University of Reading

Augmenting visual information in knowledge graphs for recommendations

Conference or Workshop Item

Accepted Version

Markchom, T. and Liang, H. (2021) Augmenting visual information in knowledge graphs for recommendations. In: ACM International Conference on Intelligent User Interfaces, 13-17 April 2021, Texas. doi: https://doi.org/10.1145/3397481.3450686 Available at https://centaur.reading.ac.uk/97210/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1145/3397481.3450686

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

University of Reading

Reading's research outputs online

Augmenting Visual Information in Knowledge Graphs for Recommendations

THANET MARKCHOM, Department of Computer Science, University of Reading, United Kingdom
HUIZHI LIANG, Department of Computer Science, University of Reading, United Kingdom

Knowledge graphs (KGs) have been popularly used in recommender systems to leverage high-order connections between users and items. Typically, KGs are constructed based on semantic information derived from metadata. However, item images are also highly useful, especially for those domains where visual factors are influential such as fashion items. In this paper, we propose an approach to augment visual information extracted by popularly used image feature extraction methods into KGs. Specifically, we introduce visually-augmented KGs where the extracted information is integrated by using visual factor entities and visual relations. Moreover, to leverage the augmented KGs, a user representation learning approach is proposed to learn hybrid user profiles that combine both semantic and visual preferences. The proposed approaches have been applied in top-N recommendation tasks on two real-world datasets. The results show that the augmented KGs and the representation learning approach can improve the recommendation performance. They also show that the augmented KGs are applicable in the state-of-the-art KG-based recommender system as well.

CCS Concepts: • **Information systems** → **Recommender systems**; *Collaborative filtering*; *Clustering and classification.*

Additional Key Words and Phrases: knowledge graph, heterogeneous information network, image feature, user profiling

ACM Reference Format:
Thanet Markchom and Huizhi Liang. 2021. Augmenting Visual Information in Knowledge Graphs for Recommendations. 1, 1 (April 2021), 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recommender systems have become essential tools to tackle information overload issues. They are designed to suggest items or information that potentially preferable or relative to users' personal interests [23]. One widely used approach for building recommender systems is collaborative filtering (CF). In CF-based recommender systems, user-item interactions are used to profile users' interests to make recommendations [7, 10]. In general, CF is highly effective [7, 10]. However, the performance becomes poorer when user-item interactions are sparse. Leveraging side information to enrich or augment user-item interactions becomes one popular research direction to address this problem [23].

Knowledge graphs (KGs) have been ubiquitous in recommendation research due to its capability of providing contextual information that can overcome the sparsity problem [8, 19–22]. A KG is a directed graph whose nodes are entities, and edges denote relations. In a recommendation scenario, such nodes usually represent users, items and semantic factors such as item attributes while edges denote user-item interaction (e.g., purchased) and item-semantic factor relations (e.g., belong to this category). Besides user-item interactions and item metadata, images are one kind of

Authors' addresses: Thanet Markchom, Department of Computer Science, University of Reading, PO Box 217, Reading, United Kingdom, RG6 6AH, t.markchom@pgr.reading.ac.uk; Huizhi Liang, Department of Computer Science, University of Reading, PO Box 217, Reading, United Kingdom, RG6 6AH, huizhi.liang@reading.ac.uk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
Manuscript submitted to ACM

important and popularly available side information. They are intuitively capable of providing rich information about users' preferences. For example, in clothing or fashion recommender systems, some users may prefer buying items with similar/complementary visual appearances rather than those that have the same category with their purchased items. Due to advances in computer vision and image processing, it is possible to extract these features from images using both feature extraction methods [1, 14, 17] and deep learning models such as convolutional neural networks (CNNs) [9]. However, most KG-based recommender systems mainly focus on semantic information and ignore visual information. How to incorporate visual information to improve recommendations still remains an open research question.

In this work, we introduce visual factor entities and visual relations to KGs and explore how to construct visually-augmented KGs with various kinds of popular image feature extraction methods. These KGs are supposed to be useful for learning latent representations of users and items to profile users' preferences in both semantic and visual perspectives. However, the majority of existing approaches [4, 18] were proposed to consider only semantic information. This work also proposes an approach to learn user latent representations from a hybrid context contains both semantic and visual information. A novel type of meta-paths called visually-annotated meta-paths are introduced. They are used to form a hybrid context to learn the representations for being used in recommender systems. Both approaches of constructing visually-augmented KGs and learning user representations from them are the main contributions of this work.

