

Bayesian Knowledge-driven Critiquing with Indirect Evidence

Armin Toroghi
University of Toronto
Toronto, ON, Canada
armin.toroghi@mail.utoronto.ca

Zhenwei Tang
University of Toronto
Toronto, ON, Canada
josephtang@cs.toronto.edu

Griffin Floto
University of Toronto
Toronto, ON, Canada
griffin.floto@mail.utoronto.ca

Scott Sanner*
University of Toronto
Toronto, ON, Canada
ssanner@mie.utoronto.ca

ABSTRACT

Conversational recommender systems (CRS) enhance the expressivity and personalization of recommendations through multiple turns of user-system interaction. Critiquing is a well-known paradigm for CRS that allows users to iteratively refine recommendations by providing feedback about attributes of recommended items. While existing critiquing methodologies utilize direct attributes of items to address user requests such as ‘I prefer Western movies’, the opportunity of incorporating richer contextual and side information about items stored in Knowledge Graphs (KG) into the critiquing paradigm has been overlooked. Employing this substantial knowledge together with a well-established reasoning methodology paves the way for critique-based recommenders to allow for complex knowledge-based feedback (e.g., ‘I like movies featuring war side effects on veterans’) which may arise in natural user-system conversations. In this work, we aim to increase the flexibility of critique-based recommendation by integrating KGs and propose a novel Bayesian inference framework that enables reasoning with relational knowledge-based feedback. We study and formulate the framework considering a Gaussian likelihood and evaluate it on two well-known recommendation datasets with KGs. Our evaluations demonstrate the effectiveness of our framework in leveraging *indirect* KG-based feedback (i.e., preferred relational properties of items rather than preferred items themselves), often improving personalized recommendations over a one-shot recommender by more than 15%. This work enables a new paradigm for using rich knowledge content and reasoning over indirect evidence as a mechanism for critiquing interactions with CRS.

CCS CONCEPTS

• Information systems → Recommender systems; Personalization.

*Affiliate to Vector Institute of Artificial Intelligence, Toronto

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 the author(s) 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.

SIGIR '23, July 23–27, 2023, Taipei, Taiwan

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9408-6/23/07...\$15.00

<https://doi.org/10.1145/3539618.3591954>

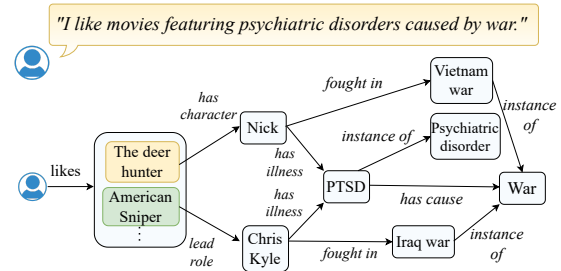


Figure 1: An example of knowledge-based feedback containing indirect evidence of user’s interest and using the side information of knowledge graph to address it

KEYWORDS

Conversational Recommendation, Critiquing, Knowledge Graph-enhanced Recommendation

ACM Reference Format:

Armin Toroghi, Griffin Floto, Zhenwei Tang, and Scott Sanner. 2023. Bayesian Knowledge-driven Critiquing with Indirect Evidence. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*, July 23–27, 2023, Taipei, Taiwan. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3539618.3591954>

1 INTRODUCTION

The growth of interest in recommender systems has given rise to demands and expectations for more expressive user-system interactions and as a response, Conversational Recommender Systems (CRS) have been introduced to facilitate a deeper understanding of a user’s preferences through a sequence of user-system interactions [7, 9, 10, 15]. Critiquing is an evolving CRS paradigm that enables personalized recommendations by adapting to user’s feedback in an iterative process [3]. In each critiquing session, the system recommends a set of items and in response, the user either accepts them or provides feedback about her preference toward a subset of recommended items or their attributes. While previous studies have presented techniques to incorporate evidence of a user’s interests in other items [6, 13], items from other domains [12, 14], and item properties [11], the ability to personalize recommendations using knowledge-based feedback containing indirect evidence of the user’s interests has not been addressed. An example of this type of feedback and the associated relational knowledge needed to address it are shown in Figure 1. This type of feedback is referred to

as *indirect evidence* since the provided information does not target the user’s preferred items but a distribution of items that share those relational properties to different extents.

