Research article

Recommendation model based on intention decomposition and heterogeneous information fusion

Suqi Zhang[1],*, Xinxin Wang[2], Wenfeng Wang[2], Ningjing Zhang[2], Yunhao Fang[3] and Jianxin Li[3]

[1] School of information engineering, Tianjin University of Commerce, Tianjin 300134, China
[2] School of science, Tianjin University of Commerce, Tianjin 300134, China
[3] School of IT, Deakin University, Burwood, VIC 3125, Australia

* Correspondence: Email: suqizhang@tjcu.edu.cn; Tel: +18602290977; Fax: 02260438123.

Abstract: In order to solve the problem of timeliness of user and item interaction intention and the noise caused by heterogeneous information fusion, a recommendation model based on intention decomposition and heterogeneous information fusion (IDHIF) is proposed. First, the intention of the recently interacting items and the users of the recently interacting candidate items is decomposed, and the short feature representation of users and items is mined through long-short term memory and attention mechanism. Then, based on the method of heterogeneous information fusion, the interactive features of users and items are mined on the user-item interaction graph, the social features of users are mined on the social graph, and the content features of the item are mined on the knowledge graph. Different feature vectors are projected into the same feature space through heterogeneous information fusion, and the long feature representation of users and items is obtained through splicing and multi-layer perceptron. The final representation of users and items is obtained by combining short feature representation and long feature representation. Compared with the baseline model, the AUC on the Last.FM and Movielens-1M datasets increased by 1.83 and 4.03 percentage points, respectively, the F1 increased by 1.28 and 1.58 percentage points, and the Recall@20 increased by 3.96 and 2.90 percentage points. The model proposed in this paper can better model the features of users and items, thus enriching the vector representation of users and items, and improving the recommendation efficiency.

Keywords: recommendation model; intention decomposition; heterogeneous information fusion; attention mechanism; graph neural networks

1. Introduction

In recent years, with the rapid development of mobile internet and big data, people can access a lot of online content, such as news, movies and various items, but a lot of information can make users confused. The recommendation model can learn the user's interests and preferences according to the user's attributes and historical behaviors, and select the items that the user may be interested in from the mass of information to recommend to the user [1,2]. It can improve the efficiency of information screening, solve the problem of information overload in the era of big data, and improve the user experience. Collaborative filtering algorithm is the most widely used in all kinds of recommendation systems [3,4]. However, the existing recommendation model uses user-item interaction data and considers that the existence of an interaction is the existence of a preference relationship, thus modeling the representation of users and items, while ignoring the different intentions of users behind these items. As shown in Figure 1, Figure 1(a) is an existing recommendation model that assumes that user and item interactions arise from interest r_0, so user interactions with items represent user preferences. In fact, user-item interaction may also be due to crowd psychology or shopping for others, as shown in Figure 1(b). Therefore, an in-depth analysis of the user-item interaction, considering how the user intended to interact with the item, can improve the accuracy of user and item modeling, thus improving the accuracy and interpretability of recommendations.

(a) existing recommendation models   (b) multi-intention recommendation models
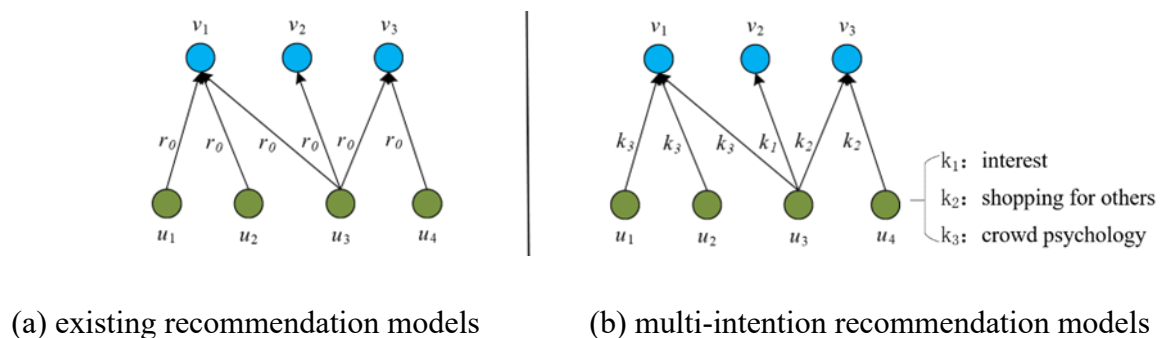
Figure 1. Comparison of recommended model.

On the one hand, the present research on the intention-based recommendation model focuses on the single modeling of intention. In 2020, Zhu et al. [5] put forward the key-array memory neural network (KA-MemNN) model, which regards the type of item as the intention, through the two-layer attention mechanism to get the representation of users. In 2020, Guo et al. [6] proposed intention modeling from ordered and unordered facets (IMFOU), which divides user intentions into long-term interest preferences and recent purchase motivations on a temporal axis. Finally, the two parts are combined to get the representation of the user. In 2021, Wang et al. [7] proposed a knowledge graph-based intent network (KGIN), in which each intention is considered as a combination of the relationships in the Knowledge graph. Moreover, during the node aggregation process, the intention is aggregated as a relationship to be represented by the user.

On the other hand, consider the multiple intentions behind user and item interactions. In 2020, Wang et al. [8] proposed the disentangled graph collaborative filtering (DGCF) model, which decomposes the reason behind the interaction into multiple intents and obtains the representation of user-and item-specific intents, at the same time, the distance correlation is used to guarantee the

intention independence. In 2020, Hu et al. [9] proposed a Graph neural News recommendation model with unsupervised preference disentanglement (GNUD), which considers intention in news recommendation, by decomposing the reasons behind the interaction into multiple intents, the user and the news are represented with specific intents. In 2021, Chen et al. [10] proposed the decomposed collaborative filtering (DCF) model, which treats an intention as the implicit relationship between users and items, and incorporates a knowledge graph. The item representation is constructed by combining the knowledge graph and the intention decomposition. In 2022, Huang et al. [11] proposed a recommendation model based on neural embedding singular value decomposition (NESVD). The model considers the intention factor behind the interaction, and gets more abundant user vector representation through intention decomposition.

In summary, the proposed models focus on the reasons behind user and item interactions, but do not take into account the timeliness of intent. First, from the user's point of view, the user's long-term interactive item reflects the user's stable interest preference, so the interest preference intention is valid for a long time. The intention to crowd psychology and shop for others is only valid in the recent term and is likely to change over time. For example, in Figure 2(a), users have crowd psychology a lot of popular online products over the last year, and users interact crowd psychology with Li Ning's short sleeves, Li Ning's sneakers and Erke's sneakers. Compared to last year's crowd psychology, recent crowd psychology items determine what users are likely to interact with next. Second, from the point of view of an item, the users of a long-term interaction item embody the long-term stable attribute features of the item, while the user intention of a recent interaction item will make the item have the feature of crowd psychology in the near future. For example, in Figure 2(b), for a pair of Erke shoes, in the long run, it is a pair of Erke shoes, but the recent Zhengzhou Rainstorm, the Erke Group as a patriotic enterprise donation, by enabling users to buy the brand's products in succession, Erke's sneakers have recently taken on a crowd psychology character. So, we think about the intent behind the interaction in the recent term, decompose the intent of the recent interaction, and think about what the user and the item are interacting with in the near term because of that intent.

