KB4Rec: A Data Set for Linking Knowledge Bases with Recommender Systems Postprint

Authors: Zhao, Xin, He, Gaole, Yang, Kunlin, Dou, Hongjian, Huang, Jin, Ouyang, Siqi, Wen, Ji-Rong, Zhao, Xin

Date: 2022-11-27T00:00:00+00:00

Abstract

To develop a knowledge-aware recommender system, a key issue is how to obtain rich and structured knowledge base (KB) information for recommender system (RS) items. Existing data sets or methods either use side information from original RSs (containing very few kinds of useful information) or utilize a private KB. In this paper, we present KB4Rec v1.0, a data set linking KB information for RSs. It has linked three widely used RS data sets with two popular KBs, namely Freebase and YAGO. Based on our linked data set, we first perform qualitative analysis experiments, and then we discuss the effect of two important factors (i.e., popularity and recency) on whether a RS item can be linked to a KB entity. Finally, we compare several knowledge-aware recommendation algorithms on our linked data set.

Full Text

Preamble

KB4Rec: A Data Set for Linking Knowledge Bases with Recommender Systems

Wayne Xin Zhao[1]†, Gaole He[1], Kunlin Yang[1], Hongjian Dou[1], Jin Huang[1], Siqi Ouyang[2] & Ji-Rong Wen[1]
[1]School of Information, Renmin University of China, Beijing 100872, China
[2]Jacobs Technion-Cornell Institute, Cornell Tech, New York 10044, USA

Keywords: Knowledge-aware recommendation; Recommender system; Knowledge base

Citation: W.X. Zhao, G. He, K. Yang, H. Dou, J. Huang, S. Ouyang & J.-R. Wen. KB4Rec: A data set for linking knowledge bases with recommender systems. Data Intelligence 1(2019), 121-136. doi: 10.1162/dint_a_{00008}

Received: November 10, 2018; **Revised:** November 28, 2018; **Accepted:** December 2, 2018

Abstract

To develop a knowledge-aware recommender system, a key challenge is obtaining rich and structured knowledge base (KB) information for recommender system (RS) items. Existing datasets and methods either rely on side information from original RSs (which contains very limited useful information) or utilize private KBs. In this paper, we present KB4Rec v1.0, a dataset that links KB information for RSs. Our dataset connects three widely used RS datasets with two popular KBs, namely Freebase and YAGO. Using this linked dataset, we first perform qualitative analysis experiments, then discuss the effect of two important factors (popularity and recency) on whether an RS item can be linked to a KB entity. Finally, we compare several knowledge-aware recommendation algorithms on our linked dataset.

1. Introduction

Recommender systems (RS), which aim to match users with items of interest, have become increasingly important in various online applications. Traditional recommendation algorithms primarily focus on learning effective preference models from historical user-item interaction data, such as matrix factorization [1]. With the rapid development of Web technologies, various kinds of side information have become available in RSs. In early stages, the contextual information used was typically unstructured, and its availability was limited to specific data domains or platforms.

Recently, both research and industry communities have made growing efforts to structure world knowledge or domain facts across various data domains. One of the most typical organizational forms is the knowledge base (KB) [3]. KBs provide a general and unified way to organize and associate information entities, which has proven useful in many applications, including recommender systems —referred to as knowledge-aware recommender systems [4]. To develop such systems, a key challenge is obtaining rich and structured KB information for RS items. Existing studies offer two main solutions. First, some collect side information from the RS platform to use as contextual features [5, 6, 7, 8, 9], with some studies further constructing small, simple KB-like structures [10, 11, 12]. However, the number of attributes or relations is usually small, and much useful item information is likely missing. Second, several works propose linking RSs with private KBs [13, 14, 15], but the linkage results are not publicly available. We are also aware of related studies [16, 17] that link RS items with DBpedia entities. In comparison, our focus is on Freebase [18] and YAGO [19], which are widely used in natural language processing (NLP) and related domains [20, 21, 22].

To address the need for linked RS-KB datasets, we present a dataset that connects two public KBs with recommender systems, named KB4Rec v1.0, freely available at https://github.com/RUCDM/KB4Rec. Our basic idea is to heuristically link items from RSs with entities from large-scale public KBs.

