ABSTRACT
Massive Open Online Courses (MOOCs) which enable large-scale open online learning for massive users have been playing an important role in modern education for both students as well as professionals. To keep users’ interest in MOOCs, recommender systems have been studied and deployed to recommend courses or videos that a user might be interested in. However, recommending courses and videos which usually cover a wide range of knowledge concepts does not consider user interests or learning needs regarding some specific concepts. This paper focuses on the task of recommending knowledge concepts of interest to users, which is challenging due to the sparsity of user-concept interactions given a large number of concepts. In this paper, we propose an approach by modeling information on MOOC platforms (e.g., teacher, video, course, and school) as a Heterogeneous Information Network (HIN) to learn user and concept representations using Graph Convolutional Networks based on user-user and concept-concept relationships via meta-paths in the HIN. We incorporate those learned user and concept representations into an extended matrix factorization framework to predict the preference of concepts for each user. Our experiments on a real-world MOOC dataset show that the proposed approach outperforms several baselines and state-of-the-art methods for predicting and recommending concepts of interest to users.

Keywords
User Modeling, MOOC, Learning Analytics, Knowledge Concept, Recommender Systems

1. INTRODUCTION
MOOCs (Massive Open Online Courses), which are free online courses available to anyone to enroll around the world, have gained a lot of popularity in the past decade. By the end of 2018, popular MOOC platforms such as edX3 and Coursera2 have provided 11,400 courses with 101 million users/learners on those platforms3. Previous studies have shown that MOOCs do have a real impact 24, 8. For example, Chen et al. 8 showed that 72% of survey respondents reported career benefits and 61% reported educational benefits. Despite of the popularity, one main challenge of MOOCs is the overall completion rate of those courses is normally lower than 10% 12, 30. Therefore, understanding and predicting user behaviors and learning needs are important to keep users learning on MOOC platforms.

To this end, previous studies have focused on understanding dropout or procrastination behavior 28, 12, 38, 14 and recommending content such as courses and learning paths that a user might be interested in 15, 26, 4. A MOOC can be seen as a sequence of videos where each video is associated with some knowledge concepts. For example, a video in a computer science MOOC can cover several concepts such as “software” and “hardware”. More recently, Gong et al. 13 argued that course or video recommendations overlook user interests regarding specific knowledge concepts. For example, data mining courses taught by different teachers can be quite different in a microscopic view, and a user who is interested in some specific concepts such as “association rules” might be interested in various video clips or learning materials from different teachers covering those concepts from different perspectives. Therefore, understanding a user’s learning needs from a microscopic view and predicting knowledge concepts that the user might be interested in are important.

In this work, we focus on predicting and recommending knowledge concepts that might be interesting to users on MOOC platforms. Based on the interaction history between users and concepts (i.e., a user has interacted with a concept if the user has learned that concept), traditional recommendation approaches such as collaborative filtering (CF) — which recommends similar items (concepts) based on a user’s interaction history or interesting items from similar users — can be applied. However, the sparsity of user-item (user-concept) relationships can limit the performance of CF-based methods. In addition to users and concepts, MOOC platform data normally contain other entities such as courses, videos, and teachers as well as the relationships among those entities.

To cope with the sparsity problem, we model those enti-
ties and their relationships as a heterogeneous information network (HIN) [33], consisting of the entities and relationships inspired by [13], which can be used for learning user (concept) representations/embddings by exploring indirect user-user (concept-concept) relationships with Graph Convolutional Networks (GCNs). Figure 1 illustrates such a HIN which we discuss in detail in Section 3. For example, one can derive a homogeneous user graph based on an indirect path in the HIN, e.g., a graph with users and edges between two users if they have taken the same course. Given such a homogeneous graph, traditional GCNs can be applied to the graph to learn the representations/embddings of users and concepts with respect to the chosen path.

