Information Processing and Management 60 (2023) 103402

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Content recommendation with two-level TransE predictors and interaction-aware embedding enhancement: An information seeking behavior perspective

Chen Yang, Ruozhen Zheng, Xuanru Chen, Hong Wang *

College of Management, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:
Content recommendation
Information seeking
Translating embedding
Engagement in online communities

ABSTRACT

In community-based social media, users consume content from multiple communities and provide feedback. The community-related data reflect user interests, but they are poorly used as additional information to enrich user-content interaction for content recommendation in existing studies. This paper employs an information seeking behavior perspective to describe user content consumption behavior in community-based social media, therefore revealing the relations between user, community and content. Based on that, the paper proposes a Community-aware Information Seeking based Content Recommender (abbreviated as CISCRec) to use the relations for better modeling user preferences on content and increase the reasoning on the recommendation results. CISCRec includes two key components: a two-level TransE prediction framework and interaction-aware embedding enhancement. The two-level TransE prediction framework hierarchically models users' preferences for content by considering community entities based on the TransE method. Interaction-aware embedding enhancement is designed based on the analysis of users' continued engagement in online communities, aiming to add expressiveness to embeddings in the prediction framework. To verify the effectiveness of the model, the real-world Reddit dataset (4,868 users, 115,491 contents, 850 communities, and 602,025 interactions) is chosen for evaluation. The results show that CISCRec outperforms 8 common baselines by 9.33%, 4.71%, 42.13%, and 14.36% on average under the Precision, Recall, MRR, and NDCG respectively.

1. Introduction

Social Recommender Systems (SRS) are recommenders applied to the social media domain to address information overload along with the rapid growth of uploaded data (Zheng et al., 2023). Among them, Content Recommendation is a key area of SRS. There are two prevalent methods for content recommendation: collaborative filtering (CF) and content-based (CB). CF approaches recommend content by measuring user or item similarity based on other individuals' feedback (Lin et al., 2023). Despite its prevalence, this approach has inherent problems, such as data sparsity (Duan et al., 2022) and rating imbalance (Cai et al., 2022). CB methods recommend content by analyzing content descriptions and users' past actions (Anandhan et al., 2018). However, this approach largely depends on the richness of context and it is unable to well infer user preferences using a single source of context. In that case, researchers have proposed utilizing auxiliary information, such as community-related information, to improve content recommendation.

* Corresponding author.
E-mail address: ms.hongwang@gmail.com (H. Wang).

https://doi.org/10.1016/j.ipm.2023.103402
Received 27 January 2023; Received in revised form 1 April 2023; Accepted 22 April 2023
Available online 6 May 2023

The community-related data contains a wealth of useful information. Online communities are defined as groups of individuals coming together around a specific topic, interest, or goal (Zhang et al., 2022). They usually have a unique name on different social media platforms, e.g., "lists" on Twitter, "subreddit" on Reddit, "groups" on Flickr and Facebook, and "boards" on Pinterest. In such community-based social media, users consume content from multiple communities and provide feedback and this data has been used to infer users' latent characteristics and interests (Gasparetti et al., 2021, Li et al., 2022, Zarrinkalam et al., 2018). Since community-related information reflects the unique interests of individual users, there have been studies utilizing this information to improve content recommendations. For example, Zheng et al. compute the interest linkage density of nodes and analyze the membership of a user in communities and that of a subject distributed in communities to provide personalized content recommendations (Zheng et al., 2019). Campos et al. use content-based and topic-labeling methods to recommend courses in multiple MOOC communities (Campos et al., 2022). These studies demonstrate the usage of community benefits content recommendation, but the existing using techniques toward community are insufficient since only explicit features are considered. We will further illustrate this via Fig. 1.

Fig. 1 (a) represents the role community plays in previous works. For users, the community is usually considered as a result of

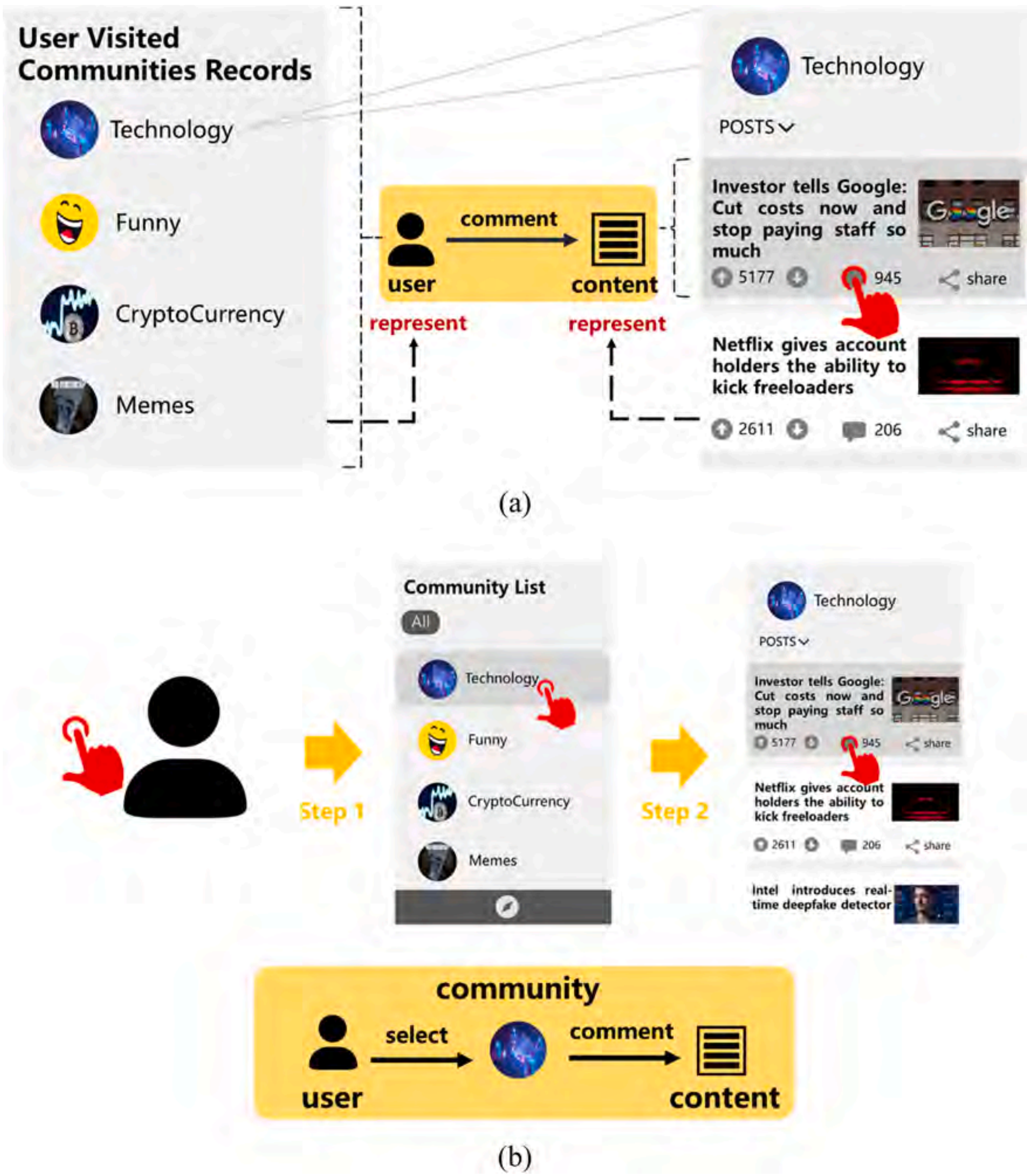


Fig. 1. Comparison between the concept of previous studies and our study.

interest-based clustering, used to model users' diverse interests. Similarly, for content, community-related information is also used to increase the expressiveness of content representation. The methods only care about the user-content interaction relation and employ the community information as useful additional information to better represent the user and content, therefore enriching the user-content interaction. In our study, the concept of "relation" originates from the Knowledge Graph domain, and it is the predicate that represents facts about two entities (Zhang et al., 2018). In this user-content interaction context, entities are the user and content, and the relation can be commenting, rating, or voting for specific content. However, besides the intuitive user-content interaction relation, there are implicit, hierarchical relations if introducing community entities. The community apparently conveys semantic conceptual information about content, reflecting users' perception of the content. If users participate in one community, it means that they are interested in what this community provides. From an Information-Seeking point of view, when users have no special information need or interest, they tend to select communities and then interact with possibly interested content among the options. The process is shown in Fig. 1 (b). Due to user-community interaction and content-community affiliation, the community is like a bridge that connects user and content and the user-content interaction can be further decomposed into the transaction from user to community and transaction from community to content. Such structure introduces communities to the original entity set (users and content), adding connections and relations between entities, but previous works have failed to utilize the hierarchical relations brought by the community entity.

