Information Sciences 654 (2024) 119815

Contents lists available at ScienceDirect

Information Sciences

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

Graph neural networks with deep mutual learning for designing multi-modal recommendation systems

Jianing Li [a,1], Chaoqun Yang [b,*,1], Guanhua Ye [c], Quoc Viet Hung Nguyen [b]

[a] School of Business Administration, Northeastern University, China
[b] Institute for Integrated and Intelligent Systems, Griffith University, Australia
[c] Research and Development Department, Deep Neural Computing Company Limited, China

ARTICLE INFO

Keywords:
Investment recommendation
Institutional investors
Deep mutual learning
Multi-modal recommendation
Knowledge distillation

ABSTRACT

Recommendation services play a pivotal role in financial decision-making and multimedia content services, as they suggest investment operations and personalized items to users, typically characterized by multi-modal features such as visual, textual, and acoustic attributes. Graph Neural Networks (GNNs), demonstrating the immense potential for graph representation learning and recommendation systems, are capable of learning user/item embeddings by taking into account the graph topological structure and the multi-modal node features. Yet, a substantial number of multi-modal recommendation studies have seemingly ignored the inherent bias among different modalities during feature fusion, consequently leading to sub-optimal embeddings for items with multi-modal features. To mitigate this issue, we propose a novel multi-modal recommendation framework that integrates GNNs with deep mutual learning techniques, termed GNNMR. GNNMR uses the mutual knowledge distillation technique to collaboratively train multiple uni-modal bipartite user-item graphs. Each GNN is trained specifically on the uni-modal user-item bipartite graph, which is separated from the original multi-modal user-item bipartite graph, to generate uni-modal embeddings. These uni-modal embeddings then act as mutual supervision signals, allowing the model to uncover and synchronize the latent semantic relationships among different modalities. Subsequently, the model can conduct inference in an ensemble manner, leveraging uni-modal embeddings from diverse modalities. Experimental results on two real-world datasets demonstrate that the proposed GNNMR outperforms other multi-modal recommendation methods in the Top-K recommendation task.

1. Introduction

Effective prediction and recommendation services can help improve the service quality of financial decision-making and multimedia content services. In the capital market, many institutional investors make equity investments in listed companies, which may involve information in the form of text, images, and sound. Analyzing such complex multi-model information is a huge challenge for retail investors. Therefore, recommending the investment trends of institutional investors to retail investors with limited atten-

* Corresponding author.
E-mail addresses: lijianing@mail.neu.edu.cn (J. Li), chaoqun.yang@griffith.edu.au (C. Yang), rex.ye@dncc.tech (G. Ye), henry.nguyen@griffith.edu.au (Q.V.H. Nguyen).

[1] First author and second author contribute equally to this work.

https://doi.org/10.1016/j.ins.2023.119815
Received 16 July 2023; Received in revised form 31 August 2023; Accepted 23 October 2023

Available online 27 October 2023
0020-0255/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

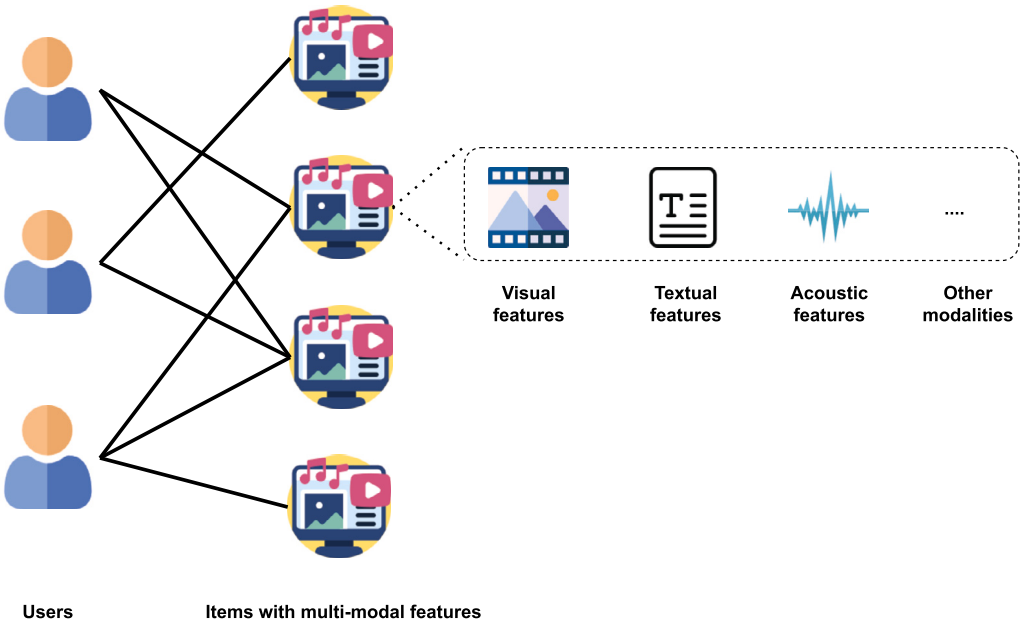

Fig. 1. Example of a user-item bipartite graph with multi-modal item features.

tion can save investment costs and improve investment efficiency. For online multimedia platforms, such as TikTok and YouTube, recommendation services play a crucial role due to the massive amount of content available, like short videos. The provision of personalized recommendations for each user can greatly enhance their user experience. Typically, these recommendation models, which are often referred to as collaborative filtering (CF)-based models, are trained over user-item bipartite graphs to learn user and item embeddings [12,4,18,42,24]. The recommendation list for each user can then be obtained by calculating the similarities between user and item embeddings. However, a major hurdle lies in the sparsity of the user-item bipartite graph, which complicates the training process for accurate user and item embeddings.

To address this issue, one potential solution involves leveraging the abundant multi-modal features of the items to mitigate the sparsity problem in interaction data, which is a strategy often associated with multi-modal recommender systems [43]. Taking the short-video recommendation as an example [35], as shown in Fig. 1, users interact with items (e.g., short videos) by clicking on the ones that pique their interest. Each item is coupled with different modality features. For example, a short video can be broken down into frames (representing visual features), titles (representing textual features), sounds (representing acoustic features), and other modal features like ID-related features. In this paper, we posit that effectively incorporating this auxiliary information (i.e., the multi-modal features of items) could enrich user and item embeddings, thereby leading to improved recommendation performance.