Based on different indirect paths chosen, we can derive various user (concept) representations, and those representations of users (concepts) regarding different paths can be aggregated, e.g., using the mean of those representations. Instead of the straightforward mean aggregation, we propose and investigate different attention mechanisms to derive aggregated user (concept) representations based on different paths. The intuition behind using an attention mechanism is that different paths might have different importance for each user. Afterwards, those learned user and concept representations can be used for predicting the preference scores of concepts for recommendations. Our contributions in this work are as follows: (1) We propose an end-to-end framework [14] for predicting and recommending knowledge concepts of a user’s interest in Section 4; (2) We investigate two attention mechanisms for aggregating information from different meta-paths (the definition can be found in Section 3) to derive user and concept representations. We then incorporate those representations into our extended matrix factorization framework for predicting the preference score of a concept with respect to a user; (3) Finally, we evaluate our approach with several baselines and state-of-the-art approaches in terms of well-established evaluation metrics, and show the effectiveness of our proposed approach in Section 6.

2. RELATED WORK

Recommender Systems and User Modeling on MOOC Platforms. There has been growing interest in recommender systems on MOOC platforms since 2013 with respect to different aspects such as course, video, and learning paths. For instance, the authors in [3] proposed YouEDU, which is a pipeline for classifying MOOC forum posts and recommending instructional video clips that might be helpful for resolving confusion detected in those posts. In [21], the authors showed that peer recommendations can improve users’ engagement significantly in the context of a Project Management MOOC. Dai et al. [9] proposed analyzing course content for recommending personalized learning paths on MOOC platforms. Khalid et al. [18] provides a comprehensive survey on recent advances regarding different recommender systems in the context of MOOCs. More recently, researchers have started modeling user interests in the context of MOOCs while user modeling has been widely studied in other domains such as social media [42]. For example, Li et al. [22] investigated the impact of acquiring user interests via surveys or questionnaires on course recommendations. In [2], the authors proposed LeCoRe which exploits user interest modeling for recommending courses as well as similar users for promoting peer learning in enterprise environment. Gong et al. [13] argued that course recommendations overlook user interests regarding specific knowledge concepts, and studying users’ online learning interests from a microscopic view and recommending knowledge concepts can capture user interests better and provide the flexibility of choosing learning resources of their interest. In this work, we also focus on the microscopic view for knowledge concept recommendations.

Recommendation Approaches with HIN. The basic idea of early recommendation approaches with HIN is to leverage path-based semantic relatedness between users and items over HINs, e.g., leveraging meta-path-based similarities for recommendation. For example, Shi et al. [32] proposed predicting item ratings based on those from similar users measured via different meta-paths. With the advances of graph representation learning, the authors in [31] proposed using pre-trained user and item embeddings based on meta-path information with random walk, and incorporated those pre-trained embeddings as features into an extended matrix factorization framework. The most similar work to ours is Gong et al. [13], which is one of the first works for recommending knowledge concepts on MOOC platforms in a heterogeneous view. The authors showed that their proposed approach outperforms other CF-based baselines as well as metapath2vec [11], which uses learned node representations of a given HIN for knowledge concept recommendations by measuring the similarities between two nodes. Our work differs from [13] in several aspects. First, we formulate interacted concepts for each user as implicit feedback while [13] treated the number of clicks as ratings and formulated the problem as rating prediction for recommending top-k unknown concepts with higher ratings. Secondly, we investigate different attention mechanisms including the one incorporating the latent features of users (items) from matrix factorization. Thirdly, the prediction layer (Eq. 6) for estimating the preference score of a concept is different from [13] which uses the user (item) representations as features for the final prediction.

3. PRELIMINARIES

In this work, we consider the task of predicting and recommending concepts that a user might be interested in based
on their learning history, which includes a set of learned concepts and their contextual information such as courses, videos, etc. With $n$ users $\mathcal{U} = \{u_1, \ldots, u_n\}$, and $m$ concepts $\mathcal{C} = \{c_1, \ldots, c_m\}$, we define an implicit feedback matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ with each entry $r_{u,c} = 1$ if user $u$ has learned concept $c$ and $r_{u,c} = 0$ otherwise. The task can be framed in the context of HIN which is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a set of relations $\mathcal{V}$ and a link set $\mathcal{E}$. A HIN is also associated with an object type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{O}$ and a link type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$. $\mathcal{O}$ and $\mathcal{R}$ denote the set of predefined object and link types, where $|\mathcal{O}| + |\mathcal{R}| > 2$. The MOOC data in our study can be represented as a HIN. The HIN consists of six types of entities such as user, concept, video, course, school, and teacher. In addition, there is a set of links describing the relationships among those entities. On top of the definition of HIN, the concept of network schema is used to describe the meta structure of a network [31].