Therefore, this study aims at providing a content recommender to fill the gap above by describing user content consumption behavior in community-based social media from an information seeking behavior perspective. A novel content recommendation model Community-aware Information Seeking based Content Recommender (CISCRec) is proposed in this paper. Compared with previous studies directly modeling the single relation between users and contents, CISCRec introduces the community to form a new content consumption behavior structure based on Information Seeking behavior related theories, therefore utilizing the hierarchical relations inside to better capture user preferences on contents and facilitating the reasoning on content recommendation. To model user preferences by the hierarchical content consumption behavior, a two-level TransE-based prediction framework is developed, where TransE is a canonical model, learning vector embeddings of the entities and the relations based on translation principles (Bordes et al., 2013). Specifically, on the first level, CISCRec aims to predict user-community interaction by three elements (user, community, and their relation, which is represented by the user's historically interacted contents) and estimate the probability user engages in each community in the next timestamp. The second level tends to predict user-content interaction between three elements (user, content, and their relation, which is represented by the content-community affiliation). Finally, results from the two levels are multiplied together to obtain the user preferences for the target content. Furthermore, to improve the effectiveness of the two-level TransE-based prediction framework, inspired by the user's continued engagement in online communities, we propose interaction-aware embedding techniques to enhance the user and community embeddings by investigating the usage of community interaction factors.

To sum up, the main contributions of this paper are listed as follows:

- We propose a novel content recommendation model CISCRec, which describes and leverages the hierarchical relations in the content consumption behavior structure by considering the community, to recommend content in community-based social media.
- We design a two-level TransE-based prediction framework to predict user-content interactions from a micro-level perspective. It on one hand hierarchically models users' content consumption behavior and on the other hand, explains why the content is recommended.
- We develop interaction-aware embedding enhancement techniques for entity embeddings in the prediction framework by integrating community interaction-related factors. It not only adds more expressiveness to user and community embeddings but also facilitates the reasoning of user engagement in communities.
- Experimental results on a real-world dataset from Reddit demonstrate that our proposed model significantly outperforms other commonly used methods on the top-k content recommendation.

The rest of the paper is organized as follows. Section 2 introduces some related works. Section 3 details the CISCRec model. Section 4 presents and discusses the experimental results. Section 5 concludes this paper.

2. Related work

In this section, we review four bodies of related work. In Section 2.1 "Content Recommendation", we sort out the relevant research on content recommendation and discuss the existing research gaps. We find that previous works have ignored the hierarchical relations between user, content, and community. Then, in Section 2.2 "Information Seeking Systems Development", we further illustrate the hierarchical relations in the user content consumption behavior from an information-seeking perspective and describe how recommender systems incorporate information-seeking tools, justifying the introduction of community in our study. Section 2.3 "Translating Embedding" introduces relevant methods of entity and relation modeling, which provides technical support for our proposed two-level TransE-based prediction framework. Section 2.4 "User Engagement in Online Communities" presents interaction-related factors affecting user engagement in Online Communities, which is the theoretical basis for our proposed embedding enhancement mechanism.

2.1. Information seeking systems development

Information Seeking Behavior is intentionally seeking for information to meet some goal. In the course of seeking, the individual may interact with information systems (Lee et al., 2022). Whiting and Williams conduct in-depth interviews with individuals ranging in age from 18 to 56 years old to explore why they use social media (Whiting and Williams, 2013). They find most respondents reported using social media to seek information or entertainment. In the domain of information behavior research, researchers examine the motivations for surfing the Internet for entertainment (Matni and Shah, 2014). They conclude that some information seeking activity on the Internet can be classified as entertainment since informational and entertaining elements are intertwined while surfing the Internet for entertainment. Therefore, common user content consumption in social media can be seen as a kind of information seeking behavior (Shah et al., 2023).

Users tend to browse when they don't have specific information needs. In the browsing mode, randomly exposed themselves to a large amount of content, individuals are passively available to absorb information and discover their potential information needs in the process (Bates, 2002). Obviously, it is inefficient. Hence, folksonomy (e.g., social tagging and ontology approach) has emerged and changed users' content consumption behavior (Liu et al., 2020, Poli, 1996). That is, with folksonomy, users can find a topic of interest to browse information, thus narrowing the scope of the content resources related to their interests and improving seeking efficiency. However, problems associated with the completely unsupervised nature of such user-dependent information organizations may reduce their effectiveness in content indexing and searching, thereby hindering the content consumption of users (Zhou et al., 2023). To address the problem, some researchers have applied recommendation techniques to folksonomies to help users find relevant content. For example, some tag-based recommendation methods are proposed, demonstrating the value of tags in describing content concisely (Chen et al., 2020, Yadav et al., 2021, Agrawal et al., 2021).

Close to tags in terms of information organization, communities assist users in browsing content along aggregate thematic indexes, apparently conveying the semantic conceptual information on resources collaboratively generated by users. Therefore, we believe that the recommender system will benefit from the introduction of community and our recommender engine tends to consider the inner hierarchical relations between users, content and communities in the content consumption behavior based on information seeking behaviors, aiming to represent user preferences accurately. The detail will be described in part 3.

2.2. Content recommendation

Social Recommender Systems (SRS) are recommender systems that target the social media domain and social media content recommendation is one of the key areas of SRS (Ricci et al., 2015). Social media enables users to create and share different types of content. Exposed to a huge volume of content, users are having difficulty judging the validity of so much content and choosing which content to consume (Tarus et al., 2018). Therefore, content recommendation aims at coping with the content overload issue by presenting the most relevant and attractive data to the user according to their preferences.

There are two basic approaches commonly used in content recommendation, namely content-based (CB) (Walek and Fajmon, 2023) and collaborative filtering (CF) (Lin et al., 2023).

The CB approach recommends items that are similar to those in which the user has shown interest in the past. The recommendation process includes building item representation by comments or feedback and user representation by corresponding actions provided by the user to express his interests (Karydi and Margaritis, 2016). But this approach has shortcomings in identifying user preferences from user actions on one content source among all other types and assigning object attributes due to the limitation of resources (Anandhan et al., 2018).

The CF approach recommends items by other individuals' similar tastes or item similarity. In traditional CF methods, only the feedback matrix, which contains either explicit (e.g., ratings) or implicit feedback (e.g., tagging, clicks, purchases) on the items given by users, is used for training and prediction. Typically, the feedback matrix is sparse, leading to unsatisfactory performance (Seyedhoseinzadeh et al., 2022). In that case, many researchers have proposed utilizing auxiliary information to alleviate the data sparsity in CF.

Among that auxiliary information, community-related information is commonly exploited, especially in community-based social media. Li et al. propose a content recommender system, which provides a novel interest-based clustering and cluster-based content recommendation solution based on an analysis of content popularity and user interests in social communities (Zheng et al., 2019). Chao et al. improve M-learning content recommendation by learning user behavioral characteristics within communities (Campos et al., 2022). Fu et al. improve the CF algorithm by integrating it with the K-means clustering algorithm and community factors, which achieves better recommendation accuracy and speed (Fu et al., 2019). Lee and Brusilovsky explore the feasibility and the value of using users' community membership as a source of personalized recommendations for individual users and find that it generates recommendations as accurately as CF, but with better efficiency (Lee and Brusilovsky, 2017).

These studies demonstrate the effectiveness of community-related information in enhancing content recommendation. Meanwhile, we also see that these studies consider community as the result of interest-based clustering, focus on information within the community and use it to enrich user-content interaction, which is the only relation these methods care about. However, there exist relations beyond the user-content interaction, such as user-community interaction and content-community affiliation. They are inherent in community-based social media systems because of their taking communities as the content organization form. Users usually select communities and then select possibly interested content among the options and their preferences are inherited during the process. Without leveraging such relations, it is difficult to accurately capture user preferences.

2.3. User engagement in online communities

Communities are collections of individuals, typically with a common interest (Joyce and Kraut, 2006). Users consume content from multiple communities and each community has multiple users interested in its content. There are researchers considering communities as a means to infer latent user characteristics and interests (Gasparetti et al., 2021, Zarrinkalam et al., 2018, Pochampally and Varma, 2011), and it is believed that community information reflects user preferences, and can be applied to recommender systems for content consumption (Velichety and Ram, 2020).

Despite this characteristic, user engagement in social communities is often sparse and uneven (Sun et al., 2020). Jones et al. conduct empirical research on a sample of 578 Usenet newsgroups active and find that only 11.5% of people participate in the group again after their initial post (Jones et al., 2004). That means user attention and interests are dynamic and changing in community-based social media, bringing challenges to model user preferences in recommender systems.