On the RS side, we selected three widely used datasets (MovieLens [5], LFM-1b [6], and Amazon book [7]) covering three domains: movie, music, and book. On the KB side, we selected two well-known KBs: Freebase and YAGO. We maximized the applicability of our linked dataset by choosing very popular RS datasets and KBs. We do not share the original datasets, as they are maintained by their original researchers or publishers and are easily accessible online.

In KB4Rec v1.0, we organized the linkage results as linked ID pairs, consisting of an RS item ID and a KB entity ID. All IDs are internal values from the original datasets. Once such linkage is established, existing large-scale KB data can be reused for RSs. For example, the movie "Avatar" from the MovieLens dataset [5] has a corresponding entity entry in Freebase, allowing us to obtain its attribute information by retrieving all associated relation triples in Freebase. Based on the linked dataset, we first perform qualitative analysis experiments, then discuss the effect of two important factors (popularity and recency) on whether an RS item can be linked to a KB entity. Finally, we compare several knowledge-aware recommendation algorithms on our linked dataset.

With our linkage results and original data copies, researchers can easily develop evaluation sets for knowledge-aware recommendation algorithms. We believe this dataset will benefit the development of knowledge-aware recommender systems.

2. Existing Data Sets and Methods

Early knowledge-aware recommendation algorithms were also called context-aware recommendation algorithms, where side information from the original RS platform was considered context data. For example, social network information from the Epinions dataset was utilized in [23, 24], POI property information from the Yelp dataset in [11], movie attribute information from the MovieLens dataset in [10], and user profile information from microblogging datasets in [25, 8]. These datasets typically contain very few kinds of side information, ignoring relations between different information types.

To obtain more structured side information, Heterogeneous Information Networks (HIN) have been proposed as a technique for modeling complex connections between different object types [26]. In HINs, we can effectively learn underlying relation patterns (called meta-paths) and organize side information via meta-path-based representations. HIN-based recommendation systems have been applied in PER [10], HeteRecom [27], and MCRec [28]. However, HIN-based algorithms usually rely on graph search algorithms, which struggle with large-scale relation pattern discovery.

More recently, KBs have become popular data resources for storing and organizing world knowledge or domain facts. Many studies have addressed KB construction, inference, and applications [3]. Several pioneering studies [13, 14, 15] have leveraged existing KB information to improve recommendation performance, applying heuristic methods to link RS items with KB entities. However, these studies use private KBs for linkage, which are not publicly accessible. We are also aware of related studies [16, 17] that link RS items with DBpedia entities. Nevertheless, our focus is on Freebase and YAGO, which are widely used in NLP and related domains [20, 21, 22]. Additionally, our datasets contain more linked entities and involved relations.

3. Linked Data Set Construction

Our work requires preparing two types of datasets: RS and KB. We first describe the original RS and KB datasets, then discuss the linkage method.

3.1 RS Data Sets

We consider three popular RS datasets for linkage: MovieLens, LFM-1b, and Amazon book, from the movie, music, and book domains, respectively.

The MovieLens dataset [7] describes user preferences on movies. A preference record takes the form <user, item, rating, timestamp>, indicating a user's rating score for a movie at a specific time. Four MovieLens datasets have been released: 100K, 1M, 10M, and 20M, reflecting the approximate number of ratings in each. We selected the largest, MovieLens 20M, for linkage.

The LFM-1b dataset [8] describes user interaction records on music, providing information including artists, albums, tracks, and users, as well as individual listening events. It records listening events of users on songs but does not contain rating information.

The Amazon book dataset [9] describes user preferences on book products, with data in the form <user, item, rating, timestamp>. The dataset is very sparse, containing 22 million ratings from 8 million users across nearly 23 million items.

All three datasets provide several kinds of side information, such as item titles (all datasets), IMDB ID (movie), writer (book), and artist (music). We utilize this side information for subsequent KB linkage.

3.2 KB Data Sets

We adopt two large-scale public KBs: Freebase and YAGO.