The network schema [35] is denoted as $\mathcal{S} = (\mathcal{O}, \mathcal{R})$. It is a meta template for an information network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the object type mapping $\phi : \mathcal{V} \rightarrow \mathcal{O}$ and the link type mapping $\psi : \mathcal{E} \rightarrow \mathcal{R}$. $\mathcal{O}$ and $\mathcal{R}$ denote the set of predefined object and link types. Given the network schema, we can extract semantic meta-paths between a pair of entities. A meta-path can be formally defined as follows:

A meta-path [34] $MP$ is defined on a network schema $\mathcal{S} = (\mathcal{O}, \mathcal{R})$ and is denoted as a path in the form of $O_1 \xrightarrow{R_1} O_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} O_{l+1}$, which describes a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between object $O_1$ and $O_{l+1}$, where $\circ$ denotes the composition operator on relations.

### 4. PROPOSED APPROACH

In this section, we introduce our proposed approach MOOCIR (MOOC Interest Recommender) based on meta-paths in the MOOC HIN. In high level, our approach extends the matrix factorization (MF) $\hat{y}_{u,c} = \mathbf{x}_u^\top \mathbf{z}_c$, where $\hat{y}_{u,c}$ denotes the predicted preference score of concept $c$ regarding user $u$, and $\mathbf{x}_u$ and $\mathbf{z}_c$ refer to latent features of $u$ and $c$, respectively. We extend the MF with user (concept) representations/embeddings that are learned by applying GCNs to meta-path-based graphs. Fig. 2 shows an overview of our approach, which consists of four main components. In the following, we describe each component in detail.

<table>
<thead>
<tr>
<th>Type</th>
<th>Meta-path</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>user → concept ( \xrightarrow{-1} ) user</td>
</tr>
<tr>
<td></td>
<td>user → course ( \xrightarrow{-1} ) user</td>
</tr>
<tr>
<td></td>
<td>user → video ( \xrightarrow{-1} ) user</td>
</tr>
<tr>
<td></td>
<td>user → course → teacher ( \xrightarrow{-1} ) course ( \xrightarrow{-1} ) user</td>
</tr>
<tr>
<td>Concept</td>
<td>concept → user ( \xrightarrow{-1} ) concept</td>
</tr>
<tr>
<td></td>
<td>concept → course ( \xrightarrow{-1} ) concept</td>
</tr>
</tbody>
</table>

#### Table 1: Meta-paths selected for extracting user-user and concept-concept relationships.

**Meta-path selection.** As discussed in Section 3, meta-paths provide the capability to derive entity-entity relationships through those paths. Similar to previous studies [13, 31], we consider user-user and concept-concept relationships via different meta-paths. To fairly compare with [31] in our experiments, we use the same set of meta-paths used in [31] for our study. Table 1 summarizes six meta-paths used in our work where four paths for users and two for concepts. For each meta-path, a homogeneous graph with respect to users (concepts) can be extracted, which is depicted as its corresponding adjacency matrix in Fig. 2. As one might expect, each entry in the adjacency matrix $\mathbf{A}$ regarding a meta-path is equal to one if two users (concepts) can be connected via that meta-path, and zero otherwise. Afterwards, we can learn user (concepts) representations for each meta-path using GCNs.

**Graph Convolution Networks (GCNs).** GCNs learn node representations of a graph by inspecting neighboring nodes. In this work, we adopt the following layer-wise propagation rule to learn user (concept) representations/embeddings with respect to a meta-path.