Although considering the timeliness of user-item interaction intention can improve the accuracy of user and item modeling, only using user-item interaction information to the model can improve the accuracy of user item modeling, still faced with data sparse [12–14] and cold start [15–18] problems.

With the popularization of social platforms, people's decision-making is often influenced by their friends, and the knowledge graph contains rich information on item attributes and associations, therefore, social networks [19] and knowledge graphs are introduced into recommender system for auxiliary information, which can alleviate the problem of data sparsity and cold-start. Therefore, how to make full use of heterogeneous information such as user-item interaction information, social information and knowledge graph content information, is a hot research topic in the recommendation field.
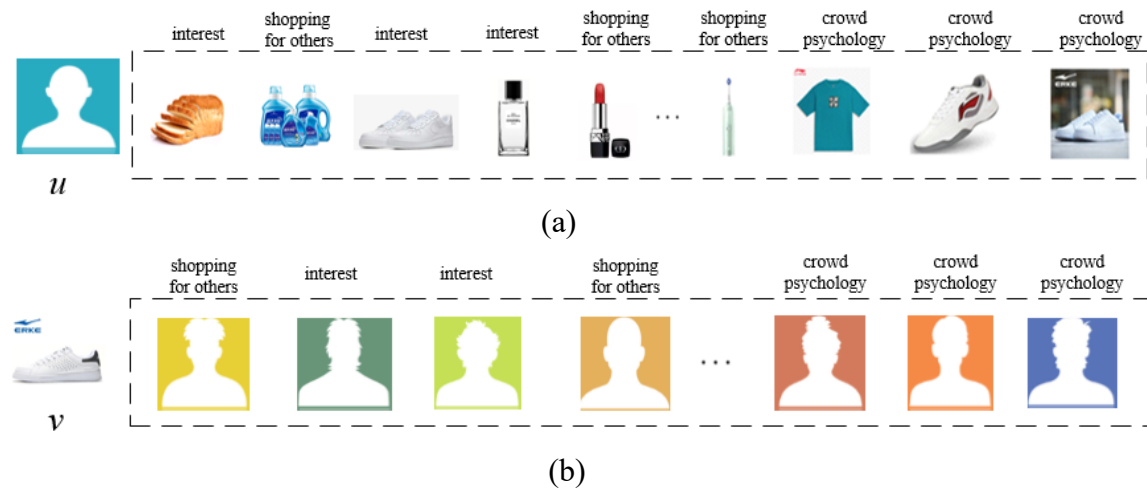
Figure 2. Timeliness of intent.

In 2018, Wang et al. [20] put forward a Ripple Net model, which takes the target item as the center, spreads and aggregates the information of the surrounding nodes along the relationship in the knowledge graph, and constructs the feature vector of users and items; Wang et al. [21] put forward the KGCN model, which introduces the graph convolution network into the domain of knowledge graph recommendation. KGCN combines the information of entity nodes and neighborhood nodes to compute the entity embedding vector containing high-order association information; In 2019, Wang et al. [22] put forward the KGAT model, which fuses the knowledge graph and the user item interaction graph (UIG) into the collaborative knowledge graph (CKG), based on the attention mechanism, the neighborhood information of nodes is aggregated on the collaborative knowledge graph, and the high-order association information between entities is mined through the stack multi-layer network structure, and the embedded vector representation of users and items is constructed. In 2020, Wang et al. [23] put forward the CKAN model, which embeds the content information of knowledge graphs to get the feature vector of the item based on the attention mechanism, and then represents the target user as the sum of all the item feature vectors in their interaction history. According to the similarity between the user vector and the item vector, the click rate of the user to the item is predicted. In 2019, GraphRec [24] studied the embedding representation of users in the user social network and the UIG, respectively, and finally combined the user embedding representation in the two graph spaces to get the final user embedding representation.

The above models use social networks or knowledge graphs to model users and items. However, in the real world, the content information contained in knowledge graphs cannot cover all the features of items, and user interaction with items is more than just item content preferences. For example, when a user watches a movie, in addition to being interested in the content of the movie, they may be influenced by factors such as recommendations from friends.

As a result, recent research has begun to fuse heterogeneous information such as user-item interactions, social information, and knowledge graph content information. In 2021, T-MRGF [25] model is the first to creatively incorporate heterogeneous information, fusing user vectors from user-item interaction graphs and user social networks, fusing item vectors from user-item interaction graphs and item-item knowledge graphs, aggregate information from high-level user-item interaction graphs, social networks, and knowledge graphs. However, the T-MRGF model directly fuses the interactive

features in the user item interaction graph, the content features in the knowledge graph and the social features in the social network, the noise caused by heterogeneous information fusion is ignored, which results in inaccurate recommendation results. In 2022, the KGRFSF [26] model projects heterogeneous information of user item interaction graph and knowledge graph into the same preference feature space, reducing the noise of fusion.

In this paper, the timeliness of user-item interaction intention and the noise problem caused by heterogeneous information fusion are studied. The main work is as follows:

1) We proposed a recommendation model based on intention decomposition and heterogeneous information fusion (IDHIF), in which the short-term feature representation from the intention decomposition (ID) module and the long-term feature representation from the heterogeneous information fusion (HIF) module are concatenated together to get the final user and item representation.

2) We designed the ID module to address the timeliness problem of user-item interaction intention. To obtain the item and the user representations under different intentions, this module first decomposes the user's recent interaction history item and the user of the recent interaction item. We employed the long short-term memory (LSTM) network to mine the representation of users and items under various intentions. We also exploited attention mechanisms to aggregate the short-term features of users and items.

3) We devised a HIF module to deal with the noise problem caused by heterogeneous information fusion. It gets the user's interactive features, the social features, the item's interactive features, and the content features through graph convolution networks on interaction graphs, social networks, and knowledge graphs. By contacting and multilayer perceptron, long-term feature representations of users and items are obtained using different feature vectors projected into the same feature space.

4) We conducted several experiments on Last-FM and MovieLens-1M datasets, where it has been shown that the proposed model cannot only solve the timeliness problem of the intention behind the user's item interaction, but also solves the noise problem that appears in the simple fusion of behavior, content, and interaction, and improves recommendation precision.

2. The design of IDHIF model

In this paper, we design a recommendation model based on intention decomposition and heterogeneous information fusion. The framework of the model is shown in Figure 3. The model mainly includes as follows:

1) **Input layer**: it consists of the input user social network, the user-item interaction graph and the item-item knowledge graph to generate the initial user and item representation.

2) **User Item Feature Modeling Layer**: it includes two modules, the intention decomposition and the heterogeneous information fusion.