Addressing this type of feedback requires abundant side information about the users and/or items and their relational dependencies, and Knowledge Graphs (KG) as a rich source of side information have been widely used to assist in recommendation [5]. Despite the considerable attention devoted to KG-enhanced recommendation in recent years [16–18], most works in this realm have focused on enhancing recommendations in a one-shot paradigm, and the question of how KG side information could be utilized to address user’s feedback in a conversational setting has been understudied. Aligning KG-enhanced recommendation with a CRS requires a proper methodology for reasoning over the user’s interests based on the KG information, which this work aims to address.

In this work, we make the following contributions: (i) We introduce the Gaussian variant of a popular tensor factorization approach for KG-enhanced recommendation. (ii) We propose Bayesian Critiquing with Indirect Evidence (BCIE), a knowledge-driven critiquing framework, and formulate a Bayesian closed-form user belief updating methodology to enable critiquing CRSs to address indirect feedback. (iii) We empirically show that BCIE results in considerable improvement of personalized recommendation over one-shot recommendation by evaluation on two datasets and a KG.

2 BAYESIAN CRITIQUING WITH INDIRECT EVIDENCE

The overall workflow of the BCIE framework is shown in Figure 2. Here, we first define the problem of Bayesian knowledge-driven critiquing with indirect evidence, and next, illustrate how BCIE operates in the pre-critiquing and critiquing phases.

Problem Definition In the conversational critiquing problem setting that we investigate, denoting the set of available items by I , preference records of a user $u \in U$ are extracted from the user preference matrix $R \in \mathbb{R}^{|U| \times |I|}$ and transformed to a set of KG triples $H = \{(u, \textit{likes}, i) | i \in I\}$, that along with KG triples of item side information form a user-item KG as shown in Figure 2(a). By performing link prediction on the user-item KG, a set of items R_u that seem pertinent to the user u are recommended. Link prediction is performed using a scoring function $\Phi(\langle h, r, t \rangle)$ that estimates the plausibility of the existence of the relation r from h to t . Along with the recommended items, the set of facts $F = \{(h, r, t) | t \in R_u \vee h \in R_u\}$ related to the recommended items is presented to the user, and the user may either accept the recommendation or critique a fact $f \in F$ (e.g., the critique presented in Figure 2(c)). The recommender’s duty in the next step is to update its belief in the user’s interests and refine R_u given d_n , the evidence of the user’s taste observed from the critique at iteration n . Hence, the recommender needs a critique-modified recommendation function f_m , such that $\hat{R}_u = f_m(R_u, d_n)$. This process continues either for a fixed number of iterations or until the user accepts the recommendation or leaves the conversation.

Pre-critiquing phase The initial stage of BCIE’s workflow involves a standard knowledge-based recommendation problem, where the objective is to leverage side information of a KG to improve recommendations. We build our recommender upon Simple, a well-known

tensor factorization-based KG embedding model, because of its efficient computations and *full-expressiveness* [8]. This model assigns two embedding vectors h_e and t_e to each entity e and two vectors v_r and v_{r-1} to each relation r , and defines its scoring function for a triple (e_i, r, e_j) as $\Phi(e_i, r, e_j) = \frac{1}{2}(\langle h_{e_i}, v_r, t_{e_j} \rangle + \langle h_{e_j}, v_{r-1}, t_{e_i} \rangle)$, in which $\langle v, w, x \rangle = (v \odot w) \cdot x$ where \odot is element-wise and \cdot represents dot product. The model parameters are trained by forming a batch of size N from a number of KG triples and negative samples and solving the following optimization problem:

$$\min_{\theta} \sum_{n=1}^N \log(1 + \exp(-y_n \cdot \Phi(e_i, r, e_j)_n)) + \lambda \|\theta\|_2^2 \quad (1)$$

in which $y \in \{+1, -1\}$ denotes the triple label, θ stands for the model parameters, n is the iterator of batch samples, and λ is the regularization hyperparameter. Using the learned embeddings of entities and relations, the set of items yielding the highest plausibility scores for $(user, \textit{likes}, item)$ triples are picked for recommendation.