With such a phenomenon, researchers have investigated factors that contribute to user continued engagement. Moreland and Levine conclude that when newcomers return to a group by posting again, this action represents a minimal expression of commitment (Moreland and Levine, 2006). Joyce considers interaction as the basis of commitment and develops some routes through which users' interaction with a group may increase their commitment to it (Joyce and Kraut, 2006). One of them is described by the reinforcement model (Ferster and Skinner, 1957), which comes from Skinner's findings that people are conditioned to respond to messages to them. Getting a positive response, and getting a response that fulfills some explicit needs are all reinforcing events. In the case of social communities, people will be more likely to continue their participation in the community if they experience some reinforcing events (Wang et al., 2022). We can also assume that the reinforcement events cultivate the user's preferences related to the community. Besides, another interaction-related factor is contributions. Moreland and Levine believe that whether an individual eventually becomes committed to the group is likely to depend on the contributions the individual makes to the group. The more a user contributes to the community, the more likely he is to continue participating in the community in the future.

2.4. Translating embedding

Translating embedding technique is derived from the Knowledge Graph domain and used to embed entities and relations in a continuous vector space. It considers relations as translations in the embedding space and represents them by low-dimensional vectors. There are many related methods proposed. For example, Bordes et al. propose a simple and easy-to-train model TransE to learn vector embeddings of entities and relations (Bordes et al., 2013). The basic motivation behind TransE is that translations are natural transformations for representing the relationship between two entities. Specifically, the embedding of the tail entity should be close to the embedding of the head entity with the relation vector. TransH is proposed based on TransE, considering more mapping properties of relations (Wang et al., 2014). Furthermore, TransR is proposed to model entities and relations in distinct spaces (Lin et al., 2015). As a very promising method, translating embedding is applied to the recommender system domain and many translation-based recommendation models have come out, such as TransRec (Garcia-Duran et al., 2020) and LRML (Tay et al., 2018). Its design motivation fits with the aim of our study and one of the classic models TransE is chosen as the basis. As for choosing TransE instead of TransR and TransH, it is because the latter two models are proposed based on TransE by further considering mapping properties and semantic spaces separately while these two aspects are not the focus of this study. Additionally, TransE is scalable and lightweight with a

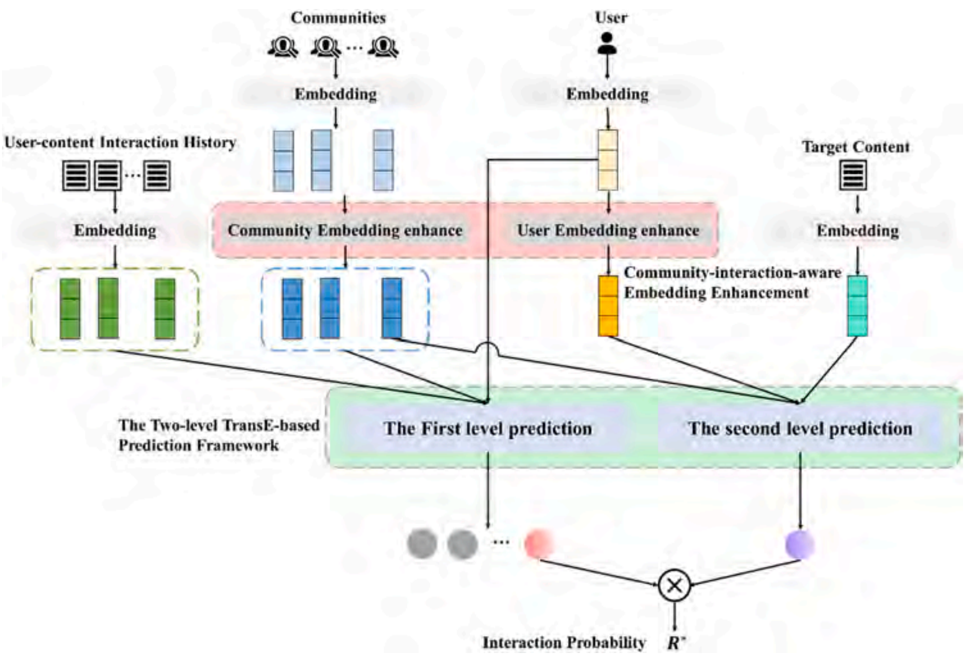


Fig. 2. The structure of CISCRec.

reduced number of parameters (Bordes et al., 2013), thus making the TransE-based models easy to train and scale up to other large databases.

Information-seeking behavior explains the existence of hierarchical relations between user, content, and community in the content consumption behavior, which contains user preference information, while such information has not been used in previous studies. Thus, we tend to model the implicit relations and learn entity and relation embedding based on the TransE model. Further, inspired by the research on user engagement in online communities, we enhance embeddings with interaction-related factors, aiming to improve recommendation and reasoning. More details about the proposed content recommendation model are discussed in Section 3 "METHODOLOGY".

3. Methodology

In this section, we first introduce the content recommendation and the notations used in this paper. Then, the proposed model Community-aware Information Seeking based Content Recommender (CISCRec) is described in detail. Fig. 2 shows the structure of CISCRec. From the figure, we can see that CISCRec consists of two main modules: a two-level TransE-based prediction framework and an interaction-aware embedding enhancement mechanism, which will be introduced in Section 3.2 separately.

3.1. Problem definition and notation

We aim to improve the content recommendation problem with relational data. The problem can be illustrated as follows. There are three key entity types, namely: the user set $U = u_1, u_2, u_3, \ldots, u_N$, the content set $I = i_1, i_2, i_3, \ldots, i_M$ and the community set $C = c_1, c_2, c_3, \ldots, c_K$. For each user u, $H_u = h_{u,1}, h_{u,2}, h_{u,3}, \ldots, h_{u,V}$ is defined as her/his interaction history, representing a set of interacted contents for users u with size V at time step $t-1$. Since every content is posted in a certain community, we further obtain the user-to-community interaction history $G_u = g_{u,1}, g_{u,2}, g_{u,3}, \ldots, g_{u,K}$, where $g_{u,K} = 1$ indicate the user u have engaged the community c, otherwise $g_{u,K} = 0$. Based on the above, we summarize three relation pairs around three entities: user-to-content interaction (u, i), user-to-community interaction (u, c) and content-to-community affiliation (i, c). Given the user-content interaction history H_u, the user-community interaction history G_u as well as the intrinsic relations, we aim to predict whether a user u has a potential interest in a community c or content i which he has not seen before.

Furthermore, interaction-related factors on user engagement in communities $\boldsymbol{F} = F_1, F_2, \ldots, F_P$ can be used to better capture users' preferences for communities, thus, improving the effectiveness of information retrieval services. The corresponding interaction-related factor scores of each community are estimated during the process for user u. Take $F_{1,u_1} = f_{1,u_1,c_1}, f_{1,u_1,c_2}, \ldots, f_{1,u_1,c_K}$ as an example, f_{1,u_1,c_K} refers to the factor score on F_1 of community c_K for user u_1. And all factor scores must be mapped to between 0 and 1.

Before the items in the set (e.g. $u_1, u_2, u_3, \ldots, u_N$ in U and $i_1, i_2, i_3, \ldots, i_M$ in I) are fed into the model, we first transfer them into a list of indices and multiply the indices with a matrix of a predetermined size, therefore generating embedding vectors of each item. In the process of model training, the matrix parameters are updated, leading to item embedding changing as well.

3.2. Community-aware information seeking based content recommender (CISCRec)

3.2.1. Two-level TransE-based prediction framework

In community-based social media, content is organized in the topic-based community. Based on the studies about information seeking behavior described in Section 2.2, we assume that users consume content in community-based social media based on class hierarchies, that is, they find a topic community they prefer on the home page to find content related to their interests. The behavior can be divided into two phases, selecting the community within the community set and selecting the content in the community. By decomposing the content consumption behavior in the user information search process, we propose a two-level TransE-based prediction framework, in which basic TransE is utilized to model users' content consumption behavior by learning entity and relation embedding to improve the prediction accuracy of user preferences at the micro-level.

The first level aims to predict user preferences for communities by modeling user-to-community interaction relations through translation. There exists a relation between the user and community, and we use user-content interaction history H_u to represent it by the thought that users' diverse interests can be captured from their past behaviors (Li et al., 2021). Hence, the user-to-community prediction process based on a translational operation derived from TransE can be described as follows:

$$R(c|u, H_u) = bias_c - d(\theta_u + \theta_{H_u}, \theta_c) \tag{1}$$

where θ_u is the embedding vector of user u, θ_{H_u} is the embedding vector of user interaction history in the timestep $t-1$, θ_c is the embedding vector of community c and $bias_c$ is a bias term. Each entity and relation vector will be learned during the training process, and $R(c|u, H_u)$ is finally obtained. It refers to the value of which community is selected in the next timestep t.

Particularly, we normalize $R(c|u, H_u)$ into the final predicted weight $P(c|u,H_u)$, using the SoftMax function, to ensure $\sum_{c \in C} P(c|u,H_u) = 1$. The normalization is shown as:

$$P(c|u, H_u) = \frac{\exp(R(c|u, H_u))}{\sum_{c' \in \hat{C}} \exp(R(c'|u, H_u))} \tag{2}$$

Likelihood values of the target user selecting candidate communities in their future interaction are generated in this phase. It aims at improving the quality of the community choice, thus, improving the effectiveness of information seeking, that exploits communities as data sources.