$$ h^{(l+1)} = g(\Phi h^{(l)} \mathbf{W}^l) $$

(1)

where $g(\cdot)$ is an activation function which we use ReLu [25] here. $\mathbf{P} = \mathbf{D}^{-1} \mathbf{A}$ where $\mathbf{D}$ is the diagonal node degree matrix of $\mathbf{A}$ to normalize the matrix $\mathbf{A}$, and $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is an adjacency matrix with self-loops in a graph based on a specific meta-path. $\mathbf{W}^l$ refers to a trainable weight matrix at layer $l$ for all nodes. $h^{(l)}$ can be fed with features of each node if there is a set of features for each node or can be initialized and learned afterwards as well. The output representation of the last layer can be used as user (concept) representations. For example, when $l = 2$, the representation of a user $u$ for a meta-path $MP_l$ will be $h^{3}_{u,MP_l} = e^{MP_l}_{u,MP_l} = h^{1}_{u,MP_l}$ where $h^{3}_{u,MP_l}$ is the output of the last layer of GCNs for the meta-path $MP_l$, with respect to $u$. In our study, we use a single layer GCN where $h^0$ is initialized randomly and learned during the training process, but one can easily extend it with more
layers or using existing features for \( h^0 \).

**Attention.** Attention mechanism [36] is motivated by how we pay visual attention to different regions of an image or relevant words in one sentence, and has been used widely to advance various fields such as natural language processing and recommender systems [37][13]. In our context, different meta-paths can have different importance with respect to each user, and incorporating the importance of each meta-path differently for each user can be beneficial when aggregating user representations from different meta-paths, i.e., \( \{e_u^{MP_1}, \ldots, e_u^{MP_{|MP|}}\} \rightarrow e_u \). In our work, we apply the attention mechanism from [6] for our context as follows:

\[
\alpha_u^{MP_i} = \frac{\exp(V_u^T \sigma(W_u e_u^{MP_i}))}{\sum_{j \in |MP|} \exp(V_u^T \sigma(W_u e_u^{MP_j}))}
\]

(2)

where the output \( \alpha_u^{MP_i} \) indicates the weight (or importance) for \( e_u^{MP_i} \), and \( V_u \) and \( W_u \) are trainable matrices for users. The attention mechanism can be formulated in the same manner for concepts. Next, the user representations coming from different meta-paths can be aggregated as follows:

\[
e_u = \sum_{j \in |MP|} \alpha_u^{MP_j} e_u^{MP_j}
\]

(3)

The above-mentioned attention mechanism takes into account different meta-paths but does not consider any context in the extended MP, which can be the latent features of users and concepts for MF. Therefore, we also investigate the following attention mechanism which considers the latent features of user \( x_u \), which has not been explored in previous studies. In this case, a meta-path based embedding \( e_u^{MP_i} \) and \( x_u \) are concatenated together when calculating the attention scores as follows.

\[
\beta_u^{MP_i} = \frac{\exp(V_u^T \sigma(W_u [e_u^{MP_i} ; f(x_u)]))}{\sum_{j \in |MP|} \exp(V_u^T \sigma(W_u [e_u^{MP_j} ; f(x_u)]))}
\]

(4)

where \( f(x_u) \) applies non-linearity with a single layer feed-forward neural networks to \( x_u \) instead of using it directly, which is inspired by [31] where the authors showed that non-linear fusion is required when combining latent features from matrix factorization and entity embeddings from GCNs. Afterwards, the final user representation can be obtained in the same manner as Eq. 5:

\[
e_u = \sum_{j \in |MP|} \beta_u^{MP_j} e_u^{MP_j}
\]

(5)

The attention mechanism can be formulated in the same manner for concepts.

**Prediction.** Given those learned user and concept representations/embeddings \( e_u \) and \( e_c \). The preference score of a concept \( c \) for a user \( u \) can be calculated as follows by extending the matrix factorization framework:

\[
\hat{y}_{u,c} = x_u^T z_c + \gamma \cdot e_u^T M e_c + b_c
\]

(6)

where \( \hat{y}_{u,c} \) is the preference score, \( x_u \) and \( z_c \) are the latent features for the matrix factorization, and \( b_c \) is a bias term. In addition, \( M \) is a trainable matrix to let \( e_u \) in the same space with \( e_c \), and \( \gamma \) is a trainable parameter for the trade-off between the prediction scores from matrix factorization and the user and concept embeddings.

