wang et al. [2021a]), which can not only better integrate factual knowledge into PLMs but also produce effective text-enhanced Knowledge Embedding with the strong PLMs
- **ERNIE-Baidu**: ERNIE-Baidu(Sun et al. [2019]) is designed to learn language representation enhanced by knowledge masking strategies, which includes entity-level masking and phrase-level masking.
- **LUKE**: LUKE(Yamada et al. [2020]) is based on the transformer, using RoBERTa as the base pre-training model, trained on a large number of entity-annotated corpora obtained from Wikipedia. It treats both words and entities as separate tokens, and computes the intermediate and output representations of all tokens through the Transformer. Since entities are also used as tokens, LUKE can model the relationship between entities.
- **CoLAKE**: Contextualized Language and Knowledge Embedding (Sun et al. [2020]) jointly learns contextualized representation for both language and knowledge with the extended MLM objective
- **ERINE 3.0**: ERINE 3.0(Sun et al. [2021]) was trained with 10 billion parameters on a 4TB corpus consisting of plain texts and a large-scale knowledge graph.

[1] https://grouplens.org/datasets/movielens/1m/
[2] https://www.bing.com/news
[3] http://www.millionsongdataset.com/lastfm/
[4] http://jmcauley.ucsd.edu/data/amazon/

- **PLM**(Pre-trained Language Model) backbone, including BERT(Devlin et al. [2018]) and RoBERTa(liu2019roberta) is given for fair comparison.

There are many different methods targeting on different types of question answering datasets. Therefore, we perform model comparisons separately on some of the most representative data to find our state-of-the-art model.

- **KEAR**: The Knowledgeable External Attention for commonsense Reasoning (KEAR) (Xu et al. [2022]) is augmented with an external attention mechanism, combining context with external knowledge, which achieves the same level as humans on the CommonsenseQA.
- **DEKCOR**: DE-scriptive Knowledge for Commonsense Reasoning(Xu et al. [2020]), leverage external entity descriptions to provide contextual information for graph entities. For the CommonsenseQA, extract concepts in questions and options and find triples related to those concepts, then extract the descriptions of these concepts from the Wkitionary and feed it to the pretrained model as an additional input of triples.
- **LinkBERT**: This is an LM pretraining method(Yasunaga et al. [2022]) that exploits links between documents, such as hyperlinks. Given a text corpus, they treat it as a document graph and create LM inputs by placing linked documents in the same context. They then pre-train the LM with two joint self-supervised objectives: masked language modeling and our new scheme document relation prediction.
- **CogQA**: CogQA(Ding et al. [2019]) has two systems. First, composed of BERT, the function is to extract key information, such as the entity node of the next hop, the current possible answer node. Second, GCN, used for reasoning, build a graph from the input of system 1, and reason on this graph to get the final answer.
- **GLM**: Shen et al. [2020] aim at equipping pre-trained LMs with structured knowledge through novel self-supervised tasks.

For recommdation scenario, some quite different types of models are proposef to utilize the Knowledge Graph, the model are listed as follows.

- **CKE**: Collaborative Knowledge Base Embedding(Zhang et al. [2016]) jointly learn the latent representations in collaborative filtering as well as items' semantic representations from the knowledge base.
- **DKN**: Deep knowledge-aware network (Wang et al. [2018c]) incorporates knowledge graph representation into news recommendation. DKN is a content-based deep recommendation framework for click-through rate prediction.
- **MKR**: Multi-task feature learning approach for Knowledge graph enhanced Recommendation(Wang et al. [2019b]). MKR is a deep end-to-end framework that utilizes knowledge graph embedding task to assist recommendation task.
- **RippleNet**: RippleNet(Wang et al. [2018b]) is an end-to-end framework that naturally incorporates the knowledge graph into recommender systems. Similar to actual ripples propagating on the water, RippleNet stimulates the propagation of user preferences over the set of knowledge entities.
- **KPRN**: Knowledgeaware Path Recurrent Network(Wang et al. [2019e]) to exploit knowledge graph for recommendation. KPRN can generate path representations by composing the semantics of both entities and relations.
- **PGPR**: Policy-Guided Path Reasoning(Xian et al. [2019]), which couples recommendation and interpretability by providing actual paths in a knowledge graph.
- **KGAT**: Knowledge Graph Attention Network(Wang et al. [2019d]) explicitly models the high-order connectivities in KG in an end-to-end fashion.
- **KGIN**: Knowledge Graph-based Intent Network(Wang et al. [2021b]) models each intent as an attentive combination of KG relations, encouraging the independence of different intents for better model capability and interpretability.