Freebase [18] is a knowledge graph announced by Metaweb Technologies, Inc. in 2007 and acquired by Google Inc. on July 16, 2010. Freebase stores facts as triples of the form <head, relation, tail>. Since Freebase shut down its services on August 31, 2016, we use its latest public version.

YAGO [19] is a large semantic KB automatically constructed based on information from Wikipedia, WordNet, GeoNames, and other data sources. It contains 447 million facts about 9.8 million entities in 10 different languages, with above 95% accuracy based on manual evaluation. In this paper, we use the YAGO version from [29].

3.3 RS to KB Linkage

With two KB datasets and three RS datasets, we can form six linkage results. Next, we describe the heuristic method for data linkage.

All three RS datasets provide item title information. For Freebase, we use offline KB search APIs to retrieve KB entities using item titles as queries. Our heuristic linkage method follows a similar approach to [30]. If no KB entity with the exact same title is returned, the RS item is rejected in the linkage process. If at least one KB entity with the exact same title is returned, we incorporate one kind of side information as a refined constraint for accurate linkage: IMDB ID, artist name, and writer name are used for the movie, music, and book domains, respectively. We found only a small number (about 1,000 for each domain) of RS items that cannot be accurately linked or rejected via this procedure, and we simply discard them.

For YAGO, a KB entity is named similarly to its corresponding Wikipedia URL, composed of the item title and related information such as type. For example, the film "Titanic" is marked as "<Titanic_(1997_{film})>" in YAGO, and the corresponding Wikipedia page can be accessed through https://en.wikipedia.org/wiki/Titanic_(1997_{film}). Therefore, we first compare the title of RS items with the prefix of KB entities. If at least one KB entity is returned, we leverage the "rdf:type" relation and suffix (if available) to filter out entities from other domains. We find that most linkages in LFM-1b and Amazon book datasets can be determined accurately (either linked or non-linked) this way. However, some ambiguous cases exist in the MovieLens 20M dataset, which are further evaluated through year restriction.

During the linkage process, we addressed several problems affecting string match algorithms, such as lowercase, abbreviation, and family/given name order. Since the LFM-1b dataset is extremely large, we removed all music items with fewer than 10 listening events. Even after filtering, it still contains about 6.5 million music items.

We present an illustrative example of our linkage results in Figure 1 [Figure 1: see original paper]. This example shows two pairs of items from MovieLens 20M and their linked entities from Freebase: the movies "Spider man" and "Spider man 2." Both movies share many common attributes in Freebase. With such linkage results, rich KB information about RS items can be easily obtained, which is likely useful for recommendation performance.

3.4 Basic Statistics

We summarize the basic statistics of the three linked datasets in Table 1 . For the MovieLens 20M dataset, we observe a very high linkage ratio: about 95.2% and 79.5% of items can be accurately linked to an entity from Freebase or YAGO, respectively. For the other two domains, the linkage ratios are much lower, especially when using YAGO. The high linkage ratio for MovieLens 20M is likely because it contains fewer items than the other two datasets, which themselves are refined by original releasers. Additionally, we speculate there may be domain bias in KB construction. Overall, more RS items can be linked with Freebase entities than YAGO. Although linkage ratios for the latter two datasets are not high, the absolute numbers of linked items are large. We also report the number of overlapping linked entities for the two KBs in the last row of Table 1, showing more linked items in the movie domain. Such a linked dataset is feasible for research-purpose studies.

3.5 Shared Data Sets

We name the above linked KB dataset for recommender systems KB4Rec v1.0, freely available at https://github.com/RUCDM/KB4Rec. In KB4Rec v1.0, we organize the linkage results as linked ID pairs, consisting of an RS item ID and a KB entity ID. All IDs are internal values from the original datasets. For Freebase, we have 25,982, 1,254,923, and 109,671 linked ID pairs for MovieLens 20M, LFM-1b, and Amazon book, respectively. For YAGO, we have 21,688, 49,608, and 17,607 linked ID pairs for MovieLens 20M, LFM-1b, and Amazon book, respectively.