The second level tends to predict user preferences for content by modeling user-to-content interaction relations through translations. With the precondition that the user selects community first, we choose the intrinsic content-to-community affiliation to represent the relation between the user and content. Similarly, a translation-based predictor is proposed to estimate whether the target user selects the target content with the context of the user and the community to which the content corresponds. The user-to-content predictor can be described as follows:

$$R(i|u,c') = bias_i - d(\theta_u + \theta_{c'}, \theta_i) \tag{3}$$

where u is the target user, i is the target content, c' is the community target content corresponding to, θ_u is the embedding vector of user u, θ_i is the embedding vector of content, $\theta_{c'}$ is the embedding vector of community c' and $bias_i$ is a bias term. $R(i|u,c')$ captures the transformation between target content and user in the embedding space.

The introduction of community is for relation extension between users and content to better capture user preferences for the target content. Thus, predictors from two levels are naturally integrated into the final user-to-content predictor, which is defined as:

$$R^*(i|u,H_u) = P(c'|u,H_u) \times R(i|u,c') \tag{4}$$

3.2.2. Interaction-aware embedding enhancement

From the two-level TransE-based prediction framework above, community entity plays an important role in predicting user preferences for content. Therefore, accurately predicting users' choices of communities has become the key to improving the performance of CISCRec. Considering that users' interaction behaviors influence their continued participation in the corresponding communities, we integrate interaction-related factors into the embedding vectors, aiming to add expressiveness and reasoning. The interaction-aware embedding enhancement mechanism is mainly imposed on community and user embedding in the two-level TransE-based prediction framework. The structure is shown in Fig. 3.

(1) Community Embedding Enhancement

We aim to enhance the community embedding vector in the two-level TransE-based prediction framework by utilizing implicit relationships about how users are influenced to continually participate in communities. The implicit relationships can be represented via multiple types of community interaction-related factors. In this paper, we use two commonly available factors:

- *Comment Score*. The score that the comment submission has accumulated, which is usually the number of upvotes minus the number of downvotes.
- *Comment Length*. The number of words in comments posted by the user within a specific community.

According to the literature review, reinforcement events and contributions are found to be strongly related to the user's preferences for the community. The factor "comment score" can be used to quantify the level of reinforcement a user obtains in a specific community. In general, positive votes motivate users' participation, while negative votes significantly decrease their participation (Chen et al., 2019). And the factor "comment length" can be used to measure users' contribution to the community (Wang et al., 2021). Therefore, the selection of these two factors helps convincingly verify our assumptions about user engagement in online communities. Most importantly, they are commonly available, making this method applicable to a variety of social media scenarios.

These two factors have different scales, so we apply the min-max normalization to map the factor score into [0,1] and add them to the initial community embeddings. The introduction of these factors specifies different weights to different communities of each user in the set and the enhancement process of a given community factor imposed on the original community embedding can be described as follows:

$$V_{u,f_{p,u}} = \{\theta_{c_1} f_{p,u,1}, \theta_{c_2} f_{p,u,2}, \ldots, \ \theta_{c_K} f_{p,u,K}\} \tag{5}$$

where $\theta_{c_1, \ \ldots, K}$ are the embedding vectors of communities, and K is the size of the community set. $f_{p,u,K}$ denotes the factor score of user u on community c_K on a given factor type p. $V_{u,f_{p,u}}$ denotes the factor p based embedding of each community for user u.

Interaction-related factors may play different roles in representing users' preferences for communities. Thus, a factor attention mechanism is proposed. Every user is associated with a community factor weighted vector $\varnothing_{u,}$ with size P, which is the number of factors. With the factor attention mechanism, the embedding of each community for user u on factor p is defined as follows:

$$V^*_{u,f_{p,u}} = \varnothing_{u,p} V_{u,f_{p,u}} \tag{6}$$

$V^*_{u,f_{1,u}}, V^*_{u,f_{2,u}}, \ldots, V^*_{u,f_{p,u}}$ are combined through a concatenation aggregator to get V^*_u, as shown in:

$$V^*_u = \mathrm{CONCAT}\left(V^*_{u,f_{1,u}}, V^*_{u,f_{2,u}}, \ldots, V^*_{u,f_{p,u}}\right) \tag{7}$$

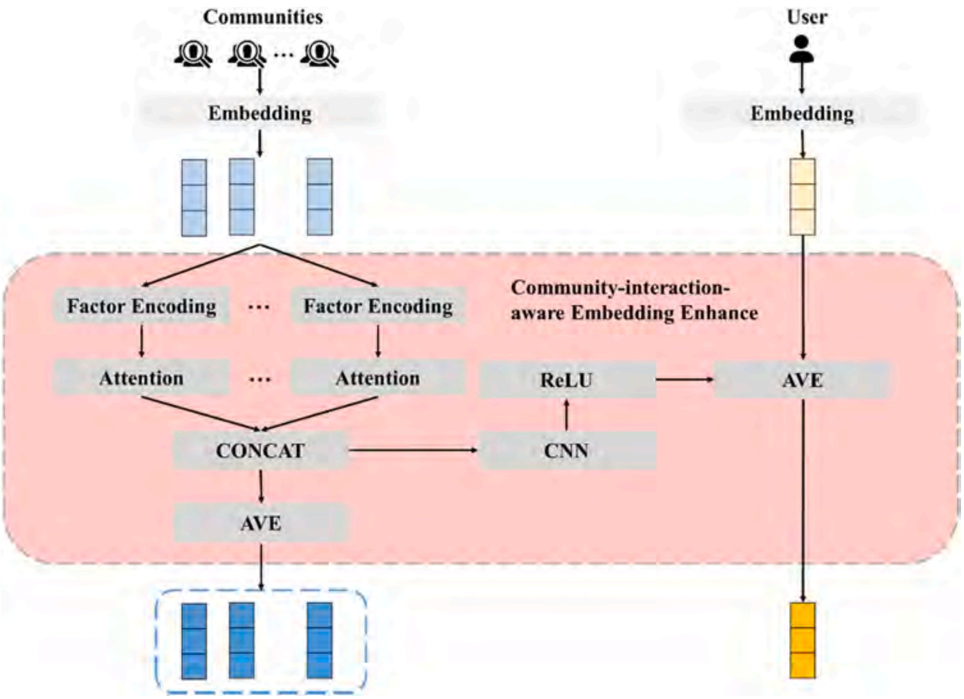


Fig. 3. The structure of interaction-aware embedding enhancement.

The same concatenation is imposed on the identification embedding vector of all communities to get θ_C. which is shown as:

$$\theta_C = \mathrm{CONCAT}(\theta_{c_1}, \theta_{c_2}, \ldots, \theta_{c_K}) \tag{8}$$

Next, we aggregate θ_C and V_u^* by an average aggregator to obtain final differentiated community embeddings toward the user, which is defined as:

$$\theta_{C,u}^* = \mathrm{AVE}\left(\theta_C, V_u^*\right) \tag{9}$$

After this operation, each user has heterogeneous community embeddings, carrying more preference information.

(1) User Embedding Enhancement

The next step is to enhance user embeddings in the second level of the framework. We feed the above V_u^* to a 1D convolutional neural network to generate the latent community embedding vector $V_u^{'}$. It aims to enrich the feature representation and extract complex information and $V_u^{'}$ is a high-dimensional representation of user preferences. To add more expressiveness, an activation function is applied to process $V_u^{'}$, which is shown as

$$V_u^{''} = \mathrm{ReLU}\left(V_u^{'}\right) \tag{10}$$

$V_u^{''}$ is generated by Eq. (10). Similarly, to capture users' global characteristics, we combine $V_u^{''}$ with the identification embedding vector of user θ_u via an average aggregator to get the enhanced user embedding, which is described as

$$\theta_u^* = \mathrm{AVE}\left(\theta_u + V_u^{''}\right) \tag{11}$$

3.2.3. Model learning

CISCRec aims to address s ranking-based recommendation task, that is, to rank contents user likely to be of interest higher than irrelevant content. Bayesian Personalized Ranking (BPR) loss function is commonly used to optimize ranking models (Rendle et al., 2012). Thus, we design our loss function following the BPR's loss function. It is defined as follows:

$$min\mathrm{L} = \sum_{(u,H_u,i,i^-)\in D} \ln\sigma(R^*(i|u, H_u) - (R^*(i^-|u, H_u)) \tag{12}$$

where σ denotes the sigmoid function, making each term a "probability" between 0 and 1. $D = \{(u, H_u, i, i^-) | u \in U \wedge H_u \in I_{t-1,u} \wedge i \in I_{t,u} \wedge i^- \in I - I_{t,u} - I_{t-1,u}\}$

The above objective function encourages the value of the positive content i larger than that of the negative content i^- given user u

and his interaction history H_u.