Critiquing phase: Bayesian User Belief Updating with Indirect Evidence In each critiquing session, the user provides knowledge-based feedback containing *indirect* evidence of her preference. Notably, the type of critique studied in this work is not limited to being directly attributed to a particular item, rather, the user may provide indirect feedback about various relational attributes from different items that she likes. Hence, in the BCIE framework, we need to consider a distribution over representations of items that cover the user’s interest, which is denoted by z_m . To this end, we require to maintain a belief state over the user preferences, hereafter called *user belief*, which is initially centered at the learned embedding of the user entity and update it conditioned on the user critiques. By exponentiating Simple’s loss function in Equation 1 to form the maximum likelihood problem, it could be shown that the Simple transforms the link prediction problem to a tensor factorized view of logistic regression by maximizing the sigmoid likelihood. However, in BCIE, we introduce a new variant of Simple in which the sigmoid likelihood is replaced by a Gaussian, forming a conjugate pair with the Gaussian prior assumed over the user belief; this, in turn, enables tractable closed-form belief updates using Gaussian Belief Propagation (BP) [19].

To update the user belief conditioned on the observed indirect evidence, we first need to update the item distribution. Assuming the evidence distribution as $z_d \sim \mathcal{N}^{-1}(\mathbf{h}_d, \mathbf{J}_d)$ and prior over item distribution as $z_m \sim \mathcal{N}^{-1}(\mathbf{h}_m, \mathbf{J}_m)$, where \mathbf{h} and \mathbf{J} are the potential vector and precision matrix of the corresponding distribution in the information form [1], the posterior item distribution becomes:

$$p(z_m | z_d) \sim \mathcal{N}^{-1}(\mathbf{h}_d + \mathbf{h}_m, \mathbf{J}_d + \mathbf{J}_m) \quad (2)$$

At the next stage, we pass on this updated belief to the user distribution $z_u \sim \mathcal{N}^{-1}(\mathbf{h}_u, \mathbf{J}_u)$. First, we can write the joint distribution of z_u and z_m conditioned on the evidence as:

$$p(z_u, z_m | z_d) \propto \exp \left\{ -\frac{1}{2} z_u^T \mathbf{J}_u z_u + \mathbf{h}_u^T z_u - \frac{1}{2} z_m^T \mathbf{J}_z z_m + \mathbf{h}_z^T z_m - z_u^T \mathbf{J}_{u,z} z \right\} \quad (3)$$

The next challenge is obtaining $\mathbf{J}_{u,z}$. Note that by adopting the Gaussian variant of Simple, the likelihood factor between

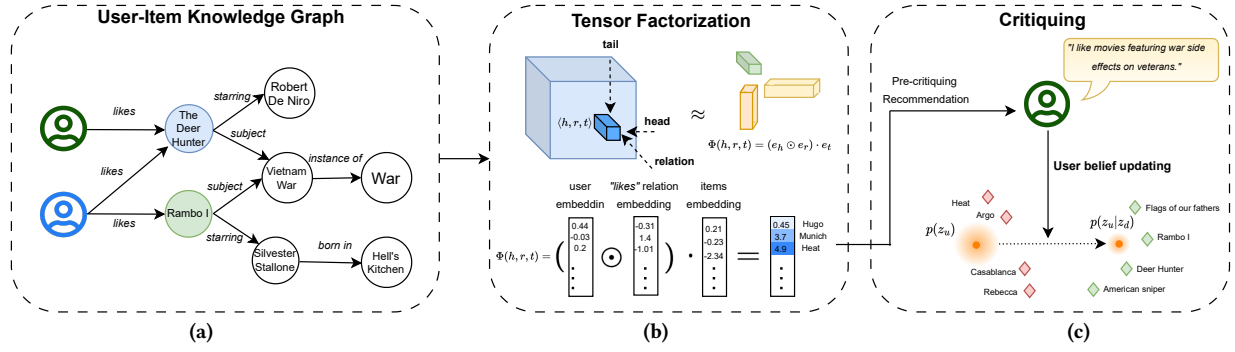


Figure 2: General workflow of the BCIE framework. (a) user interaction data is combined with KG side information to create a user-item KG (b) A tensor factorization model is trained to learn entity and relation embeddings, which are then utilized to generate initial recommendations. (c) Indirect evidence of user’s interest is observed from the critique and conditioned on the embeddings of the indirect evidence (z_d), posterior update is performed on the user belief ($p(z_u)$) to refine it toward representations of items satisfying the critique and make refined recommendations in the next session.