4. Linkage Analysis

We previously showed the linkage ratios for different datasets and found that a considerable number of RS items cannot be linked to KB entities. It is worthwhile to study what factors affect the linkage ratio. We consider two factors for analysis.

4.1 Effect of Popularity on Linkage

Intuitively, a popular RS item should be more likely to be included in a KB than an unpopular item, as it is reasonable to incorporate more "important" RS items rated by users into KBs. KB construction usually involves manual efforts, making it difficult to avoid human attention bias. To measure RS item popularity, we adopt a simple frequency-based method by counting the number of users who have interacted with the item. This measure characterizes item attractiveness from RS users. We first sort items ascendingly by popularity value, then equally divide all items into five ordered bins with the same number of items in each bin. Thus, an item with a larger bin number is more popular than one with a smaller bin number. We compute the linkage ratio for each bin, with results reported in Figure 2 [Figure 2: see original paper]. We observe that bins with larger numbers have higher linkage ratios than those with smaller numbers, indicating that popularity likely has a positive effect on linkage.

4.2 Effect of Recency on Linkage

The second factor we consider is recency: the time when an RS item was created. Our assumption is that items created or released earlier are more likely to be included in KBs, since human attention aggregation is a gradually growing process and RS items usually require considerable time to become popular. To verify this assumption, we need item release dates. However, only the MovieLens 20M dataset contains this attribute information, so we report analysis results only for this dataset. We first sort items by release date ascendingly, then equally divide all items into 10 ordered bins following the popularity analysis procedure. Finally, we compute linkage ratios for each bin, with results reported in Figure 3 [Figure 3: see original paper]. Linkage ratios gradually decrease over time, indicating that recency likely has a negative effect on linkage—older RS items appear more likely to be included in KBs than more recent ones. In Figure 3(a), the last bin shows a dramatic drop because our MovieLens version is from April 2015.

This analysis indicates that both popularity and recency considerably affect linkage results. However, KB construction is a complex process influenced by many factors. Future research should investigate what other important factors affect this process and how they influence KB construction.

5. Experiment

We present comparisons of existing recommendation algorithms using our linked datasets.

5.1 Experimental Setup

Our purpose is to test whether incorporated KB information improves recommendation performance. Freebase contains more linked entities and associated relations, so we adopt only the Freebase linked dataset for evaluation. YAGO results are similar and omitted here.

The original linked dataset is very large, so we first generate a small evaluation set for experiments. We take the subset from the last year for LFM-1b and from 2005 to 2015 for MovieLens 20M. We perform 3-core filtering for Amazon book and 10-core filtering for other datasets, following the preprocessing steps in [31]. We then keep only items linked by our dataset. Dataset statistics are reported in Table 2 .

Following [32], we consider the last-item recommendation task for evaluation. This task is commonly used in RS evaluation and facilitates method comparison. Given a user, we sort items by interaction timestamp ascendingly, take the last item as the test set, and use the rest for training. The goal is to predict the last item given the user' s previous interaction sequence. Since enumerating all items as candidates is time-consuming, we pair each ground-truth item with 100 negative items to form a randomly ordered list. Each method returns a ranked list based on recommendation confidence. We evaluate methods using Mean Reciprocal Rank (MRR), Hit Ratio (HR), and Normalized Discounted Cumulative Gain (NDCG).

5.2 KB Information Representation

Our focus is providing rich KB information for recommender systems. A simple approach represents KB information with a one-hot vector, which is sparse and large. Here we borrow ideas from [15, 33] to embed KB data into low-dimensional vectors for subsequent recommendation algorithms. To train TransE [33], we start with linked entities as seeds and expand the graph with one-step search. Since not all KB relations are useful, we remove infrequent and general-purpose relations along with their associated triples. After this process, each linked item is associated with a learned KB embedding vector. Training statistics for TransE are reported in Table 3 .

5.3 Methods to Compare

We consider the following methods for performance comparison:

BPR [34]: Learns a matrix factorization model by minimizing pairwise ranking loss in a Bayesian framework.