Over the past few years, the application of deep neural networks for learning data representations [33,32,31] has seen significant success. This is particularly true for Graph Neural Networks (GNNs) in the realm of recommender systems [37,36,38,45]. Their popularity primarily stems from their exceptional performance when dealing with graph-structured data. Notable methods such as GC-MC [3], NGCF [34], and PinSage [39] iteratively learn user/item embeddings by capturing high-order interactions between users and items, and harness their side information (e.g., features of users/items) via the message passing mechanism. Additionally, as mentioned above, item features on multimedia platforms often encompass multiple modalities. Consequently, several studies [35,22,28] have enhanced user/item embeddings by integrating diverse modality information, demonstrating superiority over methods that consider only uni-modality features. For instance, the groundbreaking VBPR work [11] boosted item node embeddings by merging visual modality features with ID-type features using a straightforward concatenation operator. There also exist GNN-based multi-modal recommendation methods like MMGCN [35] and LATTICE [40], designed to learn not only the collaborative relationships between users and items but also the semantic relationships between modality features. However, most current multi-modal recommendation methods [43] typically adopt a process that separately learns user/item embeddings for different modalities before merging them. Nevertheless, recent empirical results [43] suggest that an improper fusion of features from various modalities could introduce additional noise due to overlooking the implicit bias among these modalities, thereby leading to model performance inferior to methods that use information from a single modality alone. This phenomenon is also known as modality failure [5].

In light of the challenges at hand, and inspired by the recent successes of deep mutual learning techniques [41,9,5] in the realm of representation learning, we propose a novel multi-modal recommendation method grounded on GNN-based mutual learning, termed as GNNMR. This method is unique in its fusion-free way during the training process, thereby eliminating the need for explicit modal fusion while maintaining the ability to learn and align the semantic relationships across different modalities. More specifically, GNNMR employs the knowledge distillation technique to concurrently train several Graph Neural Networks (GNNs), exploiting the implicit semantic relationships among different modalities. The original multi-modal user-item bipartite graph is first segmented into multiple uni-modal user-item bipartite graphs based on varying item modality features. Subsequently, GNNs function as encoders,

learning user/item embeddings for different modalities within each individual uni-modal graph. This process allows various uni-modal embeddings to act as mutual supervision signals alongside ground truths, guiding the training of each GNN model and the uni-modal embeddings themselves. Ultimately, the model is capable of making predictions based on an ensemble of these uni-modal embeddings

The main contributions of this paper are summarized as follows:

- We assert that existing multi-modal recommendation methods are plagued by a modality failure issue, originating from modality fusion during the training process, which ultimately undermines the recommendation performance.
- We propose a novel multi-modal recommendation framework that integrates GNNs with deep mutual learning techniques, named GNNMR. This framework uniquely allows for the simultaneous training of multiple GNNs on uni-modal user-item bipartite graphs, thereby leveraging and sharing knowledge across different modalities for more precise recommendations.
- To validate the effectiveness of GNNMR, we have conducted an extensive series of experiments using two real-world multi-modal datasets for the Top-K recommendation task. The experimental results clearly demonstrate the superiority of GNNMR.

The rest of this paper is organized as follows. Section 2 introduces the main related works about the research problem. Section 3 provides essential definitions associated with the problem and describes the proposed GNNMR in detail. Section 4 introduces experimental settings, results, and discussions, followed by a conclusion in Section 5.

2. Related work

This section introduces the main related works including recommender systems, multi-modal recommender systems, and deep mutual learning techniques.

2.1. Recommender systems

Presently, mainstream recommendation systems primarily employ collaborative filtering (CF) methods. Such methods generally learn from the ID embeddings of users and items, drawing from user-item interaction data. The mainstays of CF-based recommendation methods encompass matrix factorization-based methods [2,14,13] and graph neural network (GNN)-based methods [36]. Matrix factorization-based methods involve decomposing the user-item interaction matrix into a lower-dimensional user matrix and item matrix, while GNN-based methods apply message-passing mechanisms in a user-item bipartite graph to identify high-order relationships between users and items. However, due to the typically sparse nature of user-item interaction data, these models might face cold start issues [8]. A prevalent solution is to incorporate side information [2] into the learning process. In this study, we concentrate on the multi-modal feature information of items.

2.2. Multi-modal recommender systems

Multi-modal feature information related to item nodes has been shown to enhance node representations, thereby improving recommendation performance. For instance, VBPR [11] integrates visual features with ID-related features by concatenating the corresponding embeddings to form the final item embeddings. On the other hand, MMGCN [35] posits that user preferences may vary across different modality features. It, therefore, proposes a modality-specific GNN model to derive user embeddings from various modality-aware bipartite user-item graphs. PMGT [21] introduces a pre-training GNN framework that learns item embeddings by taking into account both multi-modal item features and their relationships, such as co-purchasing. Meanwhile, MKGAT [28] integrates the multi-modal knowledge graph into the recommender system, enabling the explicit learning of relations among different types of entities and the aggregation of feature information from neighboring entities through an information propagation mechanism. MGAT [29], AMVAE [17], and A^2CMHNE [16] implement various attention-based strategies to capture the influence of different modalities. CER [6] introduces a collaborative embedding regression model that is able to incorporate multi-modal features with user-item interactions via the proposed priority-based late fusion method. However, most prior work on multi-modal recommendations overlooks the implicit bias between different modalities during feature fusion, resulting in sub-optimal embeddings for items with multi-modal features.

2.3. Deep mutual learning

Knowledge distillation, as proposed in [15], aims to transfer knowledge from a powerful, larger *teacher model* to a more compact *student model*. Mutual learning is a unique knowledge distillation framework that considers all models as both teacher and student models during the training process, promoting mutual learning among them. For instance, DML [41] and Co-distillation [1] incorporate an additional distillation loss between each pair of training models to facilitate mutual learning. ONE [46] employs a gate controller to merge the predictions of different models, while CLNN [27] introduces a hierarchical multi-branch design to enhance model diversity. KDCL [9], on the other hand, suggests the dynamic generation of soft labels to improve all models. Despite these advances, few studies have leveraged deep mutual learning techniques to tackle multi-modal recommendation issues. To address this gap, our paper explores the application of the deep graph mutual learning model to the multi-modal recommendation problem.

Table 1
Description of main mathematical symbols in this paper.

Symbol	Description
$\mathcal{G}$	Multi-modal user-item bipartite graph.
V	The set of nodes in the user-item bipartite graph.
u	The user node.
i	The item node.
V_{user}	The set of user nodes.
V_{item}	The set of item nodes.
E	The set of edges in the user-item bipartite graph.
$\mathbf{F}_i$	The multi-modal features associated with item node i.
$\mathbf{f}_i^m$	The feature vector of item node i in the m-th modality.
d_m	The feature dimension of the m-th modality.
$\|M\|$	The number of modalities.
E_{miss}	The latent users' preferences for items.
$f_{tea}(\cdot)$	The teacher model in the knowledge distillation learning.
$f_{stu}(\cdot)$	The student model in the knowledge distillation learning.
y	The ground truth.
y_{stu}	The prediction from the student model.
y_{tea}	The prediction from the teach model.
$\mathcal{L}_{sup}$	The supervised loss.
$\mathcal{L}_{KD}$	The knowledge distillation loss.
λ	The weight parameter associated with the knowledge distillation loss.
l	The layer index of the GNN model.
$\mathbf{h}_{u-m}^{l+1}$	The embedding of user u at $(l+1)$ layer with modality m.
$\mathbf{h}_{i-m}^{l+1}$	The embedding of item i at $(l+1)$ layer with modality m.
d	The dimension of embeddings.
$\sigma(\cdot)$	The activation function.
$\mathcal{N}(u)$	The first order neighbor nodes of user u.
$\mathcal{N}(i)$	The first order neighbor nodes of item i.
W_m^l	The trainable parameters in the l-th layer of GNN for the modality m.
s_{ui}^m	The similarity score between user u and item i under modality m.