2 RELATED WORK

Several approaches in recommendation have been developed during the past few decades. One of the popular approaches is CF using user-item interactions, either explicit (e.g. ratings) or implicit feedback (e.g. buy, tag and watch) to recommend items. This approach can be applied with various methods to learn user/item latent factors such as the k-nearest neighbors (KNN), Matrix Factorization (MF) [10] and deep learning models [23]. Recently, KGs have been leveraged in recommender systems instead of only using user-item interactions as in CF approaches. These KG-based methods can be divided into two approaches: path-based approach [8, 21] considering paths within KGs to predict recommendations and embedding-based approach [19, 22] using low-dimensional node embeddings for the prediction. In an embedding-based approach, how to find node embeddings is normally optional. Different methods including translation-based methods [2, 13] and random-walk-based methods [5, 16] can be chosen. Meta-paths-based approaches such as Metapath2vec [4] are one of the random-walk-based methods. They have been proposed to learn latent representations for nodes based on the semantics of the networks or KGs. Nevertheless, the existing methods were proposed to form a context restricted to a given meta-path and fail to consider hybrid neighborhood context where multiple factors with different weighting probabilities need to be jointly considered. Besides these approaches, there have been also unified methods that combine both approaches such as Knowledge Graph Attention Network (KGAT) [20]. It explicitly models the high-order interdependence in an end-to-end fashion by attentively propagating the embeddings from the entity's neighbors regardless of the entity types to refine the embeddings. However, this model has only been applied to typical KGs constructed from semantic factors and has ignored the visual information.

3 AUGMENTING VISUAL INFORMATION IN A KNOWLEDGE GRAPH

A knowledge graph (KG) is defined as a directed graph $\mathcal{G} = \{\mathcal{E}, \mathcal{R}\}$ whose nodes in $\mathcal{E}$ are entities and edges in $\mathcal{R}$ are relations. Each entity can belong to a type $N_i \in \mathbb{N}$. A relation type connecting entities with the type N_i and N_j is denoted by $R_{N_iN_j}$ and its inverse is denoted by $R^{-1}_{N_iN_j}$. In a recommendation scenario, common entity types are user (U), item (P) and other semantic factors such as category (C) or tag (T) and genre (G) for movies. To incorporate visual information in a typical KG, we introduce a set of *visual factor entities* ($\mathcal{V}$) with a type V and a set of *visual relations*

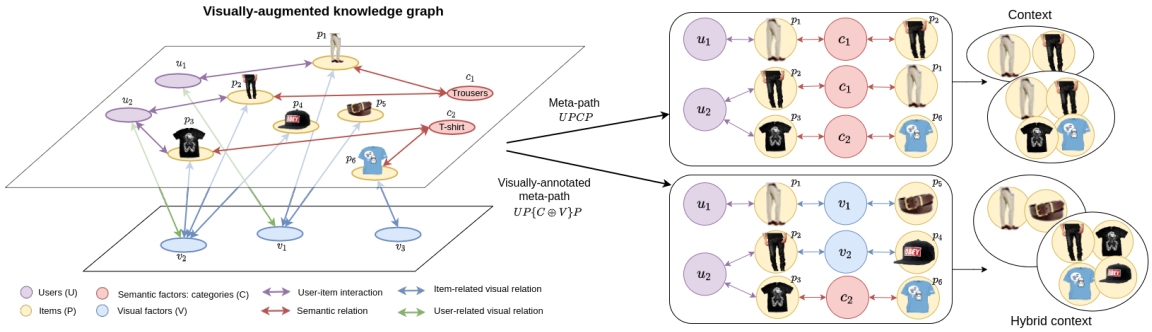

Fig. 1. An example of a visually-augmented KG and possible relationships based on a meta-path and a visually-annotated meta-path

($\mathcal{W}$). With these additional entities and relations, a *visually-augmented knowledge graph* is defined as $\mathcal{G}' = \{\mathcal{E}', \mathcal{R}'\}$ where $\mathcal{E}' = \mathcal{E} \cup \mathcal{V}$, $\mathcal{R}' = \mathcal{R} \cup \mathcal{W}$.