SVDFeature [35]: A model for feature-based collaborative filtering. We use KB embeddings as context features fed into SVDFeature.

mCKE [13]: Originally proposed to incorporate KB and other information to improve recommendation performance. For fairness, we implement a simplified CKE version using only KB information, excluding image and text information. Unlike original CKE, we fix KB representations and adopt embeddings learned by TransE.

KSR [31]: A Knowledge-enhanced Sequential Recommender that incorporates KB information to enhance semantic representation in memory networks.

5.4 Results and Analysis

Results for different methods on the last-item recommendation task are presented in Table 4 . We observe that: (1) BPR performs worst on the first two datasets but very well on Amazon book, possibly because the first two datasets are relatively dense while Amazon book is sparse—lightweight methods may outperform complex methods on sparse datasets. (2) SVDFeature implements a pairwise ranking loss function and can be roughly understood as an enhanced BPR model incorporating learned KB embeddings. Compared to BPR, SVDFeature is slightly better on MovieLens 20M, substantially better on LFM-1b, but worse on Amazon book. Since each context feature in SVDFeature incorporates parameters (depending on dimension number), it may not outperform the simple BPR model on sparse datasets. (3) For knowledge-aware methods (mCKE and KSR), mCKE does not perform as expected, working well only on LFM-1b. A possible reason is that our mCKE implementation fixes learned KB embeddings, while original CKE adaptively updates them. In contrast, the recently proposed KSR method consistently performs best across all three datasets. KSR combines RNN' s capacity for modeling sequential behavior with Memory Networks' capacity for long-term storage, further enhanced with learned KB embeddings.

6. Conclusion

This paper presents KB4Rec v1.0, a dataset linking KB information for recommender systems. It connects three widely used RS datasets with popular KBs Freebase [18] and YAGO [19]. Using our linked dataset, we perform qualitative analysis experiments and discuss the effect of two important factors (popularity and recency) on RS item linkage to KB entities. Finally, we compare several knowledge-aware recommendation algorithms on our linked dataset.

For future work, we will consider linking more RS datasets with KBs and testing more knowledge-aware recommendation algorithms on additional recommendation tasks using the linked dataset.

Author Contributions

W.X. Zhao (batmanfly@gmail.com) and J.-R. Wen (jrwen@ruc.edu.cn) led the overall work. W.X. Zhao organized the content and wrote the paper. G. He (hegaole@ruc.edu.cn), H. Dou (hongjiandou@ruc.edu.cn), and J. Huang (jin.huang@ruc.edu.cn) generated Freebase linkage results and ran experiments. K. Yang (kunliny@ruc.edu.cn) generated YAGO linkage results. All authors made meaningful and valuable contributions to revising and proofreading the manuscript.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (No. 61872369, No. 61832017, and No. 61502502).

References

[1] Y. Koren, R. Bell, & C. Volinsky. Matrix factorization techniques for recommender systems. Computer 42(8) (2009), 30–37. doi: 10.1109/MC.2009.263.

[2] S. Rendle. Factorization machines with libFM. ACM Transactions on Intelligent Systems and Technology 3(3) (2012), Article No. 57. doi: 10.1145/2168752.2168771.

[3] Q. Wang, Z. Mao, B.Wang, & L. Guo. Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering 29(12)(2017), 2724–2743. doi: 10.1109/TKDE.2017.2754499.

[4] S. Bouraga, I. Jureta, S. Faulkner, & C. Herssens. Knowledge-based recommendation systems: A survey. International Journal of Intelligent Information Technologies 10(2)(2014), 1-19. doi: 10.4018/ijiit.2014040101.

[5] F. M. Harper, & J. A. Konstan. The MovieLens data sets: History and context. ACM Transactions on Interactive Intelligent Systems 5(4)(2016), Article No. 19. doi: 10.1145/2827872.

[6] M. Schedl. The lfm-1b data set for music retrieval and recommendation. In: Proceedings of the 2016 ACM International Conference on Multimedia Retrieval, 2016, pp. 103–110. doi: 10.1145/2911996.2912004.

[7] R. He, & 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, 2016, pp. 507–517. doi: 10.1145/2872427.2883037.