The intention decomposing module decomposes the intention of the user by decomposing the l item of the recent interactive users $\{v_{n-l}, v_{n-l-1}, \ldots, v_n\}$ and the l users of the recent interactive items $\{u_{m-l}, u_{m-l-1}, \ldots, u_m\}$, generate the representation of the item under k intentions, we obtain $\{v_{n-l,1}, v_{n-l-1,1}, \ldots, v_{n,1}\}$, $\{v_{n-l,2}, v_{n-l-1,2}, \ldots, v_{n,2}\}$ (for example, two intentions) and the representations of the user under different intentions $\{u_{m-l,1}, u_{m-l-1,1}, \ldots, u_{m,1}\}$, $\{u_{m-l,2}, u_{m-l-1,2}, \ldots, u_{m,2}\}$, input into the LSTM network, we get the user with different intention to represent $u_{k,1}$、 $u_{k,2}$ and the item with different intention to represent $i_{k,1}$、 $i_{k,2}$. In the end, through self-attention aggregation, the short-term features of users and items are expressed as u_{short}, i_{short}.

The heterogeneous information fusion module can aggregate user-item interaction graphs, user

social graphs and item knowledge graphs through graph convolution networks, the user interaction feature u_i and social feature u_f, as well as the item interaction feature v_u and content feature v_i are obtained, and different feature vectors are projected into the same feature space, by concating and multi-layer perceptron (MLP), the long-term features of the user represent u_{long} and the long-term features of the item represent i_{long}, respectively.

3) **Prediction layer**: it combines the short-term feature representation and the long-term feature representation obtained by the user item feature modelling layer, gets the final representation of the user u and the final representation of the item i. Through the prediction layer output the current user to the candidate item forecast probability y.

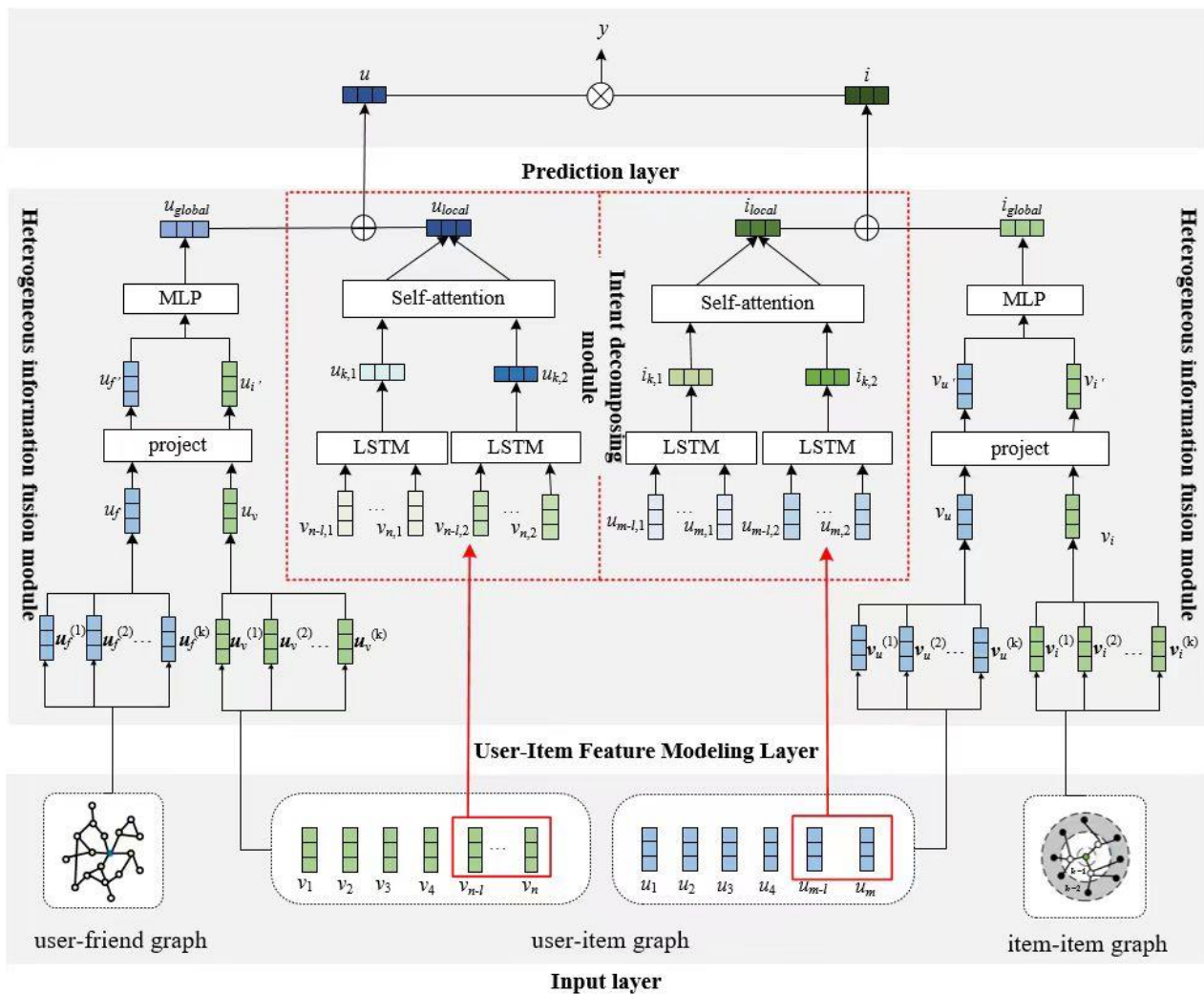The user-item feature modeling layer and prediction layer are detailed in Sections 2.1 and 2.2.

Figure 3. IDHIF model.

2.1. The user-item feature modeling layer

### 2.1.1. The intention decomposition module

There are different intents behind recent user-item interactions. To get a short-term representation of the user u_{short}, we use different short-term intents $K = \{k_1, k_2\}$ decomposes the l historical items for recent interactions $\{v_{n-l}, \cdots, v_n\}$. Finally, we can obtain the historical items that have been recently interacted with are expressed in terms of different intentions:

$$\begin{cases} v_{n-l,1} = \sigma(w_1 \times v_{n-l} + b_1) \\ \qquad \cdots \\ \quad v_{n,1} = \sigma(w_1 \times v_n + b_1) \end{cases} \tag{1}$$

$$\begin{cases} v_{n-l,2} = \sigma(w_2 \times v_{n-l} + b_2) \\ \qquad \cdots \\ \quad v_{n,} = \sigma(w_2 \times v_n + b_2) \end{cases} \tag{2}$$

where w_1, w_2, b_1, b_2 are the weights and offsets of dissimilar intent decomposition.