3. GNNMR

3.1. Preliminary

Before we delve into the details of our proposed GNNMR, we first clarify a few key concepts related to our research problem. Furthermore, for ease of understanding, we have compiled a summary of all the symbols used in this paper and presented them in Table 1.

- **Multi-modal user-item bipartite graph** is represented as a graph, denoted by $\mathcal{G} = (V, E)$. In this representation, the set of nodes V consists of two types of nodes, namely the user nodes (denoted by V_{user}) and the item nodes (denoted by V_{item}). Thus, $V = \{V_{user}, V_{item}\}$. Each individual user is represented by a node $u \in V_{user}$, while each item is represented by a node $i \in V_{item}$. The set of edges E within the graph, where $E \subseteq V_{user} \times V_{item}$, signifies the interactions between the users and items. In addition to this, each item node i is linked with multi-modal features, given by $\mathbf{F}_i = \{\mathbf{f}_i^1, ..., \mathbf{f}_i^m, ..., \mathbf{f}_i^{|M|}\}$. Here, the symbol $\mathbf{f}_i^m \in \mathbb{R}^{d_m}$ characterizes the feature vector of item node i within the m-th modality and has a feature dimension of d_m. Lastly, $|M|$ represents the total number of modalities.
- **Multi-modal recommendation systems** aim to predict the latent preferences of users towards various items within a given multi-modal user-item bipartite graph. More specifically, the primary objective of these systems is to predict the missing links, denoted as $E_{miss} \subseteq V_{user} \times V_{item}$, in the multi-modal user-item bipartite graph $\mathcal{G} = (V, E)$ as defined earlier.
- **Graph neural networks** are specific types of neural network models used to handle the graph data, such as the multi-modal user-item bipartite graph described above. The core idea of a graph neural network is to learn the representation of nodes based on the topology structure of the graph and features of the nodes via the message-passing mechanism [7,37]. Specifically, let v represent the target node, then we can calculate its representation $\mathbf{h}_v$ by the following formula:

$$\mathbf{h}_v^{(l+1)} = Update(\mathbf{h}_v^{(l)}, Agg(\{\mathbf{h}_v^{(l)}, \mathbf{h}_j^{(l)}, \forall j \in \mathcal{N}(v)\})) \tag{1}$$

  where $\mathcal{N}(v)$ represents the set of neighbor nodes of the target node v, and $\mathbf{h}_j^{(l)}$ represents the representation of the neighbor node in the l-th layer of GNN. Furthermore, $Agg(\cdot)$ and $Update(\cdot)$ are the aggregation function for aggregating features of neighbor nodes and the update function for updating the representation of the target node, respectively. It is worth noting that the main different variants of graph neural networks lie in the use of different $Agg(\cdot)$ and $Update(\cdot)$ functions [44].
- **Knowledge distillation** [15] is a learning paradigm in which knowledge is distilled from a pre-trained teacher model, denoted as $f_{tea}(\Theta_{tea}^*)$, to a student model, denoted as $f_{stu}(\Theta_{stu})$, where Θ_{tea}^* and Θ_{stu} are pre-trained parameters for the teacher model

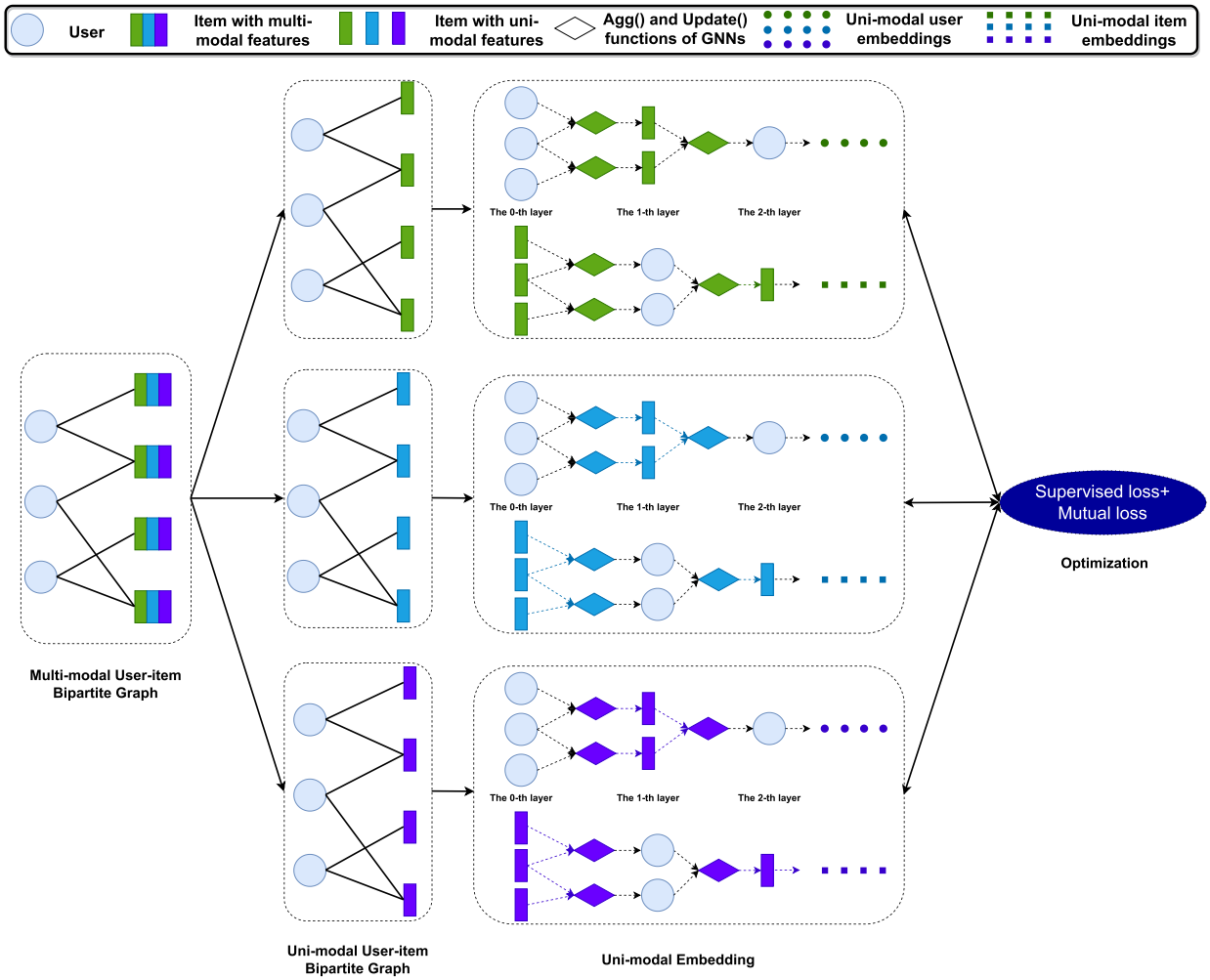

Fig. 2. The overview of the proposed GNNMR.