[8] W.X. Zhao, S. Li, Y. He, E.Y. Chang, J.-R. Wen, & X. Li. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. IEEE Transactions on Knowledge and Data Engineering 28(5)(2016), 1147–1159. doi: 10.1109/TKDE.2015.2508816.

[9] J. Huang, Z. Ren, W.X. Zhao, G. He, J. Wen, & D. Dong. Taxonomy-aware multi-hop reasoning networks for sequential recommendation. In: Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM' 19), 2019, pp. 573–581. doi: 10.1145/3289600.3290972.

[10] X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, & J. Han. Personalized entity recommendation: A heterogeneous information network approach. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, 2014, pp. 283–292. doi: 10.1145/2556195.2556259.

[11] H. Gao, J. Tang, X. Hu, & H. Liu. Content-aware point of interest recommendation on location-based social networks. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2015, pp. 1721–1727. Available at: https://dl.acm.org/citation.cfm?id=2886559.

[12] C. Shi, B. Hu, W.X. Zhao, & P.S. Yu. Heterogeneous information network embedding for recommendation. IEEE Transactions on Knowledge and Data Engineering 31(2)(2019), 357-370. doi: 10.1109/TKDE.2018.

[13] F. Zhang, N.J. Yuan, D. Lian, X. Xie, & W. Ma. Collaborative knowledge base embedding for recommender systems. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 353–362. doi: 10.1145/2939672.2939673.

[14] H. Wang, F. Zhang, X. Xie, & M. Guo. DKN: Deep knowledge-aware network for news recommendation. In: The 2018 World Wide Web Conference, 2018, pp. 1835–1844. doi: 10.1145/3178876.3186175.

[15] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, & M. Guo. Ripple network: Propagating user preferences on the knowledge graph for recommender systems. arXiv preprint. arXiv: 1803.03467v4.

[16] T.D. Noia, V.C. Ostuni, P. Tomeo, & E.D. Sciascio. Sprank: Semantic path-based ranking for top-n recommendations using linked open data. ACM Transactions on Intelligent Systems and Technology 8(1)(2016), Article No. 9. doi: 10.1145/2899005.

[17] T.D. Noia, & V.C. Ostuni. Recommender systems and linked open data. In: W. Faber, & A. Paschke (eds.) Reasoning Web 2015: Reasoning Web. Web Logic Rules. Cham, Switzerland: Springer International Publishing, 2015, pp. 88–113. doi: 10.1007/978-3-319-21768-0_4.

[18] Google. Freebase data dumps. Available at: https://developers.google.com/freebase/data.

[19] F.M. Suchanek, G. Kasneci, & G. Weikum. YAGO: A core of semantic knowledge. In: The 16th International Conference on the World Wide Web, 2007, pp. 697–706. doi: 10.1145/1242572.1242667.

[20] H. Bast, & E. Haussmann. More accurate question answering on Freebase. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015, pp. 1431–1440. doi: 10.1145/2806416.2806472.

[21] W. Cui, Y. Xiao, & W. Wang. KBQA: An online template based question answering system over Freebase. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, 2016, pp. 4240–4241. Available at: https://dl.acm.org/citation.cfm?id=3061256.

[22] P. Adolphs, M. Theobald, U. Schafer, H. Uszkoreit, & G. Weikum. YAGO-QA: Answering questions by structured knowledge queries. In: Proceedings of the 2011 IEEE International Conference on Semantic Computing, 2011, pp. 158–161. doi: 10.1109/ICSC.2011.30.

[23] M. Jamali, & M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the 4th ACM Conference on Recommender Systems, 2010, pp. 135–142. doi: 10.1145/1864708.1864736.

[24] H. Ma, I. King, & M.R. Lyu. Learning to recommend with social trust ensemble. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2009, pp. 203–210. doi: 10.1145/1571941.1571978.

[25] W.X. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, & X. Li. We know what you want to buy: A demographic-based system for product recommendation

on microblogs. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 1935–1944. doi: 10.1145/2623330.2623351.

[26] Y. Sun, & J. Han. Mining heterogeneous information networks: A structural analysis approach. ACM SIGKDD Explorations Newsletter 13(2)(2012), 20–28. doi: 10.1145/2481244.2481248.