Let p_k be an item node $p_k \in \mathcal{P}, \mathcal{P} \subset \mathcal{E}$ and i_k denote the image of p_k. To generate a set of visual factors, image features are first extracted from every i_k. Five image feature types in both local and global levels are considered. For local features, well-known features designed to capture significant shapes and textures are considered, i.e., SIFT [14], SURF [1] and ORB [17]. For global features, two popular feature types are extracted. The first one is Color Histogram (CH) indicating a distribution of hue and saturation values in HSV color space. A 2-D histogram of hue and saturation is firstly computed and then reshaped to a vector for being used as a global color histogram feature. The second global feature is a latent vector extracted from the pre-trained CNN model. As in [7], the feature is extracted from the second fully-connected layer (i.e. FC7) of Caffe reference model [9]. This feature is referred to as CNN feature. After image features are extracted, we applied the popularly used clustering method, k-means, on image features. In this way, an image i_k and its corresponding item node p_k can be allocated into one cluster in case of global features or more clusters in case of local features. Each cluster center is treated as a visual factor and added as one visual factor entity $v \in \mathcal{V}$ in $\mathcal{G}'$. Meanwhile, the relations between items and their clusters are considered as relations between item entities and the corresponding visual factor entities, i.e, R_{PV} and R_{PV}^{-1}. They are added to a set of visual relations $\mathcal{W}$ in $\mathcal{G}'$.

The relations between user and visual factor entities, i.e., R_{UV} and R_{UV}^{-1} are also considered in this work. They can be obtained by profiling users' visual preferences and identifying which visual factors are related to each user. To do so, we first aggregate all the items and their images of a given user u_j. We adopt mean pooling [12] to get the summary statistics of the visual factors of the items of u_j. Let $\mathcal{I}_j$ denote the image set of u_j. For each image $i \in \mathcal{I}_j$, we get all the image feature cluster centers $\mathbf{c}$ (represented as a vector). Then we averaged all the image feature cluster centers of $\mathcal{I}_j$ and get a vector to represent the user's visual preferences denoted as $\mathbf{u}_j$. In this paper, we compared the similarity of $\mathbf{u}_j$ with all existing image feature cluster centers and selected the most similar k^* clusters as the representative visual factors of user u_j. The relations between users and the representative visual factor entities are then added to $\mathcal{W}$ in $\mathcal{G}'$. Figure 1 shows an example of a visually-augmented KG. It has three types of semantic nodes, i.e., users (U), items (P) and categories (C), and two pairs of semantic relations R_{UP}, R_{UP}^{-1}, R_{PC} and R_{PC}^{-1}. It has one type of visual factor entities V and two pairs of visual relations R_{UV},R_{UV}^{-1},R_{PV},R_{PV}^{-1}. These nodes and relations can reveal more connections between users which cannot be achieved based on semantic factors, for example, the connections between user u_1 and item p_5.

4 RECOMMENDER SYSTEM BASED ON VISUALLY-AUGMENTED KNOWLEDGE GRAPH

Based on the generated visually-augmented KG, we can apply knowledge graph-based models such as KGAT [20] to make recommendations directly. However, learning latent representations or embedding of nodes has recently been proved to be effective for recommender systems and other applications [19, 22]. Meta-path-based random walk approaches [18] have been popularly used to generate embeddings of heterogeneous information networks, a network or a graph with multiple entity/relation types. However, these approaches only consider semantic factors to form a neighborhood context.

To form a hybrid neighborhood context that considers both semantic and visual factors, we define *visually-annotated meta-path* $m = \left(N_i\{\delta * N_x \oplus (1-\delta) * N_y\}N_l...N_j\right)$ as a sequence of node types of $\mathcal{G}'$, where $\oplus$ is a symbol that represent the "or" relation of semantic node type N_x and visual factor type N_y, δ is a probability $0 \leq \delta \leq 1$. It contains at least one visually annotated node type and one visual relation type. Starting from node type N_i, the next node type will go to semantic type N_x with probability of δ and go to visual factor type N_y with probability $1-\delta$. For simplicity, we ignore the probability in the annotation and use $\left(N_i\{N_x \oplus N_y\}N_l...N_j\right)$ to denote m.