Due to the sequence of historical items and the timeliness of user interactions, each interactive item will have an impact on users, and the intention of the recent interactive item will have a greater impact on users' choice of new items. LSTM is a network that deals with sequences and can learn the long-term dependencies of recent interaction sequences. Therefore, we exploit LSTM to process the sequence of the user's recent interaction history item under different intentions$\{v_{n-l,k}, \cdots, v_{n,k}\}$. As shown in Figure 4.
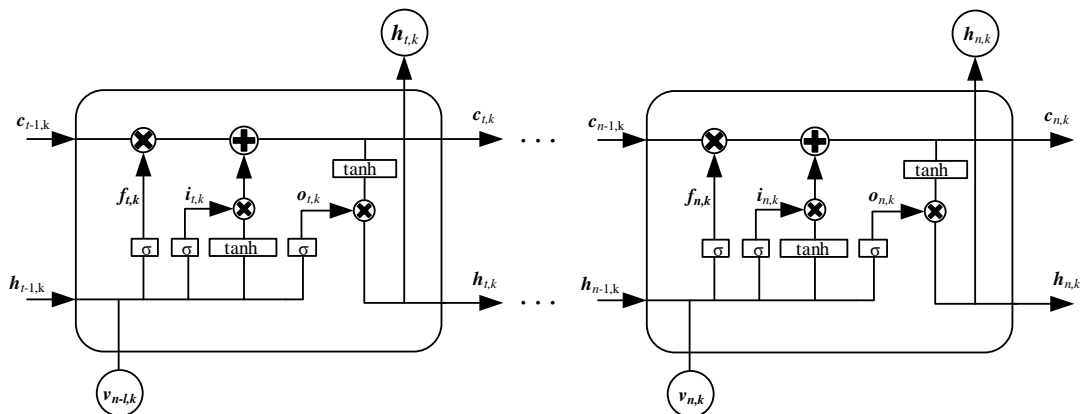
Figure 4. The processing of the item by LSTM.

Under different intents, the sequence of items within the user's recent interactions $\{v_{n-l,k}, \cdots, v_{n,k}\}$ are input into the LSTM according to the time series. The gating mechanism of LSTM is formulated as follows:

$$i_{t,k} = \sigma\left(w_{i,k}\left[h_{t-1,k}, v_{n-l,k}\right]\right) + b_{i,k} \tag{3}$$

$$f_{t,k} = \sigma\left(w_{f,k}\left[h_{t-1,k}, v_{n-l,k}\right]\right) + b_{f,k} \tag{4}$$

$$o_{t,k} = \sigma\left(w_{o,k}[h_{t-1,k}, v_{n-l,k}]\right) + b_{o,k} \tag{5}$$

$$c_{t,k} = f_{t,k} \times c_{t-1,k} + i_{t,k} \times \tilde{c}_{t,k} \tag{6}$$

$$h_{t,k} = o_{t,k} \times \tanh(c_{t,k}) \tag{7}$$

where $i_{t,k}$, $f_{t,k}$ and $o_{t,k}$ denote the memory gate, the forgetting gate and the output gate at t time, respectively. Where $c_{t,k}$ represents long-term stable information, $h_{t,k}$ indicates short-term information. $w_{i,k}$, $b_{i,k}$ are the weight and offset parameters in the memory cell. $w_{f,k}$, $b_{f,k}$ are the weight and offset parameters in the forgotten gate, σ is the sigmoid function. $w_{o,k}$, $b_{o,k}$ are the weight and offset parameters in the output gate. Then, it inputs the item features at t-time into the memory gate $i_{t,k}$ and the forgetting gate $f_{t,k}$, and selects for the features that need to be retained and forgotten in $c_{t-1,k}$.

We end up with the last-minute output of $h_{n,k}$, which is the user representation of under a different intent u_k:

$$u_k = h_{n,k} \tag{8}$$

Similarly, as shown in Figure 5, it can get the last-minute output $h_{m,k}$, which is the representation of the item under a different intent i_k:

$$i_k = h_{m,k} \tag{9}$$
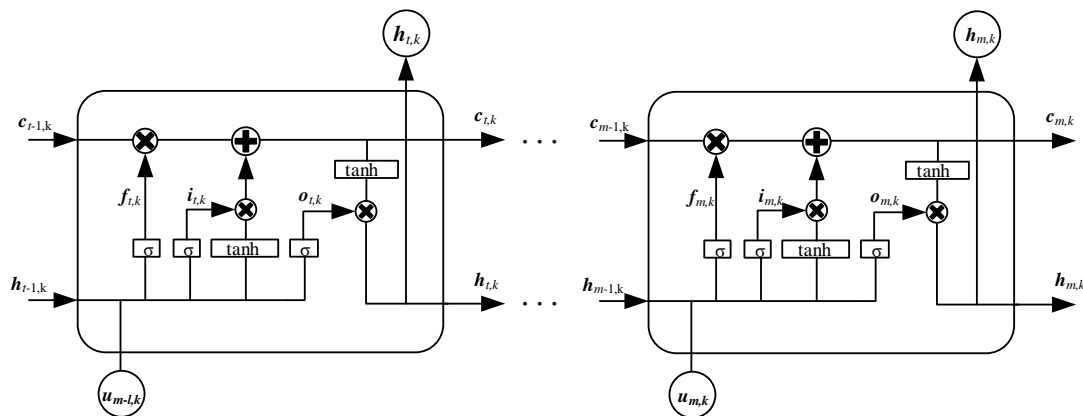
Figure 5. The processing of the user by LSTM.

u_k and i_k are used as the query, the key, and the value in the self-attention mechanism:

$$V_{r,u} = Q_{r,u} = K_{r,i} = \{u_k\} \tag{10}$$

$$V_{r,i} = Q_{r,i} = K_{r,i} = \{i_k\} \tag{11}$$

Then, the similarity matrix Cr of Q_r and K_r is calculated as:

$$Cr = Q_r^T K_r \tag{12}$$

where $Cr = R^{M \times M}$.

Finally, we get the output vector a_r of user and item under different intention after passing through the attention mechanism, that is, the short-term features of user and item represent u_{short} and i_{short}:

$$u_{short} = a_{r,u} = V_{r,u} softmax_\beta \left(\frac{Cr,u}{\sqrt{d}}\right) \qquad (13)$$

$$i_{short} = a_{r,i} = V_{r,i} softmax_\beta \left(\frac{Cr,i}{\sqrt{d}}\right) \qquad (14)$$

where the *softmax* function calculates the weight of each entity vector in V_r.

2.1.2. Heterogeneous information fusion module

In order to get the long-term feature representation $\boldsymbol{u}_{long}$ and $\boldsymbol{i}_{long}$, the neighborhood information of the nodes in the user-item interaction graph is aggregated through the graph convolution network, and the user-item interaction feature representation $\boldsymbol{u}_v$ and $\boldsymbol{v}_u$ is obtained as:

$$\boldsymbol{u}^{(k+1)} = \sum_{v \in N_u} \frac{1}{\sqrt{|N_u|}\sqrt{|N_v|}} \boldsymbol{v}^{(k)} \qquad (15)$$

$$\boldsymbol{v}^{(k+1)} = \sum_{u \in N_v} \frac{1}{\sqrt{|N_v|}\sqrt{|N_u|}} \boldsymbol{u}^{(k)} \qquad (16)$$

where $\boldsymbol{u}^{(k)}$ and $\boldsymbol{v}^{(k)}$ represent the vector representation of user u and item v after k-level convolution, Nu represents the set of interactive items of user u and Nv represents the set of interactive users of item v.