and the trainable parameters for the student model, respectively. This paradigm has proven effective in enhancing the student model's performance. Specifically, the conventional loss for the targeted student model is defined as follows:

$$\Theta^*_{stu} = \arg\min_{\Theta_{stu}} \mathcal{L}_{sup}(y, f_{stu}(\mathbf{x}, \Theta_{stu})) + \lambda \mathcal{L}_{KD}(f_{tea}(\mathbf{x}, \Theta^*_{tea}), f_{stu}(\mathbf{x}, \Theta_{stu})) \tag{2}$$

Here, $\mathbf{x}$ represents the training sample. $\mathcal{L}_{sup}$ is the supervised loss function (such as the cross-entropy loss in classification tasks) which measures the discrepancy between the student model's prediction $f_{stu}(\mathbf{x}, \Theta_{stu})$ with respect to the training sample $\mathbf{x}$, and the corresponding ground truth, y. On the other hand, $\mathcal{L}_{KD}$ is the knowledge distillation loss, aiming to transfer knowledge from the pre-trained teacher model to the student model using a weight parameter λ. It is worth mentioning that the deep mutual learning method [41] represents a unique knowledge distillation framework, which does not differentiate between teacher and student models, but rather enables models to learn reciprocally.

3.2. Proposed method

In this section, we present a detailed introduction to our proposed GNNMR, the schematic of which is depicted in Fig. 2. GNNMR is composed of two main phases: the uni-modal embedding phase, and the mutual learning phase. In the first phase, unimodal embeddings are learned by extracting user/item embeddings from separate unimodal user-item bipartite graphs, which are derived from the original multimodal user-item bipartite graph. The mutual learning phase, constituting the second phase, seeks to identify and align the semantic relationships among different modalities. Algorithm 1 provides a summary of the steps involved in this proposed method.

Algorithm 1 GNNMR.

Require: Multi-modal user-item bipartite graph $\mathcal{G} = (V, E)$
Ensure: The user embedding $\mathbf{h}_u$ and the item embedding $\mathbf{h}_i$
1: **for** $epoch \leftarrow 0, 1, \ldots$ **do**
2: **for** $m \leftarrow 1$ to M **do**
3: $\mathbf{h}_{u-m} \leftarrow$ Eq. (3)
4: $\mathbf{h}_{i-m} \leftarrow$ Eq. (4)
5: $s_{ui}^m \leftarrow$ Eq. (5)
6: **end for**
7: **for** $m \leftarrow 1$ to M **do**
8: Gradient descent $\leftarrow$ Eq. (8)
9: **end for**
10: **end for**

3.2.1. Uni-modal embedding

Given that item node features encompass multiple modalities across different feature spaces, we initially disassemble the original multi-modal user-item bipartite graph into several uni-modal user-item bipartite graphs. This step allows us to directly employ various readily available graph neural network variants[2] for obtaining user/item embeddings. In this study, to differentiate the impacts of various neighboring nodes, we harness the Graph Attention Network (GAT) [30] to learn the user/item embeddings from these uni-modal user-item bipartite graphs, respectively.

- **Uni-modal user embedding**. For each uni-modal user-item bipartite graph, each user node is connected to only the nodes of the item that the user has interacted with, and each item node is only associated with uni-modal features. In this way, we can capture the preferences of users in this modal space by aggregating the uni-modal feature information of the item nodes that users have interacted with (line 3 in Algorithm 1). Specifically, we could calculate the $(l+1)$-th layer embedding $\mathbf{h}_{u-m}^{(l+1)} \in \mathbb{R}^d$ of the user u from the item space (i.e., the uni-modal user-item graph) with modality m by GAT [30] as follows:

$$\mathbf{h}_{u-m}^{(l+1)} = \sigma\Big(\sum_{i \in \mathcal{N}(u)}^{\mathcal{N}(u)} \alpha_{ui} W_m^{(l)} \mathbf{h}_{i-m}^{(l)}\Big) \tag{3}$$

where $i \in \mathcal{N}(u)$ is the first-order item neighbor node of user u. $W_m^{(l)}$ contains the l-th layer trainable parameters, and $\mathbf{h}_{i-m}^{(l)}$ is the embedding of the item i with modality m in the l-th layer, and $\mathbf{h}_{i-m}^{(0)} = \mathbf{f}_i^m$. d is the dimension of embeddings, and $\sigma(\cdot)$ is the activation function. α_{ui} is the attention coefficient which can be used to distinguish the degree of a user's preference for different items they have interacted with (line 4 in Algorithm 1), and can be calculated as follows [30].

$$\alpha_{ui} = \frac{exp(\sigma(W_a^T(\mathbf{h}_{u-m}^{(l)} || \mathbf{h}_{i-m}^{(l)}))}{\sum_{i \in \mathcal{N}(u)}^{\mathcal{N}(u)} exp(\sigma(W_a^T(\mathbf{h}_{u-m}^{(l)} || \mathbf{h}_{i-m}^{(l)}))} \tag{4}$$

where $W_a \in \mathbb{R}^{2d}$ contains the trainable parameters. $\cdot^T$ and $||$ represent the transposition operation and concatenation operation respectively.

- **Uni-modal item embedding.** Similarly, each item node in the uni-modal user-item bipartite graph is connected only to the users who have interacted with it. In this way, we can learn the representation of each item in the given modality by aggregating the features of the users who have interacted with it (line 5 in Algorithm 1). Specifically, the $(l+1)$-th layer embedding $\mathbf{h}_{i-m}^{(l+1)} \in \mathbb{R}^d$ of the item i on the uni-modal user-item bipartite graph with modality m can be calculated by GAT as follows [30]:

$$\mathbf{h}_{i-m}^{(l+1)} = \sigma\Big(\sum_{j \in \mathcal{N}(i)}^{\mathcal{N}(i)} \alpha_{ij} W_m^{(l)} \mathbf{h}_{j-m}^{(l)}\Big) \tag{5}$$

where $j \in \mathcal{N}(i)$ is the first-order user neighbor node of item i. α_{ij} is the attention coefficient calculated by (4). $W_m^{(l)}$ contains the l-th layer trainable parameters, and $\mathbf{h}_{j-m}^{(l)}$ is the embedding of the user j on modality m.
Finally, the user/item embeddings on the other modalities can be obtained in the same way with different GNN trainable parameters.