[27] C. Shi, C. Zhou, X. Kong, P.S. Yu, G. Liu, & B. Wang. Heterecom: A semantic-based recommendation system in heterogeneous networks. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 1552–1555. doi: 10.1145/2339530.2339778.

[28] B. Hu, C. Shi, W.X. Zhao, & P.S. Yu. Leveraging meta-path based context for top-N recommendation with a neural co-attention model. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 1531–1540. doi: 10.1145/3219819.3219965.

[29] J. Hoffart, F.M. Suchanek, K. Berberich, & G. Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence 194 (2013), 28–61. doi: 10.1016/j.artint.2012.06.001.

[30] T. Scheffler, R. Schirru, & P. Lehmann. Matching points of interest from different social networking sites. In: Proceedings of the 35th Annual German Conference on Advances in Artificial Intelligence, 2012, pp. 245–248. doi: 10.1007/978-3-642-33347-7_{24}.

[31] J. Huang, W.X. Zhao, H. Dou, J. Wen, & E.Y. Chang. Improving sequential recommendation with knowledge-enhanced memory networks. In: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR' 18), 2018, pp. 505–514.

[32] R. He, W. Kang, & J. McAuley. Translation-based recommendation: A scalable method for modeling sequential behavior. In: Proceedings of the 11th ACM Conference on Recommender Systems, 2017, pp. 161–169. doi: 10.1145/3109859.3109882.

[33] A. Bordes, N. Usunier, A. García-Durán, J. Weston, & O. Yakhnenko. Translating embeddings for modeling multi-relational data. In: The 27th Conference on Neural Information Processing Systems (NIPS 2013), 2013, pp. 2787–2795. Available at: http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf.

[34] S. Rendle, C. Freudenthaler, Z. Gantner, & L. Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 2009, pp. 452–461. Available at: https://dl.acm.org/citation.cfm?id=1795114.1795167.

[35] T. Chen, W. Zhang, Q. Lu, K. Chen, Z. Zheng, & Y. Yu. Svdfeature: A toolkit for feature-based collaborative filtering. The Journal of Machine Learning Research 13(1)(2012), 3619–3622.

Author Biography

Wayne Xin Zhao is currently an Associate Professor at the School of Information, Renmin University of China. He received his PhD degree from Peking University in 2014. His research interests include recommender systems and natural language processing. He has published more than 50 referred papers in international conferences and journals.

Gaole He is currently a graduate student at the School of Information, Renmin University of China. He received his Bachelor' s degree from the School of Information, Renmin University of China in 2018. His research focuses on knowledge graphs, deep learning, and network embedding.

Kunlin Yang is currently an undergraduate student at the School of Information, Renmin University of China. He is expected to receive his Bachelor' s degree in Engineering in 2019. His research interests include recommender systems and knowledge graphs.

Hongjian Dou is currently a graduate student at the School of Information, Renmin University of China. He works at the Key Laboratory of Big Data Management and Analysis Methods, Beijing. His research focuses on natural language processing, deep learning, and recommender systems.

Jin Huang is currently a graduate student at the School of Information, Renmin University of China. She works at the Key Laboratory of Big Data Management and Analysis Methods, Beijing. Her research interests include recommender systems, deep learning, and knowledge bases.

Siqi Ouyang is currently a graduate student at the Jacobs Technion-Cornell Institute, Cornell Tech. She received her Bachelor's degree from Renmin University of China in 2018. Her research focuses on deep learning in natural language processing and recommendation systems.

Ji-Rong Wen is a Professor at the School of Information, Renmin University of China. He is also Director of the Key Laboratory of Big Data Management and Analysis Methods, Beijing. Previously, he was a Senior Researcher and Group Manager of the Web Search and Mining Group at Microsoft Research Asia (MSRA). His main research interests include big data management and analytics, information retrieval, data mining, and machine learning. He is currently Associate Editor of ACM Transactions on Information Systems (TOIS) and a Senior Member of IEEE.

Note: Figure translations are in progress. See original paper for figures.