4. Experiments

4.1. Dataset and evaluation metrics

To examine the effectiveness of the proposed model, we conduct experiments with the Reddit datasets. Reddit is a discussion website that covers a variety of topic communities called “subreddits”. Users can submit content and comment or vote on submissions. Given that Reddit is one of the most used social media platforms (Medvedev et al., 2019), having 430 million monthly active users in 2020[1], and most other platforms have communities like Reddit subreddits, we believe that our proposed model could be extended and reused for content recommendations on another community-based social media if it is verified on Reddit.

In the collection of Reddit data, we use the data provided by the social media data collection platform Pushshift (Baumgartner et al., 2020). Pushshift has collected all the submissions and comments posted on Reddit since June 2005 and makes them available to researchers. We use two sequential comment datasets which include all submissions and comments on Reddit in January 2009 and February 2009. There are important keys in data files such as comment_id, author_name, submission_id, subreddit_id, comment_text, score, created_utc, etc. Each submission is considered as a posted item, and commenting actions are considered as implicit feedback. Specifically, the datasets include 1.3M users, 9.6M submissions 48M comments.

In the data preprocessing, inactive users which have fewer than 15 associated actions respectively in two datasets will be removed from the dataset. For a specific user, we get interaction history from the January 2009 data and randomly select 80% of interacted contents in the February 2009 data as the training set and the remaining 20% as the test set. Table 1 lists the statistics of our dataset.

We focus on recommending top-N items, so we employ several widely used evaluation metrics precision, recall, mean reciprocal rank (MRR) and normalized discounted cumulative gain (NDCG) to evaluate the performance of our model. Each metric is calculated at a given cut-off rank, denoted as @k, where $k \in \{5, 10, 15\}$. Recall measures the proportion of the relevant items returned in the top-k ranked list, while Precision measures the proportion of relevant items in the top-k ranked list.

MRR and NDCG are position-aware ranking metrics (Liu et al., 2023).

MRR evaluates a recommender by computing the rank of relevant items in the list, which is formulated as:

$$MRR@k = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{rank_i} \tag{13}$$

where n is the number of users and $rank_i$ is the rank of relevant items.

NDCG is to evaluate ranked list as it gives higher rewards for the top items in a recommended list, which is defined as:

$$NDCG@k = \frac{DCG@k}{IDCG@k} \tag{14}$$

where $DCG@k = \sum_{i=1}^{N}\frac{2^{rel_i}-1}{\log_2(1+i)}$, rel_i is 1 if the recommended item at position i in the rank list is relevant, and 0 otherwise. $IDCG@k$ is the $DCG@k$ of the sorted optimal ranked list.

4.2. Baselines

In this section, we briefly describe the commonly used content recommendation methods we adopt as baselines.

- Item-based CF (2004): This is a K-nearest neighborhood-based method recommending items based on item similarity. Cosine similarity is adopted to compute the item similarity (Deshpande and Karypis, 2004).
- User-based CF (2004): Different from Item-based CF, it recommends items based on user similarity (Jin et al., 2004).
- Content-based (2007): This method recommends items to users based on their past interests (Pazzani and Billsus, 2007).
- MF (2009): This method models the user feedback matrix as a product of two lower-rank user and item matrices and applies BPR Optimization Criterion to optimize the model (Koren et al., 2009).
- FM (2010): Compared to BPRMF, this model considers the second-order feature interactions between inputs (Rendle, 2010).
- Mostpop (2017) (2020): It is defined as a non-personalized method and recommends the most popular items in the entire training set (Otunba et al., 2017). While in this paper, we take the modified Mostpop method by taking time into account, which is proposed in (Ji et al., 2020).
- LightGCN (2020): This is a deep-learning-based method. It learns user and item embeddings by linearly propagating them on the simplified Graph Convolution Network, which is proved to be easier to implement and train (He et al., 2020).
- SNNCB (2021): As a content-based method, it is proposed by using Siamese Neural Networks (SNNs) to determine the similarity between two contents (Pulis and Bajada, 2021).

[1] https://www.businessofapps.com/data/reddit-statistics/

Table 1
Statistics of Reddit dataset after preprocessing.

Users	4868
Contents	115491
Communities	850
Interactions	602025

4.3. Parameter setting

In experiments, the entity embeddings and community factor weighted vector are randomly initialized with a Gaussian distribution (0, 0.1). Biases are initialized to a constant 0. The parameters are learned during the training process.

The results of Precision@10, Recall@10, MRR@10, and NDCG@10 when changing hyper-parameters including embedding dimensions, the learning rate of the optimization algorithm, and bias in the loss function are shown in Fig. 4 (a-c). After investigating the impact of hyper-parameters on the performance of CISCRec, embedding dimensions are set to 8, the learning rate is set to 0.01, and the bias is set to 0.0001. The other parameters are configured as follows: the batch size is set to 512 and the max epoch is set to 500. In the convolutional layer, the number of output channels is set to 2, and the kernel size is ($\sim \ \times$ *embedding dimensions* $- 3$). We use the same parameter settings for the baseline models, following a previous work (Sun et al., 2020).

4.4. Experimental results and discussion

4.4.1. Model comparison

Figs. 5-8 provide the top-k recommendation performance comparison of different models under Precision, Recall, MRR and NDCG respectively. Comparing the results achieved by CISCRec and baselines, the Precision value of the former is superior to that of baselines in all cases where k takes different values / in different *k*s. On average, CISCRec outperforms ItemCF, UserCF, CB, MF, FM, Mostpop, LightGCN and SNNCB by 7.52%, 11.93%, 6.77%, 17.70%, 9.15%, 8.07%, 7.25% and 6.28%, respectively. Concerning the Recall, on average, CISCRec achieves 3.40%, 7.88%, 3.79%, 9.86%, 3.94%, 3.03%, 2.34%, and 3.46% improvement compared to ItemCF, UserCF, CB, MF, FM, Mostpop, LightGCN, and SNNCB, respectively. They indicate the effectiveness of CISCRec in recognizing relevant items according to user preferences.

Similar results are demonstrated on two other metrics. Under the MRR, on average, CISCRec shows 41.78%, 50.49%, 27.11%,

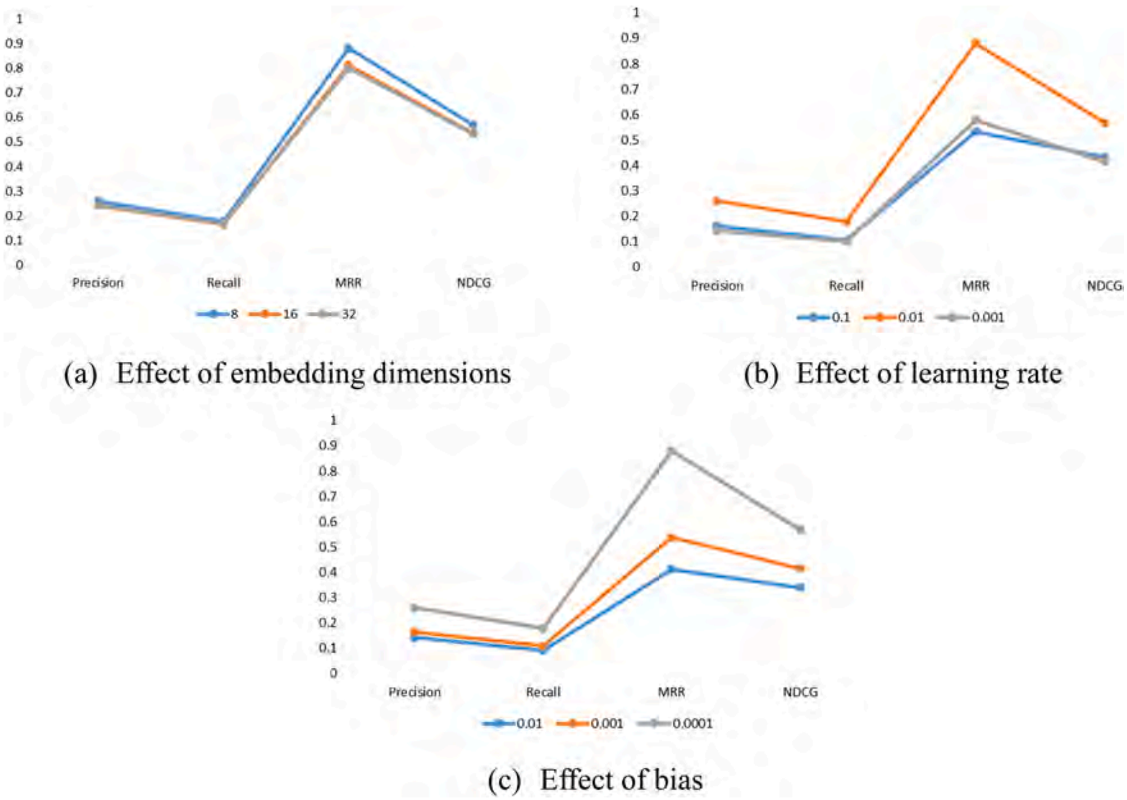
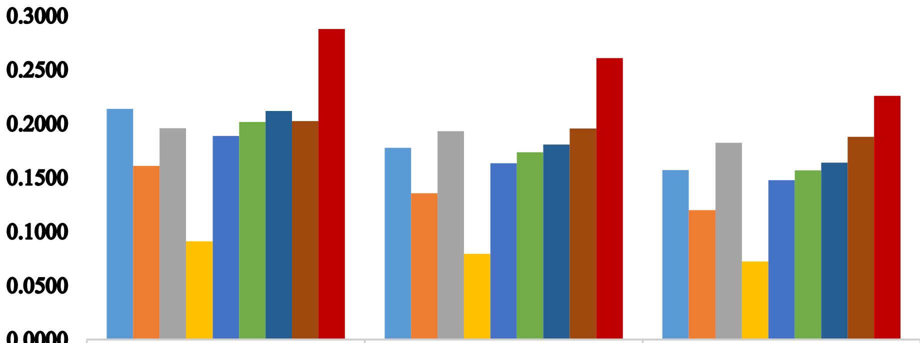
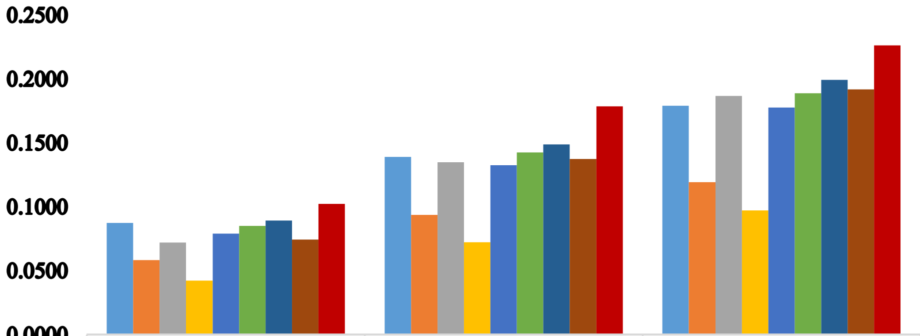