3.2.2. Deep mutual learning

Recent empirical results [43,5] have indicated that the improper fusion of features from different modalities may introduce additional noise, as it often overlooks the implicit bias between different modalities. This oversight can lead to model performance that is inferior to that achieved when using information from a single modality alone. To alleviate this issue, we propose a fusion-free method for the training process inspired by the deep mutual learning technique [41]. This approach does not necessitate explicit modal fusion, yet it remains capable of learning and aligning the semantic relationships between different modalities. As previously

[2] It's worth noting that our proposed framework is model-agnostic, implying that the proposed framework can be seamlessly integrated with other GNN or GNN-based recommendation methods.

mentioned, for a uni-modal user-item bipartite graph with modality m, we employ a modal-specific GNN model as an encoder. This enables the acquisition of the user embedding $\mathbf{h}_{u-m}$ and item embedding $\mathbf{h}_{i-m}$. Consequently, the similarity score s_{ui}^{m} between the user u and item i within modality m can be calculated as follows:

$$s_{ui}^{m} = \mathbf{h}_{u-m}^{T}\mathbf{h}_{i-m} \tag{6}$$

In this way, regarding the Top-K recommendation task, the loss function for a model with the m modality comprises two components: the supervised term and the mutual learning term.

- **Supervised term** $\mathcal{L}_{BPR}$ aims to measure the inconsistency between the model predictions and the ground truth. To this end, we utilize Bayesian Personalized Ranking (BPR) [26] as the supervised term for the modality m as follows:

$$\mathcal{L}_{BPR} = \sum_{(u,i,i')} -In\mu(\mathbf{h}_{u-m}^{T}\mathbf{h}_{i-m} - \mathbf{h}_{u-m}^{T}\mathbf{h}'_{i-m}) \tag{7}$$

  where $\mu(\cdot)$ is the sigmoid function, and i' represents the items that users have not interacted with.
- **Mutual learning term** $\mathcal{L}_{ML}$ aims to align the predictions of modality m and predictions of other modalities. Intuitively, different modal features should have consistent predictive results. Thus, for each modality m, we take the ensemble of all the other $M-1$ modalities as the teacher. Thus, we adopt the Kullback Leibler (KL) Divergence to measure the distance from the predictions s_{ui}^{m} of modality m to predictions s_{ui}^{oth} of other modalities as follows:

$$\mathcal{L}_{ML} = D_{KL}(s_{ui}^{oth}||s_{ui}^{m}) = \sum_{(u,i)} s_{ui}^{oth} log\frac{s_{ui}^{oth}}{s_{ui}^{m}}, \quad s_{ui}^{oth} = \frac{1}{|M|-1}\sum_{c=1,c\neq m}^{|M|} s_{ui}^{c} \tag{8}$$

  Finally, the final loss function $\mathcal{L}_{m}$ for the model under the modality m is defined as follows:

$$\mathcal{L}_{m} = \mathcal{L}_{BPR} + \lambda\mathcal{L}_{ML} \tag{9}$$

  where λ is a hyperparameter to control the mutual learning term.
- **Optimization** The process of mutual learning is performed throughout the model training. Specifically, at each round of iteration, the models under each modality independently compute the uni-modal user/item embeddings, and then update the parameters of corresponding GNN models and user/item embeddings successively based on formula (8) until the models converge.

3.2.3. Prediction

After training, the model can perform inference in a uni-modal embedding ensemble manner. Here, we introduce two commonly used uni-modal ensemble methods including concatenation and average.

- **Concatenation ensemble (GNNMR-con)** is the simplest ensemble method which obtains the final multi-modal user/item embeddings $\mathbf{h}_{u}$ and $\mathbf{h}_{i}$ by concatenation operation || as follows:

$$\mathbf{h}_{u} = ||_{m=1}^{M}\mathbf{h}_{u-m}, \mathbf{h}_{i} = ||_{m=1}^{M}\mathbf{h}_{i-m} \tag{10}$$

- **Average ensemble (GNNMR-ave)** hypothesizes that the different modalities have the same importance, which requires the different modality embeddings with the same dimension and could be calculated as follows:

$$\mathbf{h}_{u} = \frac{1}{|M|}(\sum_{m=1}^{M}\mathbf{h}_{u-m}), \quad \mathbf{h}_{i} = \frac{1}{|M|}(\sum_{m=1}^{M}\mathbf{h}_{i-m}) \tag{11}$$

4. Performance evaluation

To validate the effectiveness of the proposed method, we conduct extensive experiments on two real-world multi-modal bipartite graph datasets to answer the following research questions (RQs):

- **RQ1:** How is the performance of GNNMR compared with other multi-modal recommendation methods?
- **RQ2:** Does mutual learning alleviate the problem of modality failure [5]?
- **RQ3:** How do the different hyperparameters (e.g., embedding dimension) affect the performance of GNNMR?
- **RQ4:** Can the proposed method be effective in real-world application scenarios?

Table 2
The statistic information of the datasets.

Dataset		Fund	Steam
#node	#user	165	63347
	#item	5098	15864
#edge	#user-item	791025	750050
#modalities	user	1	1
	item	19	2

4.1. Experimental setup

4.1.1. Datasets

We validate the effectiveness of the proposed GNNMR on two real-world multi-modal bipartite graph datasets including one public dataset on Steam [23] and one private dataset on investment funds named Fund. The statistical information of both datasets is summarized in Table 2.

- **Steam:** Steam dataset [23] is a gaming purchasing dataset by crawling users from the Australian Steam community. It consists of users, items (i.e., games with posters and textual descriptions), and user-item interactions (i.e., purchases). We randomly select negative samples (i.e., games not purchased by users.) with an equal number of positive samples (i.e., games purchased by users.) to construct training triplets. Furthermore, we use pre-trained ResNet18 [10] to extract visual features from posters of games, and the pre-trained SentenseBert [25] is utilized to extract textual features from game descriptions. The dataset is randomly split in 8:1:1 for training, validation, and testing.
- **Fund:** Fund is a dataset of investment data for fund company-publicly traded company. In the context of the Fund dataset, the 'user node' represents the fund company, while the 'item node' represents the publicly traded company. The edges between 'user-item' represent the fund company's investment in that particular publicly traded company. In this dataset, we consider the 'user' ID as a unique feature of the user node, while the 19 different domain features associated with the 'item node' are viewed as features of different modalities. For each modality, the one-hot vector is utilized as to encode the corresponding feature. The dataset is randomly split in 8:1:1 for training, validation, and testing.

4.1.2. Comparison method

We compare the proposed GNNMR with the following multi-modal recommendation methods:

- **GAT [30]** is the representative GNN method with the attention mechanism, which is a pure collaborative filtering method that does not unitize any side information.
- **VBPR [11]** is the multi-modal recommendation method concatenating other modality features with ID-related features.
- **PMGT [20]** a pre-training GNN framework to learn item embeddings by considering both multi-modal item features and their relationships such as co-purchase.
- **A^{2}CMHNE [16]** employ various attention-based strategies to capture the influence of the different modalities.
- **MMGCN [35]** is a multi-modal recommendation method that could capture the user preference on different modalities.