Table 1: Summary of Dataset Statistics

Name	#Users	#Items	#KG entities	#KG relations
MovieLens	61,715	16,826	56,789	46
AmazonBook	74,026	21,081	79,682	38

the user belief distribution z_u and item distribution z_m becomes $\exp\{-\langle z_u, r, z_m \rangle\}$ where r is the embedding vector of the *likes* relation – this is log-bilinear in z_u and z_m and would appear to stymie closed-form Gaussian belief propagation. Serendipitously, we can rewrite $\langle z_u, r, z_m \rangle$ as $u^T D_r z$, where D_r is derived by reshaping r as a diagonal matrix. Hence, we have $J_{u,z} = D_r$. Next, to derive $p(z_u|z_d)$, the posterior user belief distribution, we marginalize the joint probability of the user belief and item distribution over z_m :

$$p(z_u|z_d) = \int p(z_u, z_m|z_d) dz_m \propto \exp\left\{-\frac{1}{2} z_u^T J_u z_u + \mathbf{h}_u^T z_u\right\} \times \int \exp\left\{-\frac{1}{2} z_m^T J_z z_m + \mathbf{h}_z^T z_m - z_u^T J_{u,z} z_m\right\} dz_m$$

In order to calculate the latter integral, we first write the exponentiated term in matrix form to obtain:

$$p(z_u|z_d) \propto \exp\left\{-\frac{1}{2} z_u^T J_u z_u + \mathbf{h}_u^T z_u\right\} \times \int \exp\left\{-\frac{1}{2} \begin{pmatrix} z_u \\ z_m \end{pmatrix}^T \begin{pmatrix} \mathbf{0} & J_{u,z} \\ J_{u,z} & J_z \end{pmatrix} \begin{pmatrix} z_u \\ z_m \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{h}_z^T \end{pmatrix} \begin{pmatrix} z_u \\ z_m \end{pmatrix}\right\} dz_m$$

Now, observing that the term inside the integral is *exactly* in the form of a jointly Gaussian distribution, we can obtain its marginal probability distribution using the Schur complement. Thus, we have $p(z_u) \sim \mathcal{N}^{-1}(\hat{\mathbf{h}}_u, \hat{J}_u)$, where:

$$\hat{\mathbf{h}}_u = \mathbf{h}_u - J_{u,z}(J_z)^{-1}(\mathbf{h}_z) \quad (6)$$

$$\hat{J}_u = J_u - J_{u,z}(J_z)^{-1}J_{u,z} \quad (7)$$

To summarize, while the use of a tensor-based likelihood introduced an unusual log-bilinear form and the need to marginalize over the latent item distribution induced by the KG critiques, we have shown that we can manipulate all necessary quantities in Gaussian form. In this way, we can perform a closed-form Gaussian user belief update w.r.t. an item distribution inferred by indirect KG properties.

3 EXPERIMENTS AND EVALUATION

Datasets We evaluate BCIE¹ on two of the most popular recommendation datasets, MovieLens 20M² and Amazon-Book³, and acquire facts about their items from Freebase KG [2]. We consider ratings greater than 3.5 to indicate that the user likes an item and extract facts about items from Freebase using entity matching data from [20]. Also, since we conduct 5 steps of critiquing, we only keep items with at least 5 KG facts to enable selection of non-repetitive critiques. Table 1 shows dataset statistics.

Baselines As prior KG-enhanced recommendation studies have not considered the conversational setting and previous critiquing works do not handle knowledge-based indirect evidence, we propose two comparison baselines, namely *‘Mapped items’* and *‘Direct’*. *Mapped items* is a heuristic method that maps each critique to a maximum of 10 relevant items from the KG and uses them for user belief updating. For example, for the critique *‘I prefer movies like Nolan’s works’*, movies that are *directed by, Nolan* are mined from the KG and used as examples of the user’s interests. Notably, *Mapped items* is anticipated to be a strong baseline, as the ground truth for many critiques, especially those referring to less popular entities is likely among the few extracted items, and thus, revealed to the recommender. *Direct* is an ablation of BCIE that treats indirect evidence of the user’s interest as direct by omitting the item distribution and directly attributing the critique fact to the user belief.