Fig. 4. Model performance for different evaluation metrics.



	Pre@5	Pre@10	Pre@15
ItemCF	0.2138	0.1778	0.1571
UserCF	0.1609	0.1356	0.1199
CB	0.1959	0.1931	0.1824
MF	0.0912	0.0796	0.0725
FM	0.1888	0.1634	0.1478
Mostpop	0.2018	0.1735	0.1569
LightGCN	0.2118	0.1809	0.1641
SNNCB	0.2024	0.1956	0.1879
CISCRec	0.2878	0.2608	0.2258

Fig. 5. Model performance for precision.

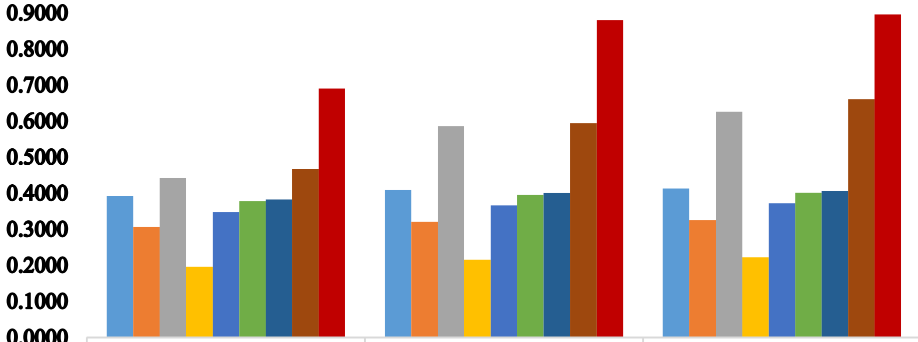


	Rec@5	Rec@10	Rec@15
ItemCF	0.0875	0.1392	0.1791
UserCF	0.0584	0.0938	0.1193
CB	0.0722	0.1351	0.1869
MF	0.0425	0.0723	0.0972
FM	0.0792	0.1326	0.1778
Mostpop	0.0852	0.1426	0.1890
LightGCN	0.0894	0.1489	0.1994
SNNCB	0.0745	0.1375	0.1920
CISCRec	0.1025	0.1788	0.2265

Fig. 6. Model performance for recall.

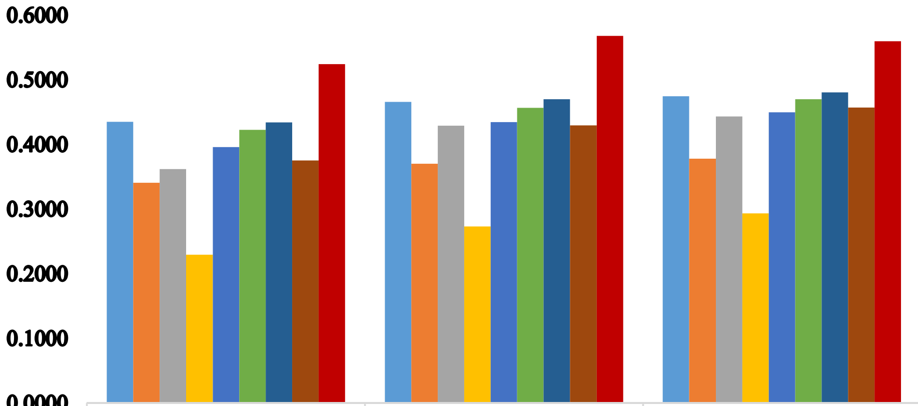
61.11%, 46.03%, 43.09%, 42.60%, and 24.83% improvement compared to ItemCF, UserCF, CB, MF, FM, Mostpop, LightGCN and SNNCB, respectively. As for the NDCG, the value of CISCRec is respectively higher than that of ItemCF, UserCF, CB, MF, FM, Mostpop, LightGCN and SNNCB by 9.21%, 18.77%, 13.94%, 28.54%, 12.40%, 10.10%, 8.93% and 13.02%. It is demonstrated that CISCRec can put relevant items at the top of the ranked list, therefore improving recommendation utility.

For a more visual comparison, the bar charts are also shown in Figs. 5-8. The proposed CISCRec achieves the best performance among all methods on the Reddit dataset regarding all different top-k values. According to the observations, we make the following



	MRR@5	MRR@10	MRR@15
ItemCF	0.3918	0.4083	0.4126
UserCF	0.3058	0.3209	0.3247
CB	0.4420	0.5853	0.6256
MF	0.1959	0.2151	0.2219
FM	0.3472	0.3661	0.3720
Mostpop	0.3771	0.3953	0.4009
LightGCN	0.3825	0.4003	0.4054
SNNCB	0.4670	0.5938	0.6604
CISCRec	0.6900	0.8803	0.8958

Fig. 7. Model performance for MRR.



	NDCG@5	NDCG@10	NDCG@15
ItemCF	0.4354	0.4662	0.4749
UserCF	0.3411	0.3705	0.3781
CB	0.3619	0.4292	0.4436
MF	0.2297	0.2734	0.2936
FM	0.3962	0.4346	0.4499
Mostpop	0.4228	0.4570	0.4700
LightGCN	0.4341	0.4701	0.4807
SNNCB	0.3753	0.4295	0.4574
CISCRec	0.5245	0.5683	0.5600

Fig. 8. Model performance for MRR.

discussions:

(1) Compared with the conventional item-based CF and user-based CF which utilize collaborative similarity, CISCRec is based on the multiple and concrete relations among users, communities, and contents. This is the major reason for CISCRec's

improvement over CF. We also find that ItemCF achieves better performance than UserCF. The result may be attributed to data sparsity, which has a varying impact on the models. UserCF doesn't perform well because the content set in social media is quite large and the percentage of users who interact with content is really low.