After K-level convolution, we get K vectors. Each vector contains neighborhood information of different ranges, and when $K = 0$, the vector contains only its own information. Obviously, the larger the K value and the wider the neighborhood, the less relevant the information is to the central node. The model synthesizes the results of K-level convolution to obtain the user and item interaction feature vector:

$$\boldsymbol{u}_v = \sum_{k=0}^{K} \alpha_k \boldsymbol{u}^{(k)} \qquad (17)$$

$$\boldsymbol{v}_u = \sum_{k=0}^{K} \alpha_k \boldsymbol{v}^{(k)} \qquad (18)$$

where α_k represents the weight of the layer K vector, the value is set to $1/(K + 1)$ because for the central node, the closer the node information is, the higher its weight.

Similarly, we can aggregate the neighborhood information of the nodes in the user's social network and item knowledge graph using the graph convolution network, where the user social feature is represented by $\boldsymbol{u}_f$ and the item content feature is represented by $\boldsymbol{v}_i$.

The integration of users' interactive features, social features, item's interactive features and content features can enrich the feature expression of users and items and improve the precision of recommendation. However, interactive features, social features and content features contain different semantic information and belong to different feature spaces, direct fusion will bring about "noise" problem.

To solve this problem, we introduce a heterogeneous information fusion module. In this module. First, by projecting, the user interaction feature $\boldsymbol{u}_v$ and social feature $\boldsymbol{u}_f$, as well as the item interaction feature $\boldsymbol{v}_u$ and content feature $\boldsymbol{v}_i$ are projected into the preference feature space,

respectively, get user interaction features representing $\boldsymbol{u}'_v$ and social features representing $\boldsymbol{u}'_f$ in the same preference space, as well as item interaction features representing $\boldsymbol{v}'_u$ and content features representing $\boldsymbol{v}'_i$; then, by feature crossing, from the interaction features, social features and content features to filter out the important feature subset, and to filter the irrelevant semantic features; The user interaction features and social features, as well as the item interaction features and content features, are concatenated together in the preference feature space, and then the features were crossed by MLP, get the long-term features of users and items $\boldsymbol{u}_{long}$ and $\boldsymbol{i}_{long}$:

$$\boldsymbol{u}'_f = \boldsymbol{u}_f \boldsymbol{M}_{uu} \tag{19}$$

$$\boldsymbol{u}'_v = \boldsymbol{u}_v \boldsymbol{M}_{uv} \tag{20}$$

$$\boldsymbol{v}'_u = \boldsymbol{v}_u \boldsymbol{M}_{vu} \tag{21}$$

$$\boldsymbol{v}'_i = \boldsymbol{v}_i \boldsymbol{M}_{ii} \tag{22}$$

$$\boldsymbol{u}_{long} = \sigma\big(\boldsymbol{W}\big(\boldsymbol{u}'_i || \boldsymbol{u}'_f\big) + \boldsymbol{b}\big) \tag{23}$$

$$\boldsymbol{i}_{long} = \sigma\big(\boldsymbol{W}\big(\boldsymbol{v}'_u || \boldsymbol{v}'_i\big) + \boldsymbol{b}\big) \tag{24}$$

where the $\boldsymbol{M}_{uv}$ and $\boldsymbol{M}_{uu}$ indicate the user's interactive feature projection matrix and social feature projection matrix respectively, $\boldsymbol{M}_{vu}$ and $\boldsymbol{M}_{ii}$ are the item's interactive feature projection matrix and content feature projection matrix respectively, $\boldsymbol{W}$ and $\boldsymbol{b}$ are the parameters to be learned.

2.2. The prediction layer

In the last layer, the short-term feature of the user $\boldsymbol{u}_{short}$ and the long-term feature $\boldsymbol{u}_{long}$ of are aggregated in the way of concatenation, which aims to get the final representation of the user $\boldsymbol{u}$. The short-term feature of the item $\boldsymbol{i}_{short}$ and the long-term feature $\boldsymbol{i}_{long}$ of are aggregated in the way of concatenation, which targets to get the final representation of the item $\boldsymbol{i}$:

$$\boldsymbol{u} = \boldsymbol{u}_{short} + \boldsymbol{u}_{long} \tag{25}$$

$$\boldsymbol{i} = \boldsymbol{i}_{short} + \boldsymbol{i}_{long} \tag{26}$$

After getting the user's vector representing $\boldsymbol{u}$ and the candidate item representing $\boldsymbol{i}$, we use the inner product function to get the user's rating of the candidate $\boldsymbol{y}$:

$$\hat{\boldsymbol{y}}(u, i) = \boldsymbol{u}^T \boldsymbol{i} \tag{27}$$

In the end, the complete loss function of the model is defined as:

$$L = \sum_{u \in U}\Big(\sum_{i \in \{i | y_{ui}=1\}} \Gamma(y_{ui}, \hat{y}_{ui}) - \sum_{i \in \{i | y_{ui}=0\}} \Gamma(y_{ui}, \hat{y}_{ui})\Big) + \lambda \|\varTheta\|_2^2 \tag{28}$$

where Γ is the cross-entropy loss function, $\boldsymbol{\theta}$ is the parameter to be trained, λ is the hyperparameter to control the $L2$ regularization. The pseudo-code of IDHIF as shown in Table 1.

Table 1. The pseudo-code of IDHIF.

Model 1: Recommendation model based on intention decomposition and heterogeneous information fusion

Input: user-item graph $G_1 = \{U, I\}$, knowledge graph $G_2 = \{V, E\}$, social graph $G_3 = \{U, F\}$.
Output: user click-through rate on items $\hat{y}(u, i) = \boldsymbol{u}^T \boldsymbol{i}$.

1 **for training times do**
2   According to the formulas (8) and (9), the vector $\boldsymbol{u}_k$ of the user and the vector $\boldsymbol{i}$ of the item under different intent are obtained;
3   According to formula (10)–(14), the user short-term feature vector $\boldsymbol{u}_{short}$ and the item short-term feature vector $\boldsymbol{i}_{short}$ are calculated;
4   According to formula (19)–(24), the user long-term feature vector $\boldsymbol{u}_{long}$ and the item long-term feature vector $\boldsymbol{i}_{long}$ are calculated;
5   The user final vector $\boldsymbol{u}$ and the item final vector $\boldsymbol{i}$ are obtained by formulas (25) and (26);
6   By using formula (27), the click rate of the user to the candidate item $\hat{y}(\boldsymbol{u}, \boldsymbol{v})$ is calculated;
7   The loss value is calculated by formula (28) and the loss training model is minimized;
8   Updating model parameters;
9 **end for**

3. Experiment

This section conducts experiments on publicly available datasets from the film and music field to verify the recommended effect of the model. The dataset is introduced in Section 3.1; the experimental environment and parameter settings and experimental metrics are presented in Section 3.2; the IDHIF proposed in this paper is compared with the other six recommended models in Sections 3.3 and 3.4, the model was validated by ablation experiments in Section 3.5, and the effect of the number of different interactions and different intents on the recommended results was discussed in Section 3.6.