4.1.3. Parameters settings

The hyperparameters of our proposed GNNMR are set as follows. We utilize Adam SGD optimizer [19] and LeakyReLU as the activation function (i.e., $\sigma(\cdot) = LeakyReLU(\cdot)$) throughout the experiments. For each uni-modal user-item bipartite graph, the two-layers GATs are utilized to learn user/item embeddings with the embedding size $d = 128$. The batch size is set to 256, and we search the learning rate in $\{0.00001, 0.00005, 0.0001, 0.0005, 0.001, 0.005\}$. We also search the mutual learning weight λ from $\{0.1, 0.5, 1, 2\}$.

4.1.4. Evaluation metrics

The *Recall@K* and *NDCG@K*, defined as below, are widely utilized as evaluation metrics for the Top-K recommendation task [43].

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$DCG = \sum_{i=1}^{K} \frac{2_i^{relevance} - 1}{log_2(i+1)}, \quad NDCG = \frac{DCG}{iDCG} \tag{13}$$

where TP and FN represent predicted labels True Positive and True Negative, respectively. *iDCG* is the ideally ordered recommendation list for DCG. Specifically, the *Recall* metric measures how many relevant items are successfully recommended in the recommendation list, and the *NDCG* metric considers both the relevance of recommended items and their positions in the list. Furthermore, both Recall and NDCG are metrics where higher values indicate better performance in the recommender system.

Table 3
The Top-K recommendation results for different methods with respect to Recall. The best results are marked in **bold**.

Method	Steam			Fund		
	Recall@10	Recall@20	Recall@50	Recall@10	Recall@20	Recall@50
GAT	0.2241	0.2542	0.3218	0.0509	0.0598	0.1026
VBPR	0.244	0.2796	0.3417	0.0527	0.0652	0.1102
PMGT	0.2409	0.281	0.3405	0.0503	0.0637	0.1078
A2CMHNE	0.2346	0.2746	0.3352	0.0536	0.064	0.1067
MMGCN	0.2412	0.2805	0.341	0.0509	0.0648	0.1072
GNNMR-con	0.2659	0.3028	**0.3699**	0.061	0.0702	0.1242
GNNMR-ave	**0.2667**	**0.3073**	0.3624	**0.0625**	**0.0717**	**0.1308**

Table 4
The Top-K recommendation results for different methods with respect to NDCG. The best results are marked in **bold**.

Method	Steam			Fund		
	NDCG@10	NDCG@20	NDCG@50	NDCG@10	NDCG@20	NDCG@50
GAT	0.1032	0.1352	0.2239	0.0292	0.0397	0.096
VBPR	0.1127	0.1487	0.2316	0.0306	0.0451	0.0917
PMGT	0.1096	0.1481	0.2345	0.03	0.0426	0.0915
A2CMHNE	0.1158	0.1471	0.2302	0.0314	0.0431	0.0932
MMGCN	0.1142	0.1488	0.2314	0.0311	0.0424	0.0978
GNNMR-con	0.1226	0.1527	**0.2578**	0.0405	0.0458	0.102
GNNMR-ave	**0.13**	**0.1559**	0.2552	**0.0417**	**0.0463**	**0.1187**

Table 5
The results of ablation studies with respect to *Recall@20* and *NDCG@20* on Steam.

Metrics	GNNMR w/o ML		GNNMR w ML		GNNMR
	Textual	Visual	Textual	Visual	
Recall@20	0.2845	0.2836	0.3042	0.2996	**0.3073**
NDCG@20	0.1476	0.146	0.1555	0.1551	**0.1559**

4.2. Top-K recommendation (RQ1)

To compare the performance of the proposed GNNMR with other multi-modal recommendation methods, we test them on the Top-K recommendation task with $K = 10, 20, 50$. Each algorithm is tested 5 times independently, and the average results are reported for comparison. The experimental results are summarized and given in Table 3 and Table 4. From the results, we can observe that:

- GNNMR outperforms all the comparison methods on both Fund and Steam datasets, which demonstrates the effectiveness of the proposed method. For example, GNNMR improves the second-best comparison method with respect to *NDCG@20* by 4.77% and 2.67% on Steam and Fund, respectively. We attribute the improvements to GNNMR learning superior node embeddings by exploiting the different semantic information among different modalities by mutual learning techniques.
- The average-based ensemble method (GNNMR-ave) outperforms the concatenation-based ensemble method (GNNMR-con) in most cases on Steam and Fund datasets. One possible reason is that the concatenation-based ensemble method has a larger embedding size leading to introducing additional noise.
- Generally speaking, the methods leveraging the side information such as multi-modal features could get better performance than GAT. One possible explanation is that the multi-modal feature information could alleviate the data sparse issue.

4.3. Ablation experiment (RQ2)

To explore the effects of different components of the proposed GNNMR, we conduct the ablation experiment and compare different components on the Steam dataset for the Top-K recommendation task. We use the performances of GNNMR-ave to represent the GNNMR performance due to its superiority on the Steam dataset. In particular, we first explore the effect of the mutual learning component. Specifically, we implement the model without the mutual learning loss, denoted as *GNNMR w/o ML*, and independently train two GNN models for two modalities (i.e., the textual and visual modalities). Finally, we report the model performance for different uni-modal models separately. In addition, we implement the model with the mutual learning loss but remove the uni-modal embedding ensemble component, denoted as *GNNMR w ML*. The experimental results are shown in Table 5, and we can observe that:

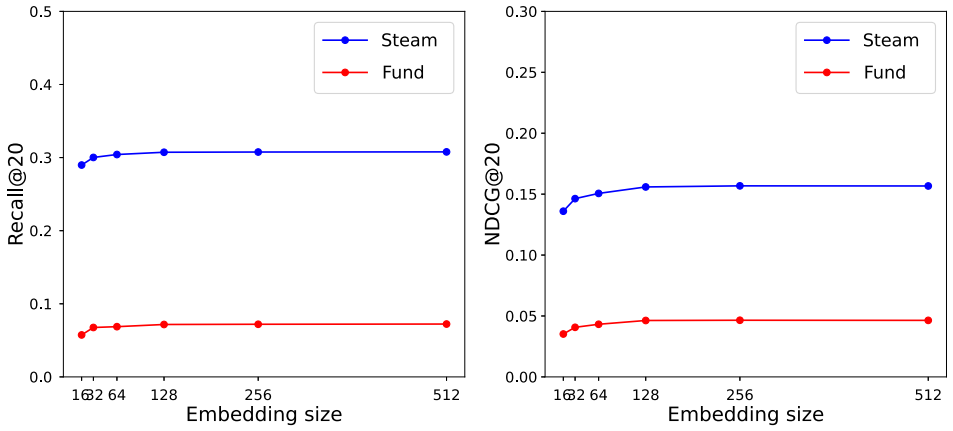

Fig. 3. Hyper-parameter sensitivity analysis of the dimension of embedding d.