- **MetaKG**: A Meta-learning based framework(Du et al. [2022]), which encompasses a collaborative-aware meta learner and a knowledge-aware meta learner, to capture meta users' preference and entities' knowledge for cold-start recommendations.

Model	PLM backbone	FewRel (F1)	Open Entity(F1)	GLUE	XNLI	SuperGLUE
BERT-base	-	84.9△	73.6△	79.7	76.3	
BERT-large	-			82.1		69.0♡
RoBERTa-base	-	85.3△	75.4△	86.4△		
RoBERTa-large	-			**88.5**		
ERNIE-Tsinghua	-	88.3	75.6			
K-BERT	BERT-base				77.0	
KEPLER	-			84.9△		
ERNIE-Baidu	-				78.4	
CoLAKE	-	**90.6**	76.4	86.3		
LUKE	RoBERTa-large		**78.2**			
ERNIE 3.0	-				**83.8**	**90.6**

Table 2: The performance on each Language Representation-related dataset, since some of the result are not from original paper, we use △ to represent the result is from CoLAKE(Sun et al. [2020]), ♡ to present result from SuperGLUE(Wang et al. [2019a]), - means the model does not use PLM backbone, traning totally from scratch.

Model	PLM backbone	CommonsenseQA (ACC)	SQuAD 1.1 (F1)	HotpotQA (F1)
BERT-large	BERT-large	55.9♡	88.9	39.1
RoBERTa-large	RoBERTa-large	72.1△		
ALBERT-large	ALBERT-large	76.5		
CogQA	BERT-large			57.7
LinkBERT	BERT-large		92.7	**80.8**
LUKE	RoBERTa-large		**95.4**	
GLM	RoBERTa-large	74.1		
DEKCOR	ALBERT-large	84.7		
KEAR	RoBERTa-large	**89.4**		

Table 3: The performance of each model on Quetion Answering datasets. Since some of the result are not from original paper, we use △ to represent the result is from GLM(Shen et al. [2020]), ♡ to represent the result is from CommonsenseQA(Talmor et al. [2018])

Multirow	PLM backbone	MovieLens-1M	Bing-News	Last-FM		Amazon-Book	
		AUC	AUC	AUC	NDCG@20	Recall	NDCG@20
CKE	-	80.1△	55.3△	74.4△		13.4♡	23.9♢
DKN	-	66.5△	65.9	60.2△			
MKR	-	91.7	**68.9**	**79.7**			
RippleNet	-	**92.1**	67.8	76.8△	16.1♢	13.4♡	17.4♢
KGAT	-				27.4♢	14.9	27.1♢
KGIN	-					16.9	
MetaKG	-				**36.0**		**32.1**

Table 4: The performance of each model on Recommendation System dataset, Since some of the result are not from original paper, we use △ to represent the result is from MKR(Wang et al. [2019b]), ♡ to represent the result is from KGAT(Wang et al. [2018b]), ♢ to represent the result is from MetaKG(Du et al. [2022]). - means the model does not use PLM backbone, training totally from scratch.

5 Performance Analysis

According to Table 2, we could conclude that ERNIE 3.0 achieved State-of-the-Art performance in both XNLI and SuperGLUE. Even though there is not published result for ERINE 3.0 over GLUE, since We think this is due to its utilization of large amount of corpora and encyclopedia(Wikipedia, Baidu Baike, feed, etc.), linked with the 50 Million facts included in Baidu Knowledge Graph and (Sun et al. [2021]).