3.1. Experimental data

In this experiment, the performance of the model was tested using the public dataset of movie recommendations and music recommendations. The data comes from Last.FM, the online music platform, and contains music interactions for about 2,000 users, 20,000 friends, and 10,000 triples of knowledge. Movielens-1M is one of the most widely used public datasets for movie recommendation scenarios, containing about one million user ratings, 40,000 friends, and 20,000 triples of the knowledge graph. The dataset was randomly divided into the training set and the testing set in a ratio of 8:2. Detailed statistical results are shown in Table 2.

Table 2. The detail of datasets.

Dataset		Last.FM	MovieLens-1M
User Social Information	Users	1872	6040
	Friends	25064	48320
User-Item Interaction Information	Users	1872	6040
	Items	3846	2347
	Interactions	42346	753772
Item Knowledge Graph Information	Entities	9366	5740
	Relationships	60	12
	Triples	15518	20195

3.2. Experimental setting

3.2.1   Experimental environment and parameter setting

The experimental environment: operating system Windows 64-bit, independent graphics card model NVIDIA GeForce GTX 1650, memory 16GB. The lab tools are PyCharm, Python 3.6, and deep learning TensorFlow 1.14.

The setting of parameters in the experiment is shown in Table 3. Where d is the vector dimension, L is the number of iterations, Lr is the learning rate, λ is the regularization weight, R is the number of recently sampled users and items, and Batch size is the size of the data entered for each batch.

Table 3. Parameter settings.

Dataset	d	L	Lr	λ	R	Batch size
Last. FM	16	1	5×10^{-4}	2×10^{-2}	20	128
Movielens-1M	16	1	2×10^{-2}	1×10^{-7}	30	128

3.2.2   Experimental metrics

In this experiment, Area Under Curve (AUC), F1 value and Recall are used as evaluation metrics. The calculation formula of evaluation metrics is as follows.

The AUC indicates that the model predicts that the probability of users rating like items is greater than the probability of users rating dislike items. The bigger the AUC, the better the prediction effect of the model. Where m' is the number of times a user has rated a favorite item more than the user has rated a disliked item, and m'' is the number of times a user has rated a favorite item as equal to the user has rated a disliked item, m is the total number of comparisons.

$$AUC = \frac{m' + 0.5m''}{m} \tag{29}$$

Recall indicates that the intersection of the collection of items recommended by the model with the collection of items that the user actually interacts with accounts for a percentage of the collection of items that the user actually interacts with. The larger the Recall value, the higher the Recall rate of the model.

$$Recall = \frac{\sum_{u \in U} |Rec(u) \cap Fav(u)|}{\sum_{u \in U} Fav(u)} \qquad (30)$$

F1 is a geometric mean of Accuracy and Recall, which provides a more comprehensive measure of the effectiveness of the algorithm. The higher the F1 value, the higher the prediction accuracy.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (31)$$

3.3. Baseline model

The performance of the IDHIF model presented in this paper is compared with six baseline models.

1) NGCF [27]: It uses only interactive information, and is an optimized collaborative filtering model, which uses item interaction graph structure to model the user-item historical interaction, in order to optimize the user and item embedding and greatly improve the recommendation performance.

2) KGCN [21]: It is a new recommendation algorithm by fusing knowledge graphs. It uses KG as a unit to model the high-level neighborhood, and introduces the attention mechanism to optimize the user and item embedding.

3) GraphRec [24]: It integrates social information and learns the embedded representation of users in the user social network and user-item interaction graph, respectively. In this model, the user embedded representation in two heterogeneous spaces is combined to obtain the final user embedded representation.

4) T-MRGF [25]: It fuses social information and knowledge graph information, fuses user vector from user-item interaction graph and user social network, fuses item vector from user-item interaction graph and item-item knowledge graph.

5) CKAN [23]: It is based on the attention mechanism, which constructs a set of user and item-related content entities through collaborative filtering propagation, and combines the content information contained in the set into a feature vector representation of user and item.

6) DGCF [8]: It considers the intention representation in interactive information. Through the intention decomposition, the representation of user and item-specific intention is obtained. Meanwhile, distance correlation is used to ensure the independence of intention.

7) KGRFSF [26]: It integrates knowledge graph information, and user item interaction information, and carries out projection fusion of feature space for heterogeneous information.

3.4. Performance comparison

Table 4 shows the experimental results for all baseline methods on the two datasets, with the best results marked in bold and the sub-best results underlined. "Improved" denotes the rate of improvement between our IDHIF model and the suboptimal model.

Table 4. Parameter settings.

Models	Last.FM			Movielence-1M		
	F1	Recall@20	AUC	F1	Recall@20	AUC
NGCF	0.6752	0.1450	0.7370	0.7526	0.1252	0.7859
KGCN	0.7186	0.1726	0.7869	0.8094	0.1429	0.8474
GraphRec	0.7206	0.1763	0.7864	0.8049	0.1451	0.8493
T-MRGF	0.7253	0.1820	0.7902	0.8152	0.1534	0.8568
CKAN	0.7306	0.1847	0.7977	0.8235	0.1625	0.8623
DGCF	0.7654	0.2163	0.8139	0.8257	0.1769	0.8652
KGRFSF	<u>0.7990</u>	<u>0.2284</u>	<u>0.8620</u>	<u>0.8260</u>	<u>0.1783</u>	<u>0.8762</u>
IDHIF	**0.8118**	**0.2680**	**0.8803**	**0.8418**	**0.2073**	**0.9165**
Improved	**1.28%**	**3.96%**	**1.83%**	**1.58%**	**2.90%**	**4.03%**

Compared with the baseline model, the IDHIF model proposed in this paper achieves the optimal values on each metric of the two datasets. Specifically, compared with the sub-optimal values of each metric, the AUC metric of IDHIF on Last.FM dataset increased by 1.83 percentage points, the F1 metric increased by 1.28 percentage points, and the Recall@20 metric increased by 3.96 percentage points The AUC on the MovieLens-1M data set improved by 4.03 percentage points, the F1 by 1.58 percentage points, and the Recall@20 by 2.90 percentage points. Further analysis of the experimental results leads to the following conclusions:

1) First, the IDHIF model proposed in this paper is better than NGCF, KGCN, GraphRec, KGRFSF, T-MRGF which are modeled by heterogeneous information. This is mainly due to the fact that NGCF cannot capture the content information of the item on the knowledge graph, and the social information of the user on the social network only by using the historical interaction of the user-item. KGCN models the high-order relationship of item on the knowledge graph to capture the content information of the item, but it cannot capture the social information of users on the social network. GraphRec models users by simply fusing user information from the user social network and the user-item interaction graph, and cannot capture the item's content on the KG. The KGRFSF models the user item information of the user item interaction graph and the item content information of the knowledge graph through feature space fusion, but it does not capture the user's social information. While T-MRGF considers all heterogeneous information of users and items, the user information in the user-item interaction graph and the user's social network is fused, and the item information in the user-item interaction graph and the item-item knowledge graph is fused. However, it is ignored that interactive information, social information and content information contain different semantic information and belong to different feature spaces respectively. Direct fusion will bring about the problem of "noise". This proves the effectiveness of making full use of heterogeneous information of users and items for fusion. At the same time, these models do not consider the multiple intentions behind user-item interactions.