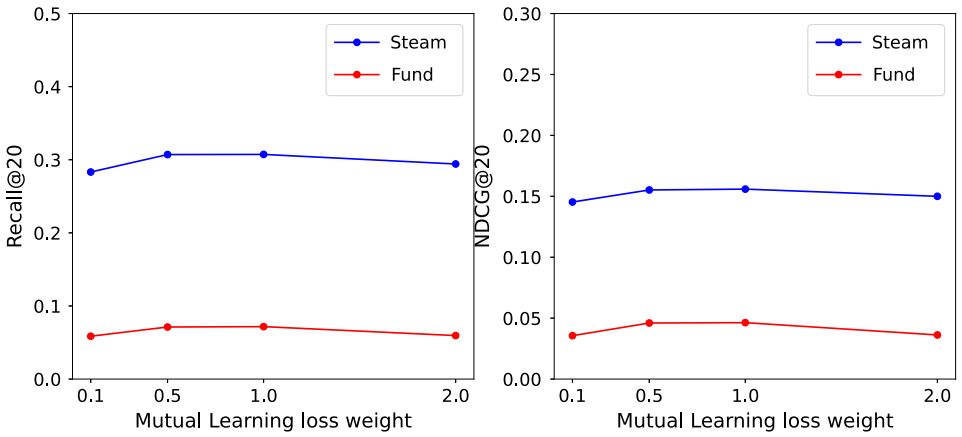

Fig. 4. Hyper-parameter sensitivity analysis of the weight of the mutual learning loss λ.

- Overall, all the components have significant effects on the proposed GNNMR, which demonstrates that all the components are essential for GNNMR.
- The different modal features have different effects on the Steam dataset. In particular, the texture features have a larger effect than visual features on the Steam dataset, which demonstrates that users of the Steam platform pay more attention to texture information than visual information.
- The mutual learning loss has significant effects on GNNMR, which shows that it is capable of well exploiting the semantic information among different modalities.
- The uni-modal embedding ensemble can achieve better performance than methods using single-modal features. It is reasonable that more feature information can better describe the user's preference.

4.4. Hyperparameter sensitivity analysis (RQ3)

To explore the effect of different hyperparameters including the dimension d of user/item embeddings and the weight λ of the mutual learning loss. We conduct experiments over two datasets for the Top-K recommendation task and report Recall@20 and NDCG@20. In particular, we explore d in $\{16, 32, 64, 128, 256, 512\}$, and λ from $\{0.1, 0.5, 1, 2\}$ respectively. The experimental results are shown in Fig. 3 and Fig. 4 respectively, and we can observe that:

- With the increasing of the dimension of embeddings, the performance increases first and then tends to be steady on both datasets. It is reasonable that the small embedding size may contain incomplete feature information resulting in bad performance.
- The proposed model achieves good performance when the weight λ of mutual learning loss lies in the range from 0.5 to 1 on both datasets. One possible reason is that the model cannot effectively share the knowledge among different modalities with a small mutual learning weight. On the other hand, for a large λ, the model would be led by mutual learning but ignore learning from the real distributions.

Fig. 5. The case study results on Fund dataset.

- Particularly speaking, the proposed GNNMR could achieve the best performance on the Top-K recommendation task with $d = 256$ and $\lambda = 0.5$ for both datasets.

4.5. *Case study (RQ4)*

To validate the proposed method for real-world multi-modal recommendation problems, we conducted experiments on the Fund dataset and provided a case study result. In order to verify that leveraging multimodal information can alleviate the cold-start problem, we focused on the fund company (i.e., CINDA SECURITIES) with fewer investment projects and provided the top-10 recommendation results in 2021. Specifically, in Fig. 5, we use boxes of different colors to represent companies belonging to different domains. The black solid line represents the ground truth, which is the actual companies invested in by the investment fund company. The black dashed line represents discrepancies between the model's predictions and the ground truth, while the red solid line represents instances where the model's predictions align with the ground truth. The experimental results are shown in Fig. 5, and we can observe that:

- The fund company primarily invests in the real estate and manufacturing industries, indicating that our method can effectively learn the multimodal semantic information among different companies and provide relevant recommendations in these two industries. For example, our model correctly recommended Gree Electric and Vanke in the manufacturing and real estate industries, respectively.

- Despite Datang Telecom Technology and Macrolink not being invested by Cinda Securities, it can be observed that our model effectively recognizes the company's investment preferences in the manufacturing and real estate industries based on modal information related to company categories.
- It's worth noting that although the fund company invests across various domains, our model can effectively capture its preferences in the different domains. One possible reason is that the mutual knowledge distillation technique we employed can effectively learn information from different modal features for the cross-domain recommendations.

5. Conclusion

In this paper, we first argue that most current multi-modal recommendation methods suffer from the modality failure issue due to overlooking the implicit modalities bias. To alleviate this issue, we propose a deep mutual learning-based fusion-free multi-modal recommendation framework, named GNNMR, which can jointly train multiple GNN models on uni-modal user-item bipartite graphs without explicit modalities fusion. To validate the effectiveness of the proposed method, extensive experiments are conducted on two real-world multi-modal graph datasets for the Top-K recommendation task, and experimental results have demonstrated that the proposed GNNMR outperforms the other multi-modal recommendation methods. Since our method can be seamlessly integrated into various late fusion methods, studying the impact of different late fusion methods on the performance of our model will be part of our future work.

CRediT authorship contribution statement

Jianing Li: Methodology, Validation, Writing – original draft. **Chaoqun Yang:** Conceptualization, Investigation, Methodology, Supervision, Visualization, Writing – review & editing. **Guanhua Ye:** Data curation, Resources, Writing – review & editing. **Quoc Viet Hung Nguyen:** Conceptualization, Investigation, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China under Grant 72202030, and the China Postdoctoral Science Foundation under Grant 2022M710630.