RQ1: Pre-critiquing recommendation performance. Although the pre-critiquing performance is not our main focus, it can indicate the quality of the trained embeddings and provide a baseline

¹<https://github.com/atoroghi/BCIE>

²<https://grouplens.org/datasets/movielens/20m/>

³<http://jmcauley.ucsd.edu/data/amazon/>

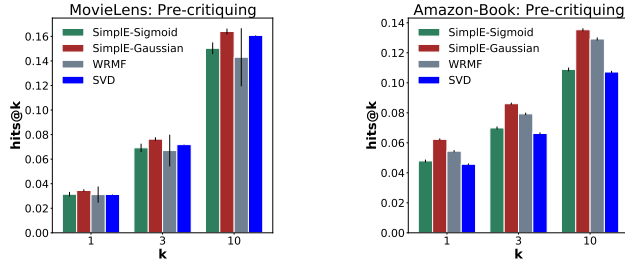


Figure 3: Initial performance of tensor factorization-based recommender compared to classical recommendation models

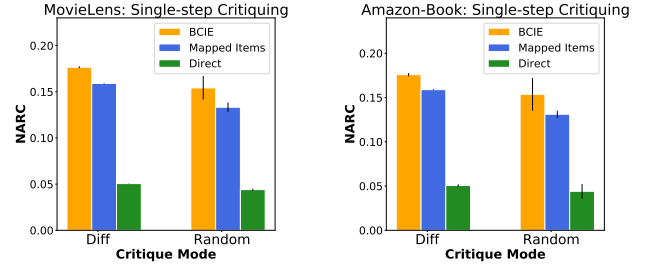


Figure 4: Single-step critiquing performance of BCIE compared to the baselines under two critique modes

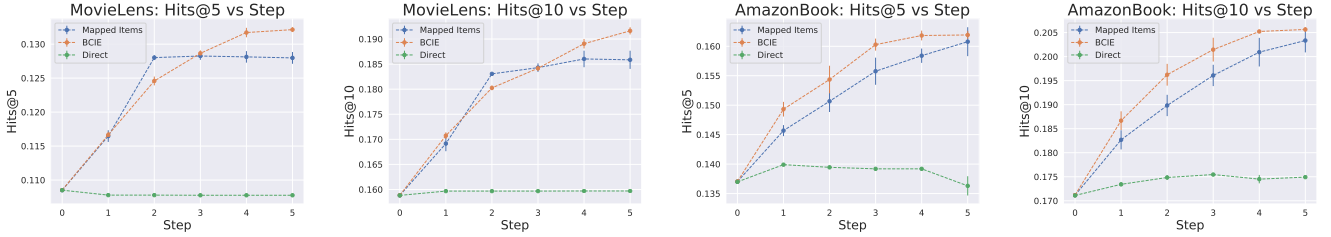


Figure 5: Comparison of change in hit rate @{5,10} during the critiquing sessions averaged over 3 independent runs.

for comparing the efficiency of critiquing in improving the recommendations. We use two classical recommendation models, namely WRMF [6] and SVD [4], as the baselines. For all experiments, hyperparameter tuning is performed using nested cross-validation with one inner loop validation fold and five test folds in the outer loop. Results of these experiments are provided in Figure 3 that show the proposed Gaussian variant of the tensor factorization-based recommender outperforms the classical models and the performance of the original SimpleE model is on par with that of these baselines.

RQ2: Effect of Single-step Critiquing. As a proof-of-concept, we investigate BCIE’s efficacy in single-step critiquing. Following a positive critique, we anticipate observing a reduction in the average rank of items satisfying the fact. To monitor this impact, we utilize the Normalized Average Rank Change (NARC) metric defined as

$$NARC = \frac{AR_{pre}(I^f) - AR_{post}(I^f)}{AR_{pre}(I^f)} \quad (8)$$

where $AR_{pre}(I^f)$ and $AR_{post}(I^f)$ denote the Average Rank of items fulfilling the critique fact f before and after updating the user belief, respectively. This experiment is conducted under two critique selection modes: *diff*, where the user critiques a fact related to ground truth that deviates the most from the recommended items facts, and *random*, where the critique fact is randomly selected. In all critiquing experiments, the only hyperparameters are the initial precision matrices which are optimized for all methods. The experimental results for both datasets are presented in Figure 4 demonstrating that BCIE is more proficient in capturing the user’s interests from their critiques and reflecting them in reducing the rank of the relevant items by more than 15 %.