2) Second, the IDHIF model proposed in this paper is better than DGCF model based on intention. This is mainly because the DGCF model only considers multiple intents behind user-item interactions, but does not consider the timeliness of intents. It is shown that the analysis of multiple intentions behind recent user-item interactions can help to more accurately model users and item features, and thus obtain more accurate recommendations.

3) Finally, the results of the IDHIF model presented in this paper are better than those of all baseline models, this is because the IDHIF model in this paper avoids the noise problem caused by simple fusion of heterogeneous information through the fusion method of projection and feature crossing when the heterogeneous information of users and items is fused. At the same time, we consider the user's intention to interact with the items recently and the intention features of the items recently.

3.5. Ablation experiment

In order to verify the impact of intent decomposition module and heterogeneous information fusion module on the performance of IDHIF, five variants of IDHIF model, *W/O* HIF, *W/O* ID, *W/O* SO, *W/O* KG, and *W/O* SOKG are used to compare the two datasets, the results are shown in Table 5, where:

W/O HIF cancels the heterogeneous information fusion module HIF, and retains the intention decomposition module; this model mainly considers the timeliness of user-item interaction intention, and obtains the long-term feature representation by using the long-term interaction training of user-item, the short-term representation is obtained by decomposing the short-term and intention, and the final user and item representation is obtained by combining the long-term representation and the short-term representation.

W/O ID cancels the intention decomposition module and retains the heterogeneous information fusion module. This model mainly considers the noise problem caused by heterogeneous information fusion, and based on the heterogeneous information fusion method, the interactive features of users and items can be found on the interaction graph, the social features of users can be found on the social network, and the content features of items can be found on the knowledge graph. Different feature vectors are projected into the same feature space and the user and item representations are obtained by concating and multi-layer perceptron, respectively.

W/O SO does not capture user social information, but retains the captured user item interaction information and knowledge graph content information for modeling. *W/O* KG does not capture item knowledge graph content information, but retains the captured user item interaction information and user social information for modeling. *W/O* SOKG does not capture item knowledge graph information and user social information, but only retains user item interaction information for modeling.

Table 5 shows the results of the ablation experiment, with all five variants showing a reduction in the IDHIF experiment relative to the overall model. It can be concluded that removing any module or any heterogeneous information in the model results in a degradation of the recommended performance of the model. It shows that considering the timeliness of user-item interaction intention and the noise caused by heterogeneous information fusion, the model can improve the performance of recommendation, in this way, users can accurately recommend items that are more in line with their preferences.

Table 5. Parameter settings.

Models	Last.FM			Movielence-1M		
	F1	Recall@20	AUC	F1	Recall@20	AUC
W/O HIF	0.7952	0.2300	0.8670	0.8295	0.1833	0.8671
W/O ID	0.8032	0.2420	0.8781	0.8377	0.2047	0.9125
W/O SO	0.8022	0.2430	0.8521	0.8231	0.1897	0.8742
W/O KG	0.8036	0.2357	0.8466	0.8253	0.1936	0.8823
W/O SOKG	0.7866	0.2146	0.8263	0.8033	0.1845	0.8732
IDHIF	**0.8118**	**0.2680**	**0.8803**	**0.8418**	**0.2073**	**0.9165**

3.6. The sensitivity of the parameter

In order to evaluate the effect of different parameters on the performance of recommendations, we present the experimental results of IDHIF under different hyperparameter settings and different datasets. The following hyperparameters are adjusted:

Number of recent interactions l: In user-item interaction sequence select l items that have recent interactions and decompose them in order to get the short-term features of user and item, however, when the selection number is small, it will not be enough to represent the recent features of users and items. When the selection number is too large, it will lose the features of recent features. Therefore, it is possible to explore the effect of the number of recently interacting items l on the experimental performance. On Last.FM and Movielens-1M, respectively, set l to range from 10 to 40, and the experimental results are shown in Table 6. It can be observed that the best performance is achieved on the Last.FM dataset when $l = 20$ and on the Movielens-1M dataset when $l = 30$.

Table 6. The effect of the number of different recent interactions on experiments.

l	Last.FM			Movielence-1m		
	F1	Recall@20	AUC	F1	Recall@20	AUC
10	0.8035	0.2563	0.8729	0.8246	0.1849	0.8921
20	**0.8118**	**0.2680**	**0.8803**	0.8321	0.1925	0.9057
30	0.8064	0.2583	0.8703	**0.8418**	**0.2073**	**0.9165**
40	0.7947	0.2458	0.8531	0.8345	0.1987	0.8972

Number of intents K: decomposing the vector representations of recent users and items into multiple intents can more accurately model the features of users and items. However, the number of key intents that affect users' decisions recently is limited, the superfluous intention will interfere with the feature modeling of the model and affect the result of the recommendation. Therefore, we can

explore the effect of intention number K on experimental performance. On Last.FM and Movielens-1M, respectively, set K to range from 1 to 4, and the experimental results are shown in Table 7. The experimental results show that on Last.FM dataset with high data sparsity, the best result is obtained when the intention number is 2, while on the Movielens-1M dataset with rich data, the best result is achieved when the number of intents is set to 3.

Table 7. The effect of the number of different intentions on experiments.

K	Last.FM			Movielence-1m		
	F1	Recall@20	AUC	F1	Recall@20	AUC
1	0.8049	0.2536	0.8769	0.8253	0.1867	0.8954
2	**0.8118**	**0.2680**	**0.8803**	0.8387	0.1935	0.9038
3	0.8032	0.2569	0.8703	**0.8418**	**0.2073**	**0.9165**
4	0.7890	0.2449	0.8531	0.8263	0.1937	0.8898

4. Conclusions

We present a recommendation model that utilizes intention decomposition and heterogeneous information fusion. The model considers user-item interaction features, item content features, and user social features in separate feature spaces. These features are projected into a single preference space for heterogeneous information fusion, which helps to reduce the noise problem that arises from simple heterogeneous information fusion. The model also decomposes the intention of the user and item based on the intention behind them and the timeliness of the intention. This approach enriches the feature representation of both the user and item. The final user and item representation is obtained by combining the short-term feature representation of the intent decomposition module and the long-term feature representation of the heterogeneous information fusion module. Our proposed model is capable of handling noise and this has been validated through experimental results on MovieLens-1M and Last-FM datasets. While the intention modeling and heterogeneous information fusion methods proposed in this paper have been effective in improving recommendation accuracy, it is worth noting that there are other datasets such as knowledge graph, social network, and interaction graph that could be explored in the future. Therefore, the next research direction of this paper is to improve the interpretability of suggestions based on semantic information in user item interaction graph and knowledge graph, and try to find more datasets to validate the model.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

[20310802D] and the Innovation Capacity Improvement Plan Project in Hebei Province [22567603H].