A more feasible and easy-to-follow result is CoLAKE, which only use pre-training on structured, unlabeled word-knowledge graphs, and integrate the language context and knowledge context in a unified data structure(Sun et al. [2020]). This model also achieved the SOTA performance among published results of FewRel in F1 score.

LUKE is another good choice, which achieved the SOTA among published results of Open Entity in F1 score, shown in Table 2, and it also achieve SOTA in SQuAD 1.1, as shown in Table 3. It use RoBERTa as the basic pre-training model, and train on a large number of entity annotation corpora obtained from Wikipedia. KEAR achieves the best results in another one-hop reasoning dataset, CommonsenseQA. It uses an external attention mechanism, combining context with external knowledge and integrating external information(Yamada et al. [2020]). It achieves human accuracy, with 89.4 accuracy compared to 88.9 human accuracy(Xu et al. [2022]).

In multi-hop reasoning scenario, LinkBERT has the best performance in HotpotQA, shown in Table 3. It learns the knowledge and dependencies between documents through pre-training tasks(Yasunaga et al. [2022]), and achieves good results in multi-hop reasoning tasks.

As shown in Table 4, MetaKG achieved the best result in Last-FM and Amazon-Book dataset. It is a good representation of combining the best practise in both knowledge represenation and knowledge reasoning. It not only used the Translation Embedding, but also utilized GNN-based feature extraction for information aggregation over the Knowledge Graph.

6 Conclusion

In this review, we disentangle the knowledge graph modeling in downstream application into two dimension: Knowledge representation and Knowledge Reasoning. We have provided the modeling mechanisms for each model case by case and summarised them chronologically with trends in both dimensions. We have provided detailed comparison among 21 most trendy models along with their PLM backbone within 11 datasets, covering topics Language Representation, Question Answering and Recommendation System. Lastly, we provided analysis based on our comparison and provide suggestion on the model selection for further research direction.

References

M. Ali, M. Berrendorf, M. Galkin, V. Thost, T. Ma, V. Tresp, and J. Lehmann. Improving inductive link prediction using hyper-relational facts. In *The Semantic Web – ISWC 2021: 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24–28, 2021, Proceedings*, page 74–92, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-030-88360-7. doi: 10.1007/978-3-030-88361-4_5. URL https://doi.org/10.1007/978-3-030-88361-4_5. 1

X. Chen, S. Jia, and Y. Xiang. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 141:112948, 2020. 1

E. Choi, O. Levy, Y. Choi, and L. Zettlemoyer. Ultra-fine entity typing. *arXiv preprint arXiv:1807.04905*, 2018. 3.1

A. Conneau, R. Rinott, G. Lample, A. Williams, S. R. Bowman, H. Schwenk, and V. Stoyanov. Xnli: Evaluating cross-lingual sentence representations. In *Proceedings of the*

2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018. 3.1

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 4

M. Ding, C. Zhou, Q. Chen, H. Yang, and J. Tang. Cognitive graph for multi-hop reading comprehension at scale, 2019. URL https://arxiv.org/abs/1905.05460. 2, 4

Y. Du, X. Zhu, L. Chen, Z. Fang, and Y. Gao. Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2022. 2, 2, 4, 4

Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, and Q. He. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 2020. 1

X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. *arXiv preprint arXiv:1810.10147*, 2018. 3.1

R. He and J. McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*, pages 507–517, 2016. 3.3

S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514, 2021. 1, 3

W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, and P. Wang. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2901–2908, 2020. 2, 2, 4

A. Newell, J. C. Shaw, and H. A. Simon. Report on a general problem-solving program. In *IFIP Congress*, 1959. 1

P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang. Squad: 100,000+ questions for machine comprehension of text, 2016. URL https://arxiv.org/abs/1606.05250. 3.2

M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks, 2017. URL https://arxiv.org/abs/1703.06103. 1