(2) Compared with the latent factor-based methods MF and FM, the performances of CISCRec improve significantly. This may be for the following reasons. Although MF and FM identify users and items by decomposing the user-item interaction matrix into the product of two lower-dimensionality rectangular matrices, they fail to explicitly represent them with auxiliary information extracted from that data. Besides, we can see that FM achieves better overall performance than MF because it considers the second-order feature interactions between inputs. Without high-order feature interaction, CISCRec still works better, which confirms the effectiveness of our two-level prediction framework and embedding enhancement mechanism.
(3) Compared with the Mostpop, the CISCRec model works better among all four indicators. This is because Mostpop is a non-personalized method, only considering the content popularity for recommendation while CISCRec models user preferences. Interestingly, we also noticed that Mostpop is slightly better than other personalized approaches like ItemCF, UserCF, MF and FM. It indicates that the popularity of content can directly affect the consumption of content in social media. For example, a major worldwide event is likely to garner wide attention while specific content relevant only to a few may not be as attention-grabbing.
(4) Although LightGCN is a strong baseline, the performance of CISCRec is superior to that of it. We can draw some inferences from our experiments. LightGCN can learn higher-order patterns from user-item interaction and add more expressiveness to user and item representations, but it is insufficient to capture user preferences. CISCRec incorporates translational embeddings to model the relations between users, contents and communities, therefore refining user preferences at different stages. We argue that this is the major resource for improvement.
(5) Although CB, SNNCB, and CISCRec all use user interaction history to model user preferences and SNNCB even adds an SNN content-based similarity metric, CISCRec achieves better results. The reason is that CB and SNNCB recommend content to users based on their interests reflected in their past content interaction behavior, while CISCRec introduces community and incorporates theories related to user engagement in online communities to effectively model user future preferences, which in turn can help in better representation learning. The SNN module improves SNNCB compared to CB, but the due to the limited sources, the marginal effect improvement from this deep learning module is not too high.

4.4.2. Ablation studies of interaction-related factors

The proposed CISCRec enhances the user and community embeddings by integrating two community interaction-related factors. To estimate the effectiveness of these two components, we conduct ablation experiments by removing the corresponding parts. Table 2 reports the results of the top 10 recommendations when removing different community factors.

It can be seen that model with the "score" factor CISCRec-s achieves better performance than CISCRec-none by an average of 8.80% over four metrics, demonstrating the effectiveness of feedback attributes. Generally speaking, the feedback a user has received in the past within the community directly influences his willingness to participate in the community in the future. Hence, the attribute helps capture user preferences and improve the recommendation effectiveness of the basic model.

As for another factor, although the introduction of the "comment-length" factor aims at capturing user preferences from a community perspective according to the premise that if a user has contributed more to the community in the past, the greater the likelihood that they will continue to participate in the community in the future, the performance of CISCRec-c is worse than the basic model. However, the model combined both "score" and "comment-length" factors, namely CISCRec, improves the CISCRec-none by 20.78% on average, much higher than that of CISCRec-s. The reason behind this observation may be the single use of the "comment-length" factor could not differentiate user preferences but instead brings the noise. But the problem can be solved by the joint use of "comment-length" and "score" factors. The introduction of the "score" factor eliminates the noise brought by the "comment-length" factor, and in turn "comment-length" factor refines user preferences for communities from another perspective and increases the expressiveness of embedding. Moreover, these results indicate the importance of selecting appropriate attributes to improve recommendation models.

4.4.3. Individual case study

We randomly select user A in the Reddit dataset to see how CISCRec makes recommendations. Fig. 9 shows the process of the top 5 recommendations. We can see that CISCRec first predicts user A's preference weights for all communities by using his last visited content data. Then CISCRec predicts the probability of user A choosing candidate contents with content-to-community affiliation known. Finally, the two values are multiplied to get the score for each candidate post. After ranking, the top 5 recommendation list is determined.

5. Conclusions

5.1. Theoretical and practical implications

In this paper, we understand how users consume content in community-based social media from an Information Seeking behavior perspective, and present a novel content recommendation model CISCRec, which extends user-content interaction relations by considering the community entity. Specifically, we build a two-level TransE-based prediction framework to predict user-item interactions from a micro and hierarchical perspective based on user content consumption behavior. The first level predicts user preferences for communities by modeling user-to-community interaction relations based on user-content interaction history. The second

Table 2
ablation experimental results on top-10 recommendation.

Model	Precision@10	Recall@10	MRR@10	NDCG@10	RI
CISCRec-none	**0.2109**	**0.1443**	**0.7089**	**0.4993**	–
CISCRec-s	0.2355	0.1624	0.7772	0.5259	+8.80%
CISCRec-c	0.2069	0.1446	0.6650	0.4708	-4.87%
CISCRec	0.2608	0.1788	0.8803	0.5683	+20.78%

Note: RI is the relative improvement on the CISCRec-none. CISCRec-none denotes that the embedding enhancement mechanism based on two interaction-related factors is removed. CISCRec-s means only considering the "score" factor for embedding enhancement. CISCRec-c means only considering the "comment-length" factor for embedding enhancement.

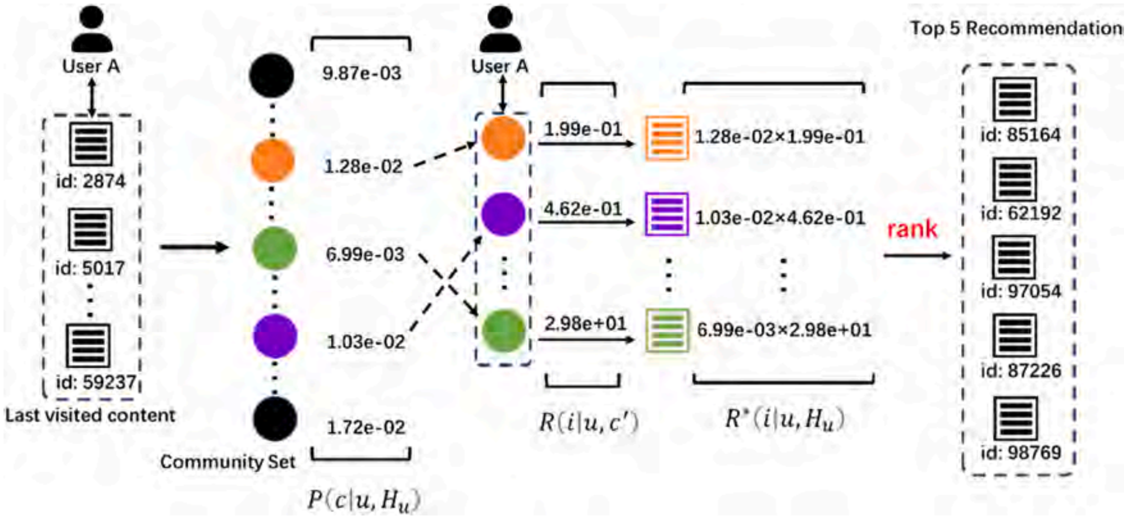


Fig. 9. An example of recommendations.

level predicts user preferences for the target content by modeling user-to-content interaction relations based on content-to-community affiliation. The two levels' outputs are naturally integrated into the final user-to-content predictor, which clearly illustrates why the content is recommended. Further, to improve the accuracy of the two-level prediction framework, we develop interaction-aware embedding enhancement techniques for entity embeddings by integrating community interaction-related factors. It not only adds more expressiveness to user and community embeddings but also facilitates the reasoning of user engagement in communities. Extensive experiments on real-world Reddit datasets demonstrate the superiority of CISCRec compared with other commonly-used content recommenders.

From a practical perspective, the proposed CISCRec also shows its great significance in creating commercial benefits. There are multiple community-based social media applications, such as Pinterest, Twitter, Douban, etc., providing content recommendation services. Our work can be applied to them to benefit the platforms by effectively modeling users' preferences and improving users' satisfaction by providing content they like quickly and accurately. In addition, techniques in our model can be employed to aid other types of recommendations, such as product, music, and movie recommendations.

5.2. Limitations and future work

In this study, our work has made some progress in the content recommendation literature. However, there is a question worth further exploration in the two-level TransE-based prediction framework, that is, how to accurately predict user preferences for new communities without any records. In our view, one of the possible approaches is to assign preference value to new communities based on the value assigned by the user to other similar communities. For future studies, we can build a community network based on semantic relevance and use neighbor structures to compute the similarity. Besides, we would like to incorporate additional factors in the interaction-aware embedding enhancement mechanism, such as dialogue factors (Duck, 1998), to better predict user preferences for communities from different perspectives and validate relevant theories in return.

Ethical statement

The authors certify that this manuscript is original and has not been published and will not be submitted elsewhere for publication while being considered.

Funding statement

This research was supported by grants from National Natural Science Foundation of China (71901152, 71701134), Guangdong Basic and Applied Basic Research Foundation (2023A1515012515, 2023A1515010919), Guangdong Province Innovation Team "Intelligent Management and Interdisciplinary Innovation" (2021WCXTD002), and Shenzhen Stable Support Grant (20220810100952001).

CRediT authorship contribution statement

Chen Yang: Conceptualization, Methodology, Writing – original draft, Project administration, Funding acquisition. **Ruozhen Zheng:** Validation, Writing – original draft. **Xuanru Chen:** Validation, Writing – original draft. **Hong Wang:** Methodology, Writing – original draft, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Data availability

No data was used for the research described in the article.