Conflict of interest

The authors declare there is no conflict of interest.

References

1. G. J. Zheng, F. Z. Zhang, Z. H. Zhang, Y. Xiang, N. J. Yuan, X. Xie, et al., DRN: A deep reinforcement learning framework for news recommendation, in *Proceedings of the 2018 World Wide Web Conference*, (2018), 167–176. https://doi.org/10.1145/3178876.3185994

2. G. R. Zhou, X. Q. Zhu, C. R. Song, Y Fan, H. Zhu, X. Zhu, Deep interest network for click-through rate prediction, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (2018), 1059–1068. https://dl.acm.org/doi/10.1145/3219819.3219823

3. H. Liu, B. Yang, D. Li, Graph collaborative filtering based on dual-message propagation mechanism, *IEEE Trans. Cybernetics*, **53** (2023), 352–364. https://ieeexplore.ieee.org/document/9515772

4. A. Hamzehei, R. K. Wong, D. Koutra, F. Chen, Collaborative topic regression for predicting topic-based social influence, *Mach. Learn.*, **108** (2019), 1831–1850. https://doi.org/10.1007/s10994-018-05776-w

5. N. J. Zhu, J. Cao, Y. C. Liu, Y. Yang, H. C Ying, H. Xiong, Sequential modeling of hierarchical user intention and preference for next-item recommendation, in *Proceedings of the 13th ACM International Conference on Web Search and Data Mining*, (2020), 807–815. https://doi.org/10.1145/3336191.3371840

6. X. L. Guo, C. Y. Shi, C. M. Liu, Intention modeling from ordered and unordered facets for sequential recommendation, in *Proceedings of the 2020 World Wide Web Conference*, (2020), 1127–1137. https://doi.org/10.1145/3366423.3380190

7. X. Wang, T. L. Huang, D. X. Wang, Y. C. Yuan, Z. G. Liu, X. N. He, et al., Learning intents behind interactions with knowledge graph for recommendation, in *Proceedings of the 2021 World Wide Web Conference*, (2021), 878–887. https://dl.acm.org/doi/10.1145/3442381.3450133

8. X. Wang, H. Y. Jin, A. Zhang, X. N. He, T. Xu, T. S. Chua, Disentangled graph collaborative filtering, in *Proceedings of the 43st International ACM SIGIR Conference on Research & Development in Information Retrieval*, (2020), 1001–1010. https://dl.acm.org/doi/10.1145/3397271.3401137

9. L. M. Hu, S. Y. Xu, C. Li, C. Yang, C. Shi, N. Duan, et al., Graph neural news recommendation with unsupervised preference disentanglement, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 4255–4264. https://doi.org/10.18653/v1/2020.acl-main.392

10. H. Chen, X. Xin, D. Wang, Y. Ding, Decomposed collaborative filtering: Modeling explicit and implicit factors for recommender systems, in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, (2021), 958–966. https://doi.org/10.1145/3437963.3441826

11. T. Huang, R. Zhao, L. Bi, D. Zhang, C. Lu, Neural embedding singular value decomposition for collaborative filtering, *IEEE Trans. Neural Net. Learn. Syst.*, **33** (2022), 6021–6029. https://doi.org/10.1109/TNNLS.2021.3070853

12. E. O. Aboagye, G. C. James, J. B. Gao, R. Kumar, R. U. Khan, Probabilistic time context framework for big data collaborative recommendation, in *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*, (2018), 118–121. https://doi.org/10.1145/3194452.3194458

13. C. Chen, X. Meng, Z. Xu, T. Lukasiewicz, Location-aware personalized news recommendation with deep semantic analysis, *IEEE Access*, **5** (2017), 1624–1638. https://doi.org/10.1109/ACCESS.2017.2655150

14. X. He, H. Zhang, M. Y. Kan, T. S. Chua, Fast matrix factorization for online recommendation with implicit feedback, in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, (2016), 549–558. https://doi.org/10.1145/2911451.2911489

15. C. Y. Liu, C. Zhou, J. Wu, Y. Hu, L. Guo, Social recommendation with an essential preference space, in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, **32** (2018), 346–353. https://doi.org/10.1609/aaai.v32i1.11245

16. S. Sedhain, A. K. Menon, S. Sanner, L. Xie, D. Braziunas, Low-rank linear cold-start recommendation from social data, *in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, (2017), 236–243. https://doi.org/10.1609/aaai.v31i1.10758

17. M. Volkoves, G. Yu, M. T Poutanen, Dropoutnet: Addressing cold start in recommender systems, in *Proceedings of the Advances in Neural Information Processing Systems*, (2017), 4957–4966.

18. Y. Gu, B. Zhao, D. Hardtke, Y. Sun, Learning global term weights for content-based recommender systems, in *Proceedings of the 25th International Conference on World Wide Web*, (2016), 391–400. https://doi.org/10.1145/2872427.2883069

19. L. Jiang, L. Shi, L. Liu, J. Yao, M. E. Ali, User interest community detection on social media using collaborative filtering, *Wireless Netw.*, **28** (2022), 1177. https://doi.org/10.1007/s11276-021-02826-5

20. H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, et al., Ripple net: Propagating user preferences on the knowledge graph for recommender systems, preprint, arXiv:1803.03467.

21. H. Wang, M. Zhao, X. Xie, W. Li, M. Guo, Knowledge graph convolutional networks for recommender systems, in *Proceedings of the 2019 World Wide Web Conference*, (2019), 3307–3313. https://doi.org/10.1145/3308558.3313417

22. X. Wang, X. He, Y. Cao, M. Liu, T. S. Chua, KGAT: Knowledge graph attention network for recommendation, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (2019), 950–958. https://doi.org/10.1145/3292500.3330989

23. Z. Wang, G. Lin, H. Tan, Q. Chen, X. Liu, CKAN: Collaborative knowledge-aware attentive network for recommender systems, in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, (2020), 219–228. https://doi.org/10.1145/3397271.3401141

24. W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, et al., Graph neural networks for social recommendation, in *Proceedings of the World Wide Web Conference*, (2019), 417–426. https://doi.org/10.1145/3308558.3313488

25. J. Guo, Y. Zhou, P. Zhang, B. Song, C. Chen, Trust-aware recommendation based on heterogeneous multi-relational graphs fusion, *Inform. Fusion*, **74** (2021), 87–95. https://doi.org/10.1016/j.inffus.2021.04.001

26. S. Zhang, X. Wang, R. Wang, J. Gu, J. Li, Knowledge graph recommendation model based on feature space fusion, *Appl. Sci.*, **12** (2022). https://doi.org/10.1016/10.3390/app12178764

27. X. Wang, X. He, M. Wang, F. Feng, T. S. Chua, Neural graph collaborative filtering, in *Proceedings of the 42nd International ACM SIGIR Conference on Research & Development in Information Retrieval*, (2019), 165–174. https://doi.org/10.1145/3331184.3331267