This paper assumes each connection (i.e., edge) between any two nodes with the same type of relation is equally important. Starting from one source node, we walk along a visually-annotated meta-path to get a set of neighbor nodes. Different from other papers only consider symmetric meta-paths, in this paper, we do not restrict the meta-paths to be symmetric, as the connectivity of different types of nodes is important in recommender systems [8]. For example, in Figure 1, $UP\{C \oplus V\}P$ is a visually-annotated meta-path. It forms a hybrid neighborhood where those items that either have the same visual factor value or have the same category are considered similar. For item p_1, following this meta-path, item p_5 that has the same visual factor value as p_1 has the probability $1-\delta$ of being put in the same context with p_1, while item p_2 that has the same category as p_1 has the probability δ of being put in the same context with p_1.

Based on the generated hybrid neighborhood context, we use self-supervised representation learning models such as skip-gram to learn item node representations/embeddings [15]. Similar to work [12], each user representation is computed by using the mean of item node embeddings of this user's items. The learned user and item representations can be used for different kinds of recommendation approaches. In this paper, the user-based CF method is adopted. The recommendation process is the same as the traditional user-based CF method except the input is the generated user representations rather than the typical user-item interaction matrix.

5 EXPERIMENTS

5.1 Experimental Setup

The experiments were conducted on two datasets: 1) **Amazon dataset**[1] [6], consisting of users' reviews and item metadata in "Clothing, Shoes and Jewelry" category. We only retained 5-rated reviews in the dataset to ensure the users' satisfaction for learning their preferences and considered user rating as implicit feedback. 2) **MovieLens dataset**[2] [3], the version of HetRec2011-MovieLens-2K. The dataset includes user tagging data, movie genres, movie tags, and movie poster images. To extract visual factors, the movie posters were crawled from OMDB[3]. To avoid the sparsity problem, 10-core data in which users and items have at least ten reviews or tagging records each were selected.

The number of visual factors is 100 (i.e., $k = 100$ in k-means clustering algorithm). The number of representative visual factors per user is 1 ($k^* = 1$). This parameter was chosen after comparing with $k* = 3, 5$ and 10. They had similar

[1] http://jmcauley.ucsd.edu/data/amazon/
[2] https://grouplens.org/datasets/hetrec-2011/, http://www.imdb.com, http://www.rottentomatoes.com
[3] http://www.omdbapi.com

Table 1. The statistics of the visually-augmented KGs

	Amazon dataset				MovieLens dataset			
Entity	users (U)	12,491	items (P)	1,019	users (U)	152	items (P)	301
	categories (C)	604	visual factors (V)	100	genres (G) / tags (T)	18 / 3,031	visual factors (V)	100
Relation	R_{UP}	207,281	R_{PC}	5,775	R_{UP}	3,870	R_{PG}/R_{PT}	871 / 11,289
	R_{UV}	12,491	R_{PV}(SIFT)	11,116	R_{UV}	152	R_{PV}(SIFT)	3,009
	R_{PV}(SURF)	1,979	R_{PV}(ORB)	45,536	R_{PV}(SURF)	3,009	R_{PV}(ORB)	3,007
	R_{PV}(CH)	964	R_{PV}(CNN)	964	R_{PV}(CH)	301	R_{PV}(CNN)	301

accuracy but the selected setting required less computational resources. The basic statistics of these KGs are shown in Table 1. To ensure that the hybrid contexts can cover sufficient information, the number of generated paths for every starting node was set to 20. We selected some popular semantic meta-paths in literature [11] to conduct experiments. The selected meta-paths are: $m_1 = UP$, $m_2 = UVP$, $m_3 = U\{P \oplus V\}UP\{U \oplus V\}P$ and $m_4 = UP\{C \oplus V\}P\{U \oplus V\}P$ (or $UP\{G \oplus V\}P\{U \oplus V\}P$) and $m_5 = UP\{T \oplus V\}P\{T \oplus V\}P$. These meta-paths were applied in the experiments on the datasets where they are applicable. The first and the second probabilities in the visually-annotated meta-paths ($m_3 - m_5$) were varied among $\{0, 0.25, 0.5, 0.75, 1\}$ to optimize the results. As for node embedding, the skip-gram method was applied with the size of embeddings set to 128 while the other settings were set as in [4]. For the user-based CF models, the number of user neighbors was set to 10 for all experiments.