RQ3: Multi-step Critiquing Performance. In this experiment, for each $(u, likes, i)$ triple in the test set, item i is once considered

as the ground truth target of recommendation for user u . At each critiquing session, the top 10 ranked items together with their KG critiques are presented to the user, and following *diff* critique selection mode, a critique is selected. Experimental results comparing BCIE with the baselines are presented in Figure 5. The hit rates at 5 and 10 reveal that BCIE often performs superior to the *Mapped Items* and *Direct* baseline, demonstrating the effectiveness of BCIE’s Bayesian update of the user belief in accurately capturing the user’s preferences and directing the top recommendations towards the items of interest. The initial hit rate increase of the *Mapped items* in the MovieLens dataset is observed to be caused by the fact that, in several cases, the ground truth is among the few items used for updating the user belief from the first critique. However, since further updating with the same evidence does not contribute new information, the performance plateaus. Also, the comparison with the direct baseline highlights the importance of incorporating item distribution in the framework. It is noteworthy that treating indirect evidence feedback (e.g., "I want movies related to Vietnam war"), as direct evidence of the user’s interest (e.g., "User likes Vietnam war") significantly alters the semantics.

4 CONCLUSION

We proposed BCIE, a novel Bayesian framework that enables incorporating rich contextual and KG-based relational side information into critique-based recommender systems to enhance their expressivity by allowing for knowledge-based indirect feedback. Experimental results exhibit the efficacy of BCIE in initial recommendation and its ability to capture user preferences by reasoning over the indirect evidence observed through sequential critiquing. This promising performance combined with the computational efficiency of its closed-form posterior user belief update opens up possibilities for versatile future extensions to address more complex and natural types of user critiquing feedback.

REFERENCES

- [1] Christopher M Bishop and Nasser M Nasrabadi. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.
- [2] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 1247–1250.
- [3] Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction* 22 (2012), 125–150.
- [4] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*. 39–46.
- [5] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3549–3568.
- [6] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE international conference on data mining*. Ieee, 263–272.
- [7] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2020. A Survey on Conversational Recommender Systems. arXiv:2004.00646 [cs.HC]
- [8] Seyed Mehran Kazemi and David Poole. 2018. Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems* 31 (2018).
- [9] Shuokai Li, Ruobing Xie, Yongchun Zhu, Fuzhen Zhuang, Zhenwei Tang, Wayne Xin Zhao, and Qing He. 2022. Self-Supervised learning for Conversational Recommendation. *Information Processing & Management* 59, 6 (2022), 103067.
- [10] Shuokai Li, Yongchun Zhu, Ruobing Xie, Zhenwei Tang, Zhao Zhang, Fuzhen Zhuang, Qing He, and Hui Xiong. 2023. Customized Conversational Recommender Systems. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part II*. Springer, 740–756.
- [11] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. *The adaptive web: methods and strategies of web personalization* (2007), 325–341.
- [12] Ahmad Nurzid Rosli, Tithrotanak You, Inay Ha, Kyung-Yong Chung, and Geun-Sik Jo. 2015. Alleviating the cold-start problem by incorporating movies facebook pages. *Cluster Computing* 18 (2015), 187–197.
- [13] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. 285–295.
- [14] Suvash Sedhain, Scott Sanner, Darius Brazianas, Lexing Xie, and Jordan Christensen. 2014. Social collaborative filtering for cold-start recommendations. In *Proceedings of the 8th ACM Conference on Recommender systems*. 345–348.
- [15] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *The 41st international acm sigir conference on research & development in information retrieval*. 235–244.
- [16] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 417–426.
- [17] 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*. 1835–1844.
- [18] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 950–958.
- [19] Yair Weiss and William Freeman. 1999. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Advances in neural information processing systems* 12 (1999).
- [20] Wayne Xin Zhao, Gaole He, Kunlin Yang, Hongjian Dou, Jin Huang, Siqu Ouyang, and Ji-Rong Wen. 2019. Kb4rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence* 1, 2 (2019), 121–136.