### 4.1 Training Details

**Loss function.** We use the Bayesian Personalized Ranking (BPR) [29] which has been widely used for recommender systems with implicit feedback [7][5][27]. The intuition behind BPR is that a learned concept for a user should be ranked higher (with a higher score) compared to a random one in the list of concepts with which the user has not interacted, which can be formulated as follows:

\[
\mathcal{L} = \sum_{(u,i,j) \in D_u} -\ln(\sigma(\hat{y}_{u,i} - \hat{y}_{u,j})) + \lambda \parallel \Theta \parallel^2
\]

(7)

where \( (u,i,j) \) refers to a triplet including a user \( u \), an interacted concept \( i \) and an unknown concept \( j \) for the user. \( \hat{y}_{u,i} - \hat{y}_{u,j} \) measures the preference difference between the interacted concept and the unknown one, \( \sigma \) denotes the sigmoid function: \( \sigma(x) = \frac{1}{1 + e^{-x}} \), \( \lambda \) is the regularization parameter for the \( L_2 \) norm, and \( \Theta \) denotes the set of parameters to be learned. The training set \( D_u \) can be constructed by paring an unknown concept randomly with an interacted concept in the training set of a user.

To learn the parameters of our proposed approach for minimizing the loss in Eq. 7 we use a mini-batch gradient decent with 1,024 as the batch size, and use the Adam update rule [20] to train the model using the training set. In addition, the learning rate is set as 0.01, the regularization parameter \( \lambda \) is set as 1e-8, and the dimension of latent features for MF and that of user (concept) embeddings are set as 30 and 100 respectively as in [13].

To overcome the overfitting problem, we further construct a validation set by using the last interacted concept for each user, and randomly pair each known concept with 99 unknown concepts. We run 500 epochs where the convergence is observed, and monitor the performance of evaluation metrics (see Section 5) on the validation set. At the end, we choose the best-performing model on the validation set in terms of MRR (Mean Reciprocal Rank), which is one of the evaluation metrics measuring how well a ground truth concept is ranked in the corresponding set of 100 concepts. Any other evaluation metric can be used for choosing the best-performing model as well based on the preference for a specific metric.

### 5. EXPERIMENTAL SETUP

**MOOC Dataset.** We use the MOOCCube dataset [39] from the XuetangX platform for our experiments. The MOOCCube dataset is one of the largest and comprehensive MOOC datasets, and provides rich information about MOOCs and user activities on the platform from 2017 to 2019 [39]. Each course or video has a set of covered knowledge concepts in the dataset. In this work, we use user activities from 2017-01-01 to 2019-10-31 for training and those from 2019-11-01 to 2019-12-31 for testing. We limit users who have learned concepts in both training and testing periods and have at least one new concept (which did not appear in the training...
which only uses meta-path-based user and concept representations for predicting the preference score of a concept.

Next, we compare MOOCIR with the following baselines and state-of-the-art methods to evaluate the performance of recommending knowledge concepts for users. TopPop is a straightforward baseline method which ranks concepts based on their popularity. Here, the popularity of a concept can be measured based on the number of users that have learned the concept. MFBPR [29] is a matrix factorization approach which optimizes a pairwise ranking loss for the recommendation task as our approach but without meta-path-based representation learning. That is, the second component in Eq. 6 based on user (concept) representations is removed. FISM [17] is an item-to-item collaborative filtering approach which provides recommendations based on the average embeddings of all interacted concepts and the embeddings of the target concept. NAIS [15] is also an item-to-item collaborative filtering approach, but with an attention mechanism, which is capable of distinguishing which historical items in a user profile are more important for a prediction. We use the author’s implementation for both NAIS and FISM metapath2vec [11]. metapath2vec is a meta-path-based representation learning model which leverages meta-path-based random walks to construct the heterogeneous neighborhood of a node and then leverages a heterogeneous skip-gram model to learn node embeddings. We use the StellarGraph [10] implementation of metapath2vec for our experiment in which the parameters of metapath2vec are set the same as in [11] except the number of random walks is set as 500 instead of 1000. AckRec [13] also models the MOOC dataset as a HIN and extracts user (concept) representations from the same set of meta-paths in Table 1. However, AckRec treats the problem as rating prediction task where the rating of a concept for a user is the number of interactions between the user and the concept. Also, it exploits user and concept representations as features while extending the matrix factorization framework. We use the author’s implementation for our experiments. MFBPR and those MOOCIR variants are implemented using Tensorflow [1]. All experiments are run on an Intel(R) Core(TM) i5-8350U processor laptop with 16GB RAM, and MOOCIR variants take less than two days for training.