T. Shen, Y. Mao, P. He, G. Long, A. Trischler, and W. Chen. Exploiting structured knowledge in text via graph-guided representation learning. *arXiv preprint arXiv:2004.14224*, 2020. 2, 4, 3

T. Sun, Y. Shao, X. Qiu, Q. Guo, Y. Hu, X. Huang, and Z. Zhang. Colake: Contextualized language and knowledge embedding. *arXiv preprint arXiv:2010.00309*, 2020. 2, 2, 4, 2, 5

Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, and H. Wu. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*, 2019. 2, 2, 4

Y. Sun, S. Wang, S. Feng, S. Ding, C. Pang, J. Shang, J. Liu, X. Chen, Y. Zhao, Y. Lu, et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*, 2021. 2, 2, 4, 5

A. Talmor, J. Herzig, N. Lourie, and J. Berant. Commonsenseqa: A question answering challenge targeting commonsense knowledge, 2018. URL https://arxiv.org/abs/1811.00937. 3.2, 3

A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018a. 3.1

A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman. Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*, 32, 2019a. 3.1, 2

H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 417–426, 2018b. 2, 4, 4

H. Wang, F. Zhang, X. Xie, and M. Guo. Dkn: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*, pages 1835–1844, 2018c. 2, 2, 4

H. Wang, F. Zhang, M. Zhao, W. Li, X. Xie, and M. Guo. Multi-task feature learning for knowledge graph enhanced recommendation. In *The world wide web conference*, pages 2000–2010, 2019b. 4, 4

Q. Wang, P. Huang, H. Wang, S. Dai, W. Jiang, J. Liu, Y. Lyu, Y. Zhu, and H. Wu. Coke: Contextualized knowledge graph embedding, 2019c. URL https://arxiv.org/abs/1911.02168. 1

X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019d. 2, 2, 4

X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5329–5336, 2019e. 2, 2, 4

X. Wang, T. Gao, Z. Zhu, Z. Zhang, Z. Liu, J. Li, and J. Tang. Kepler: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9:176–194, 2021a. 2, 2, 4

X. Wang, T. Huang, D. Wang, Y. Yuan, Z. Liu, X. He, and T.-S. Chua. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the Web Conference 2021*, pages 878–887, 2021b. 2, 4

Y. Xian, Z. Fu, S. Muthukrishnan, G. De Melo, and Y. Zhang. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 285–294, 2019. 2, 2, 4

Y. Xu, C. Zhu, R. Xu, Y. Liu, M. Zeng, and X. Huang. Fusing context into knowledge graph for commonsense question answering. In *ACL-IJCNLP 2021*, December 2020. URL https://www.microsoft.com/en-us/research/publication/fusing-context-into-knowledge-graph-for-commonsense-question-answering/. 2, 2, 4

Y. Xu, C. Zhu, S. Wang, S. Sun, H. Cheng, X. Liu, J. Gao, P. He, M. Zeng, and X. Huang. Human parity on commonsenseqa: Augmenting self-attention with external attention. In *IJCAI 2022*, July 2022. URL https://www.microsoft.com/en-us/research/publication/human-parity-on-commonsenseqa-augmenting-self-attention-with-external-attention/. human parity result on CommonsenseQA. 2, 2, 4, 5

I. Yamada, A. Asai, H. Shindo, H. Takeda, and Y. Matsumoto. Luke: Deep contextualized entity representations with entity-aware self-attention, 2020. URL https://arxiv.org/abs/2010.01057. 2, 2, 4, 5

Z. Yang, P. Qi, S. Zhang, Y. Bengio, W. W. Cohen, R. Salakhutdinov, and C. D. Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering, 2018. URL https://arxiv.org/abs/1809.09600. 3.2

M. Yasunaga, J. Leskovec, and P. Liang. Linkbert: Pretraining language models with document links, 2022. URL https://arxiv.org/abs/2203.15827. 2, 2, 4, 5

F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 353–362, 2016. 2, 2, 4

Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu. Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*, 2019. 2, 2, 4