References

[1] Rohan Anil, Gabriel Pereyra, Alexandre Passos, Robert Ormandi, George E. Dahl, Geoffrey E. Hinton, Large scale distributed neural network training through online distillation, arXiv preprint arXiv:1804.03235, 2018.
[2] Linas Baltrunas, Bernd Ludwig, Francesco Ricci, Matrix factorization techniques for context aware recommendation, in: Proceedings of the Fifth ACM Conference on Recommender Systems, 2011, pp. 301–304.
[3] Rianne van den Berg, Thomas N. Kipf, Max Welling, Graph convolutional matrix completion, arXiv preprint arXiv:1706.02263, 2017.
[4] Tong Chen, Hongzhi Yin, Guanhua Ye, Zi Huang, Yang Wang, Meng Wang, Try this instead: personalized and interpretable substitute recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 891–900.
[5] Chenzhuang Du, Tingle Li, Yichen Liu, Zixin Wen, Tianyu Hua, Yue Wang, Hang Zhao, Improving multi-modal learning with uni-modal teachers, arXiv preprint arXiv:2106.11059, 2021.
[6] Xingzhong Du, Hongzhi Yin, Ling Chen, Yang Wang, Yi Yang, Xiaofang Zhou, Personalized video recommendation using rich contents from videos, IEEE Trans. Knowl. Data Eng. 32 (3) (2020) 492–505.
[7] Jiarui Feng, Yixin Chen, Fuhai Li, Anindya Sarkar, Muhan Zhang, How powerful are k-hop message passing graph neural networks, in: S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh (Eds.), Advances in Neural Information Processing Systems, vol. 35, Curran Associates, Inc., 2022, pp. 4776–4790.
[8] Jyotirmoy Gope, Sanjay Kumar Jain, A survey on solving cold start problem in recommender systems, in: 2017 International Conference on Computing, Communication and Automation (ICCCA), IEEE, 2017, pp. 133–138.
[9] Qiushan Guo, Xinjiang Wang, Yichao Wu, Zhipeng Yu, Ding Liang, Xiaolin Hu, Ping Luo, Online knowledge distillation via collaborative learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11020–11029.
[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[11] Ruining He, Julian McAuley, Vbpr: visual bayesian personalized ranking from implicit feedback, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2016.
[12] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, Meng Wang 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, 2020, pp. 639–648.
[13] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, Tat-Seng Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182.

[14] Xiangnan He, Hanwang Zhang, Min-Yen Kan, Tat-Seng 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, pp. 549–558.
[15] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, Distilling the knowledge in a neural network, arXiv preprint arXiv:1503.02531, 2015.
[16] Jun Hu, Shengsheng Qian, Quan Fang, Xueliang Liu, Changsheng Xu, A2cmhne: attention-aware collaborative multimodal heterogeneous network embedding, ACM Trans. Multimed. Comput. Commun. Appl. 15 (2) (2019) 1–17.
[17] Feiran Huang, Xiaoming Zhang, Chaozhuo Li, Zhoujun Li, Yueying He, Zhonghua Zhao, Multimodal network embedding via attention based multi-view variational autoencoder, in: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, 2018, pp. 108–116.
[18] Nguyen Quoc Viet Hung, Huynh Huu Viet, Nguyen Thanh Tam, Matthias Weidlich, Hongzhi Yin, Xiaofang Zhou, Computing crowd consensus with partial agreement, IEEE Trans. Knowl. Data Eng. 30 (1) (2017) 1–14.
[19] Diederik P. Kingma, Jimmy Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
[20] Yong Liu, Susen Yang, Chenyi Lei, Guoxin Wang, Haihong Tang, Juyong Zhang, Aixin Sun, Chunyan Miao, Pre-training graph transformer with multimodal side information for recommendation, arXiv preprint arXiv:2010.12284, 2020.
[21] Yong Liu, Susen Yang, Chenyi Lei, Guoxin Wang, Haihong Tang, Juyong Zhang, Aixin Sun, Chunyan Miao, Pre-training graph transformer with multimodal side information for recommendation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 2853–2861.
[22] Aditya Pal, Chantat Eksombatchai, Yitong Zhou, Bo Zhao, Charles Rosenberg, Jure Leskovec, Pinnersage: multi-modal user embedding framework for recommendations at pinterest, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 2311–2320.
[23] Apurva Pathak, Kshitiz Gupta, Julian McAuley, Generating and personalizing bundle recommendations on steam, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1073–1076.
[24] Ruihong Qiu, Zi Huang, Hongzhi Yin, Zijian Wang, Contrastive learning for representation degeneration problem in sequential recommendation, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 813–823.
[25] Nils Reimers, Iryna Gurevych, Sentence-bert: sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019, p. 11.
[26] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, Lars Schmidt-Thieme Bpr, Bayesian personalized ranking from implicit feedback, in: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI'09, AUAI Press, Arlington, Virginia, USA, 2009, pp. 452–461.
[27] Guocong Song, Wei Chai, Collaborative learning for deep neural networks, Adv. Neural Inf. Process. Syst. 31 (2018).
[28] Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, Kai Zheng, Multi-modal knowledge graphs for recommender systems, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1405–1414.
[29] Zhulin Tao, Yinwei Wei, Xiang Wang, Xiangnan He, Xianglin Huang, Tat-Seng Chua, Mgat: multimodal graph attention network for recommendation, Inf. Process. Manag. 57 (5) (2020) 102277.
[30] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, Graph attention networks, arXiv preprint arXiv:1710.10903, 2017.
[31] Dianhui Wang, Caihao Cui, Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics, Inf. Sci. 417 (2017) 55–71.
[32] Dianhui Wang, Ming Li, Robust stochastic configuration networks with kernel density estimation for uncertain data regression, Inf. Sci. 412 (2017) 210–222.
[33] Dianhui Wang, Ming Li, Stochastic configuration networks: fundamentals and algorithms, IEEE Trans. Cybern. 47 (10) (2017) 3466–3479.
[34] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, Tat-Seng Chua, Neural graph collaborative filtering, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 165–174.
[35] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, Tat-Seng Chua Mmgcn, Multi-modal graph convolution network for personalized recommendation of micro-video, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1437–1445.
[36] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, Bin Cui, Graph neural networks in recommender systems: a survey, ACM Comput. Surv. 55 (5) (2022) 1–37.
[37] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, S. Yu Philip, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2020) 4–24.
[38] Hongzhi Yin, Bin Cui, Zi Huang, Weiqing Wang, Xian Wu, Xiaofang Zhou, Joint modeling of users' interests and mobility patterns for point-of-interest recommendation, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 819–822.
[39] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, Jure Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 974–983.
[40] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, Liang Wang, Mining latent structures for multimedia recommendation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 3872–3880.
[41] Ying Zhang, Tao Xiang, Timothy M. Hospedales, Huchuan Lu, Deep mutual learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4320–4328.
[42] Bolong Zheng, Kai Zheng, Xiaokui Xiao, Han Su, Hongzhi Yin, Xiaofang Zhou, Guohui Li, Keyword-aware continuous knn query on road networks, in: 2016 IEEE 32Nd International Conference on Data Engineering (ICDE), IEEE, 2016, pp. 871–882.
[43] Hongyu Zhou, Xin Zhou, Zhiwei Zeng, Lingzi Zhang, Zhiqi Shen, A comprehensive survey on multimodal recommender systems: taxonomy, evaluation, and future directions, arXiv preprint arXiv:2302.04473, 2023.
[44] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, Maosong Sun, Graph neural networks: a review of methods and applications, AI Open 1 (2020) 57–81.
[45] Xiaofei Zhu, Gu Tang, Pengfei Wang, Chenliang Li, Jiafeng Guo, Stefan Dietze, Dynamic global structure enhanced multi-channel graph neural network for session-based recommendation, Inf. Sci. 624 (2023) 324–343.
[46] Xiatian Zhu, Shaogang Gong, et al., Knowledge distillation by on-the-fly native ensemble, Adv. Neural Inf. Process. Syst. 31 (2018).