References

Agrawal, S., Roy, D., & Mitra, M. (2021). Tag embedding based personalized point of interest recommendation system. *Information Processing & Management, 58*(6), Article 102690.
Anandhan, A., Shuib, L., Ismail, M. A., & Mujtaba, G. (2018). Social media recommender systems: review and open research issues. *IEEE Access, 6*, 15608–15628.
Bates, M. J. (2002). Toward an integrated model of information seeking and searching. *The New Review of Information Behaviour Research, 3*(1), 1–15.
Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The pushshift reddit dataset. In *, 14. Proceedings of the international AAAI conference on web and social media* (pp. 830–839).
Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*, 26.
Cai, Y., Ke, W., Cui, E., & Yu, F. (2022). A deep recommendation model of cross-grained sentiments of user reviews and ratings. *Information Processing & Management, 59*(2), Article 102842.
Campos, R., dos Santos, R. P., & Oliveira, J. (2022). Providing recommendations for communities of learners in MOOCs ecosystems. *Expert Systems with Applications, 205*, Article 117510.
Chen, L., Baird, A., & Straub, D. (2019). Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community. *Decision Support Systems, 118*, 21–32.
Chen, X., Du, C., He, X., & Wang, J. (2020). Jit2r: A joint framework for item tagging and tag-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 1681–1684).
Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS), 22*(1), 143–177.
Duan, R., Jiang, C., & Jain, H. K. (2022). Combining review-based collaborative filtering and matrix factorization: A solution to rating's sparsity problem. *Decision Support Systems, 156*, Article 113748.
Duck, S. (1998). *Human relationships*. Sage.
Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement.
Fu, W., Liu, J., & Lai, Y. (2019). Collaborative filtering recommendation algorithm towards intelligent community. *Discrete and Continuous Dynamical Systems-S, 12* (4&5), 811–822.
Garcia-Duran, A., Gonzalez, R., Onoro-Rubio, D., Niepert, M., & Li, H. (2020). Transrev: Modeling reviews as translations from users to items. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42* (pp. 234–248). Springer International Publishing.
Gasparetti, F., Sansonetti, G., & Micarelli, A. (2021). Community detection in social recommender systems: a survey. *Applied Intelligence, 51*, 3975–3995.
He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval* (pp. 639–648).
Ji, Y., Sun, A., Zhang, J., & Li, C. (2020). A re-visit of the popularity baseline in recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1749–1752).
Jin, R., Chai, J. Y., & Si, L. (2004). An automatic weighting scheme for collaborative filtering. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 337–344).
Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research, 15*(2), 194–210.
Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication, 11*(3), 723–747.
Karydi, E., & Margaritis, K. (2016). Parallel and distributed collaborative filtering: A survey. *ACM Computing Surveys (CSUR), 49*(2), 1–41.
Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30–37.
Lee, D. H., & Brusilovsky, P. (2017). Improving personalized recommendations using community membership information. *Information Processing & Management, 53* (5), 1201–1214.
Lee, L., Ocepek, M. G., & Makri, S. (2022). Information behavior patterns: A new theoretical perspective from an empirical study of naturalistic information acquisition. *Journal of the Association for Information Science and Technology, 73*(4), 594–608.
Li, D., Hu, B., Chen, Q., Wang, X., Qi, Q., Wang, L., & Liu, H. (2021). Attentive capsule network for click-through rate and conversion rate prediction in online advertising. *Knowledge-Based Systems, 211*, Article 106522.
Li, Y. J., Hoffman, E., & Zhu, D. (2022). Should firms pay for online brand communities: Using lead user theory in analyzing two contrasting cases. *Decision Support Systems, 155*, Article 113729.
Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *, 29. Proceedings of the AAAI conference on artificial intelligence*.

Lin, J., He, M., Pan, W., & Ming, Z. (2023). Collaborative filtering with sequential implicit feedback via learning users' preferences over item-sets. *Information Sciences, 621*, 136–155.
Liu, X., Werder, K., & Maedche, A. (2020). Novice digital service designers' decision-making with decision aids—A comparison of taxonomy and tags. *Decision Support Systems, 137*, Article 113367.
Liu, X., Wu, K., Liu, B., & Qian, R. (2023). HNERec: Scientific collaborator recommendation model based on heterogeneous network embedding. *Information Processing & Management, 60*(2), Article 103253.
Matni, Z., & Shah, C. (2014). For the love of information: Motivations and affective dynamics of surfing the web for pleasure. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1–10.
Medvedev, A. N., Lambiotte, R., & Delvenne, J. C. (2019). The anatomy of Reddit: An overview of academic research. *Dynamics on and of Complex Networks III: Machine Learning and Statistical Physics Approaches, 10*, 183–204.
Moreland, R. L., & Levine, J. M. (2006). Socialization in organizations and work groups.
Otunba, R., Rufai, R. A., & Lin, J. (2017). Mpr: Multi-objective pairwise ranking. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 170–178).
Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. *The Adaptive Web: Methods and Strategies of Web Personalization*, 325–341.
Pochampally, R., & Varma, V. (2011). User context as a source of topic retrieval in Twitter. In *Workshop on Enriching Information Retrieval (with ACM SIGIR)* (pp. 1–3).
Poli, R. (1996). Ontology for knowledge organization. *Advances in Knowledge Organization, 5*, 313–319.
Pulis, M., & Bajada, J. (2021). Siamese neural networks for content-based cold-start music recommendation. In *Proceedings of the 15th ACM Conference on Recommender Systems* (pp. 719–723).
Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618.
Rendle, S. (2010). Factorization machines. In *2010 IEEE International conference on data mining* (pp. 995–1000). IEEE.
Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. *Recommender systems handbook*, 1–34.
Seyedhoseinzadeh, K., Rahmani, H. A., Afsharchi, M., & Aliannejadi, M. (2022). Leveraging social influence based on users activity centers for point-of-interest recommendation. *Information Processing & Management, 59*(2), Article 102858.
Shah, D., Webster, E., & Kour, G. (2023). Consuming for content? Understanding social media-centric consumption. *Journal of Business Research, 155*, Article 113408.
Sun, R., Chen, C., Wang, X., Zhang, Y., & Wang, X. (2020). Stable community detection in signed social networks. *IEEE Transactions on Knowledge and Data Engineering, 34*(10), 5051–5055.
Sun, Z., Yu, D., Fang, H., Yang, J., Qu, X., Zhang, J., & Geng, C. (2020). Are we evaluating rigorously? benchmarking recommendation for reproducible evaluation and fair comparison. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 23–32).
Tarus, J. K., Niu, Z., & Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing, 22*, 2449–2461.
Tay, Y., Anh Tuan, L., & Hui, S. C. (2018). Latent relational metric learning via memory-based attention for collaborative ranking. In *Proceedings of the 2018 world wide web conference* (pp. 729–739).
Velichety, S., & Ram, S. (2020). Finding a Needle in the Haystack-Recommending Online Communities on Social Media Platforms Using Network and Design Science. Forthcoming. *Journal of Association for Information Systems*.
Walek, B., & Fajmon, P. (2023). A hybrid recommender system for an online store using a fuzzy expert system. *Expert Systems with Applications, 212*, Article 118565.
Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In *, 28. Proceedings of the AAAI conference on artificial intelligence*.
Wang, X., Ounis, I., & Macdonald, C. (2021). Leveraging review properties for effective recommendation. In *Proceedings of the Web Conference 2021* (pp. 2209–2219).
Wang, N., Liu, Y., & Xiao, S. (2022). Which feedback matters? The role of expressions and valence in continuous high-quality knowledge contribution in the online Q&A community. *Decision Support Systems, 156*, Article 113750.
Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative market research: an international journal, 16*(4), 362–369.
Yadav, N., Mundotiya, R. K., & Singh, A. K. (2021). Tag-based Personalized Collaborative Movie Recommender System. *Journal of Information Assurance & Security, 16*(1).
Zarrinkalam, F., Kahani, M., & Bagheri, E. (2018). Mining user interests over active topics on social networks. *Information Processing & Management, 54*(2), 339–357.
Zhang, Z., Zhuang, F., Qu, M., Lin, F., & He, Q. (2018). Knowledge graph embedding with hierarchical relation structure. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3198–3207).
Zhang, B. Z., Liu, T., Corvite, S., Andalibi, N., & Haimson, O. L. (2022). Separate Online Networks During Life Transitions: Support, Identity, and Challenges in Social Media and Online Communities. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–30.
Zheng, J., Wang, S., Li, D., & Zhang, B. (2019). Personalized recommendation based on hierarchical interest overlapping community. *Information Sciences, 479*, 55–75.
Zheng, R., Qu, L., Cui, B., Shi, Y., & Yin, H. (2023). AutoML for Deep Recommender Systems: A Survey. *ACM Transactions on Information Systems*.
Zhou, K., Yang, C., Li, L., Miao, C., Song, L., Jiang, P., & Su, J. (2023). A folksonomy-based collaborative filtering method for crowdsourcing knowledge-sharing communities. *Kybernetes, 52*(1), 328–343.