5.2 Recommendation results

The top-N recommendation performance was evaluated by two commonly used metrics, the *average precision@N* (AP) and *average recall@N* (AR) with $N = 1, 5, 10, 50, 100$. They were computed as follows:

$$AP = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{|\mathcal{P}_u \cap \mathcal{P}_u^N|}{N} \quad \text{and} \quad AR = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{|\mathcal{P}_u \cap \mathcal{P}_u^N|}{|\mathcal{P}_u|},$$

where $\mathcal{U}$ is a set of users, $\mathcal{P}_u$ is the set of user u's items and $\mathcal{P}_u^N$ is the set of top-N recommended items of a user u. We compared the proposed approach **VR** with **MR** which is a user-based CF approach using *metapath2vec++* [4] method based on typical semantic meta-paths. The results of **MR** and **VR** using different meta-paths are shown in Figure 2. We can see that the proposed approach performed better than **MR** for all these meta-paths for both datasets. It suggests that the proposed approach that considers hybrid context containing both semantic (e.g., tags or categories) and visual factors is more effective than **MR** that only considers semantic factors.

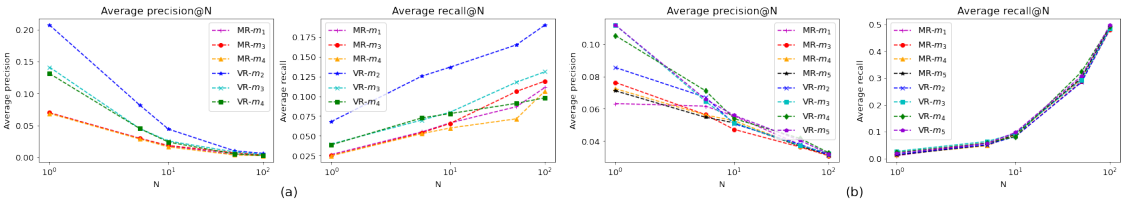

Fig. 2. Results of **VR** and **MR** approaches with different meta-paths on (a) **Amazon dataset** and (b) **MovieLens dataset**

We also compared the effectiveness of different visual factors. To discuss the effects of visual factor types, **VR** approaches with different settings are compared in Figure 3 where **VR**-m_i-**S**, **VR**-m_i-**F**, **VR**-m_i-**O**, **VR**-m_i-**H** and **VR**-m_i-**C** denote the **VR** approaches using SIFT, SURF, ORB, CH and CNN visual factors with the meta-path m_i respectively.

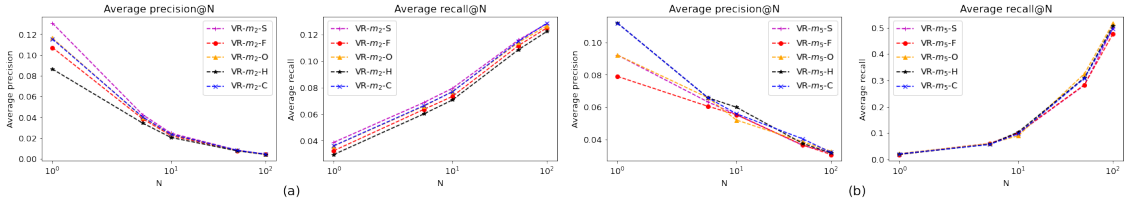

Fig. 3. Results of the proposed approach **VR** with different visual factor types on (a) **Amazon dataset** and (b) **MovieLens dataset**

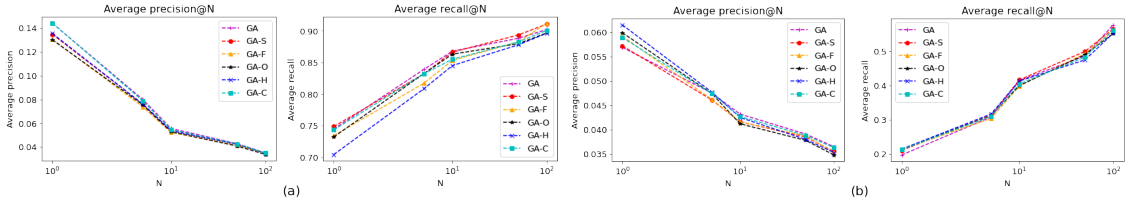

Fig. 4. Results of **GA** based on KGs augmented with different visual factor types on (a) **Amazon dataset** and (b) **MovieLens dataset**

The best meta-paths of both datasets are compared with the best probability parameter setting. For **Amazon dataset**, **VR-m_2-S** performed better than other approaches for both AP and AR. It suggests that texture and shape features are more effective in recognizing users' preferences compared to the color feature of items in this dataset. For **MovieLens dataset**, **VR-m_5-C** and **VR-m_5-H** outperformed other settings in terms of AP. It demonstrates that CNN and CH features are more effective in capturing users' visual preferences on the movie posters compared to the other features.