6. RESULTS

Table 3 summarizes the results using the variants of MOOCIR. As we can see from the table, MOOCIR$_{att}$ — which uses user and concept representations learned based on meta-paths with the HIN but without the matrix factorization component — provides worse performance compared to the other variants. The results indicate that extending the matrix factorization is necessary for MOOCIR.

Next, we compare MOOCIR$_c$ and the variants with attention mechanisms (i.e., MOOCIR$_{att}$ and MOOCIR$_{att2}$). We observe that both MOOCIR$_c$ and MOOCIR$_{att2}$ outperform MOOCIR$_c$, in terms of all evaluation metrics, which shows that using attention (Eq. 6) boosts the performance significantly.

https://github.com/JockWang/ACKRec

https://github.com/AaronRee/Neural-Attentive-Item-Similarity-Model

We noticed that using 1000 random walks took more than 10 days for training and did not improve the performance compared to using 500.

Table 2: Statistics of the MOOCCube dataset for experiments.

<table>
<thead>
<tr>
<th>Entities</th>
<th>Statistics</th>
<th>Relations</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>users</td>
<td>2,005</td>
<td>user-concept</td>
<td>930,553</td>
</tr>
<tr>
<td>concepts</td>
<td>21,037</td>
<td>user-course</td>
<td>13,696</td>
</tr>
<tr>
<td>courses</td>
<td>600</td>
<td>course-video</td>
<td>42,117</td>
</tr>
<tr>
<td>videos</td>
<td>22,403</td>
<td>teacher-course</td>
<td>1,875</td>
</tr>
<tr>
<td>schools</td>
<td>137</td>
<td>video-concept</td>
<td>295,475</td>
</tr>
<tr>
<td>teachers</td>
<td>138</td>
<td>course-concept</td>
<td>150,811</td>
</tr>
</tbody>
</table>
The attention mechanism is a non-trivial task, and other approaches should be investigated in the future.

Overall, MOOCIRa1 provides the best performance among all variants. In the following, we discuss the performance of MOOCIRa1 compared with other baselines and state-of-the-art methods.

Table [5] shows the performance of MOOCIRa1 and compared methods. We first observe that all the other methods outperform TopPop which is a baseline method recommending popular concepts. For example, MOOCIRa1 and ACKRec improves MRR over TopPop 45.8% and 43.1%, respectively. Among all the compared methods in Table [5] MOOCIRa1 provides the best performance followed by ACKRec, MFPR, and metapath2vec. ACKRec performs best in terms of nDCG and MRR, and MFPR performs best in terms of HR among compared methods. In detail, a significant improvement of MOOCIRa1 over ACKRec in MRR (+1.9%), nDCG@5 (+3.1%), nDCG@10 (+4.5%), and +nDCG@20 (4.8%) can be noticed (α < 0.01). Compared to MFPR, MOOCIRa1 improves the HR scores 6.7%, 9.2%, and 9.4% when k = 5, 10, 20, respectively (α < 0.01). The two item-item CF methods (FISM and NAIS) do not perform well compared to MFPR and ACKRec. One possible explanation might be due to the sparsity of the dataset, which makes that deriving item-item similarities based on interacted users for each item is challenging and limits the performance.

Those results indicate that the proposed approach MOOCIRa1 can achieve competitive performance in terms of those evaluation metrics for top-k concept recommendations compared to the baselines and state-of-the-art methods.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we presented MOOCIR for predicting and recommending concepts that might be of users’ interest on MOOC platforms. The comparison of MOOCIR variants in Section [6] shows that extending the matrix factorization with user and concept representations learned from different meta-paths and using attention for deriving those representations play crucial roles in achieving better performance. In addition, the results compared to other baselines and state-of-the-art methods indicate that MOOCIRa1 can improve the performance of predicting and recommending concepts significantly. The comparison between the two introduced attention mechanisms (Eq. [8] and [9]) suggests that a more comprehensive approach is required while fusing the latent features of users and concepts into the attention mechanism, which will be investigated in the near future.
8. REFERENCES