Furthermore, we also evaluated the effectiveness of visually-augmented KGs applied directly in a KG-based recommender system. The state-of-the-art system based on the KGAT [20] was adopted denoted by **GA**. Let **GA-S**, **GA-F**, **GA-O**, **GA-H** and **GA-C** denote the **GA** approaches that applied on the augmented KGs based on SIFT, SURF, ORB, CH and CNN visual factors respectively. The results of **GA** based on the original KGs and the augmented KGs with different visual factors are shown in Figure 4. We can see that **GA-C** performed as well as **GA** for **Amazon dataset**. For **MovieLens dataset**, **GA-H** performed slightly better than **GA** in terms of AP. It suggests that the augmentation may not work well in **GA** approach if there are already a number of entities and relations in the original KGs. On the other hand, augmenting entities and relations in the **MovieLens dataset** KGs is more effective since it increases the chance to discover more relationships beyond those in the original KGs.

6 CONCLUSION

This work proposes a visually-augmented KG in which image features are incorporated as visual factor entities. We investigated five different types of image feature extraction methods to augment visual factors in KGs. Moreover, we also propose a user representation learning approach to form a hybrid context that considers both semantic and visual factors from the augmented KGs. The generated user representations are applied to neighborhood-based recommender systems. We conducted experiments on two real-world datasets **Amazon** and **MovieLens**. We compared the proposed approaches with baseline models based on metapath2vec++ [4] and conducted an experiment on the recent KGAT [20] to evaluate visually-augmented KGs effectiveness. The results show that visual factors can be used to improve recommendation accuracy and the proposed approaches are effective. In the future, we would like to explore more

image feature types to construct visually-augmented KGs and investigate hybrid contexts with more factors such as temporal factors. Also, we will explore other self-learning models for more effective embedding learning and user profiling based on visually-augmented KGs.

REFERENCES

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. SURF: Speeded Up Robust Features. In *Computer Vision – ECCV 2006*, Aleš Leonardis, Horst Bischof, and Axel Pinz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 404–417.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada) *(NIPS'13)*. Curran Associates Inc., Red Hook, NY, USA, 2787–2795.

[3] Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2011. 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). In *Proceedings of the 5th ACM conference on Recommender systems* (Chicago, IL, USA) *(RecSys 2011)*. ACM, New York, NY, USA.

[4] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD '17*. ACM, 135–144.

[5] Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. New York, NY, USA, 855–864.

[6] Ruining He and Julian McAuley. 2016. 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* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 507–517. https://doi.org/10.1145/2872427.2883037

[7] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona) *(AAAI'16)*. AAAI Press, 144–150.

[8] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S. Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1 (2018), 1531–1540.

[9] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv:1408.5093 [cs.CV]

[10] Y. Koren, R. Bell, and C. Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37.

[11] H. Liang. 2020. DRprofiling: Deep Reinforcement User Profiling for Recommendations in Heterogenous Information Networks. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1.

[12] Huizhi Liang and Timothy Baldwin. 2015. A Probabilistic Rating Auto-encoder for Personalized Recommender Systems. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*. ACM, 1863–1866. https://doi.org/10.1145/2806416.2806633

[13] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (Austin, Texas) *(AAAI'15)*. AAAI Press, 2181–2187.

[14] David Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2. 1150–1157 vol.2.

[15] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality.

[16] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. *CoRR* abs/1403.6652 (2014).

[17] Ethan Rublee, V. Rabaud, K. Konolige, and G. Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision* (2011), 2564–2571.

[18] Yizhou Sun, Jiawei Han, X. Yan, Philip S. Yu, and Tianyi Wu. 2011. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. *Proc. VLDB Endow.* 4 (2011), 992–1003.

[19] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1835–1844. https://doi.org/10.1145/3178876.3186175

[20] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *KDD*.

[21] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable Reasoning over Knowledge Graphs for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019), 5329–5336. https://doi.org/10.1609/aaai.v33i01.33015329 arXiv:1811.04540

[22] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative Knowledge Base Embedding for Recommender Systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 353–362. https://doi.org/10.1145/2939672.2939673

[23] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Comput. Surv.* 52, 1, Article 5 (Feb. 2019), 38 pages. https://doi.org/10.1145/3285029