Sentiment-aware Rating-based Recommendation via Semantic-enhanced Item Alignment

Yingjie Chen[1], Xiang Li[1], Dongliang Chen[1], Guoqing Chao[2], Zhongying Zhao[3], Yanwei Yu[1*]
[1]Faculty of Information Science and Engineering, Ocean University of China
[2]School of Computer Science and Technology, Harbin Institute of Technology
[3]College of Computer Science and Engineering, Shandong University of Science and Technology
{chenyingjie, lixiang1202}@stu.ouc.edu.cn, chendongliang@ouc.edu.cn, guoqingchao@hit.edu.cn,
zyzhao@sdust.edu.cn, yuyanwei@ouc.edu.cn

Abstract

Leveraging review texts to mine deep user preferences is vital for recommendation. However, existing methods neglect the positive-negative counteraction and rely on noisy hard sentiment thresholds. Furthermore, the feature density asymmetry causes dense semantic features to overwhelm sparse collaborative signals. To address these issues, we propose the **SENSE** model for fine-grained sentiment-aware and structurally aligned recommendation. We employ a confidence-aware soft sentiment assignment and a sentiment exchange mechanism to quantify fine-grained preferences and explicitly model sentiment interaction. For feature fusion, we construct a collaborative-semantic graph and incorporate adaptive data augmentation. Finally, via anchor-guided cross-view alignment, we use the reliable behavior view as a topological anchor to enforce semantic alignment, thereby preserving semantic richness while preventing collaborative signals from being overwhelmed. Through experiments on four datasets, we verify that SENSE outperforms state-of-the-art methods. It achieves average improvements of 19.26% in MSE for rating prediction, 13.41% in Recall@10 and 13.89% in NDCG@10 for Top-N recommendation. Our source code is available at https://github.com/yingjie160/SENSE.

1 Introduction

Review-based recommender systems aim to improve user and item representation learning by utilizing semantic information within unstructured reviews. This method has been proven useful for solving the data sparsity problem of rating-only methods and improving recommendation performance [McAuley and Leskovec, 2013; Ling *et al.*, 2014; Zhang *et al.*, 2019]. Recently, with the development of Graph Neural Networks (GNNs), integrating review semantics into the user-item interaction graph has become a common approach [Guo *et al.*, 2025; Li *et al.*, 2026b; Liu *et al.*, 2026]. These methods capture both collaborative signals and semantic connections by message passing.

Existing research has evolved through three phases. Early works like DeepCoNN [Zheng *et al.*, 2017] and NARRE [Chen *et al.*, 2018] utilize deep networks for semantic extraction, which advance into Review-aware GNNs to capture high-order collaborative signals within the graph structure. Subsequently, to mitigate data sparsity and noise, Graph Contrastive Learning like RGCL [Shuai *et al.*, 2022] and ReHCL [Wang *et al.*, 2024] is introduced to enhance view representations. Recently, models like SGDN [Ren *et al.*, 2023] and LETTER [Son *et al.*, 2025] try to separate complex preferences into different factors or identify the sentiment polarities inherent in the reviews. Although these methods show satisfactory performance, they still present three main limitations in mining the complex user decision processes and handling cross-modal fusion:

(1) Neglect of Sentiment Counteraction: Existing disentangled GNNs often treat latent factors as parallel channels by simply adding them up [Ren *et al.*, 2023]. This approach oversimplifies the decision process by ignoring the conflict between positive incentives and negative resistance. In fact, a rating is usually a trade-off where a user compares the good points of an item against its bad points.

(2) Label Noise from Hard Thresholds: Even though current works try to identify sentiment types, they mostly rely on fixed hard thresholds [Son *et al.*, 2025] or focus only on explicit click signals [Xie *et al.*, 2021]. This simple strategy overlooks the subjective bias inherent in different users and neglects mixed sentiments. As a result, using a "one-size-fits-all" rule to force preferences into two categories causes label noise and hides the real user intent.

(3) Cross-Modal Feature Misalignment: There is a clear mismatch between the dense but noisy text features and the sparse but accurate interaction features [Wang *et al.*, 2025; Zhang *et al.*, 2025a; Li *et al.*, 2026a], which creates conflicts. Direct fusion allows dense text noise to hide the precise collaborative signals while the sparse interaction graph fails to provide enough structure to guide the learning of text features. This leads to a misalignment that reduces the effectiveness of the model.

To address these problems, we propose a novel **SEN**timent-aware and **SE**mantic-Enhanced framework named SENSE. *Regarding sentiment awareness*, we propose a distribution-

*Corresponding author: Yanwei Yu.

aware soft sentiment assignment mechanism. Unlike hard thresholds, this mechanism utilizes Empirical Bayes to estimate a personalized threshold and uncertainty for each user. It changes discrete ratings into continuous soft sentiment scores. This helps us generate robust positive and negative user representations. *Regarding feature fusion*, we design a collaborative-guided semantic alignment strategy. Specifically, we build a collaborative-semantic fused graph that utilizes reliable collaborative information to adjust the sampling of semantic neighbors, combined with adaptive data augmentation to balance feature scales. Furthermore, we introduce a cross-view alignment objective by using the reliable behavior view as an anchor. This approach preserves rich semantics without obscuring the collaborative signals, ensuring that the semantic space aligns well with the behavioral space. *Regarding the interaction modeling*, we design a sentiment exchange mechanism that gathers general user preferences. It also explicitly models the conflict between positive incentives and negative resistance, ultimately predicting ratings through this refined interaction. Extensive experiments on four real-world datasets demonstrate the clear superiority of our proposed model over state-of-the-art (SOTA) baselines. The main contributions of this paper are summarized as follows:

- We propose a novel framework named SENSE, utilizing personalized soft sentiment modeling and an explicit positive-negative sentiment exchange mechanism to capture the complex user decision-making process.
- We design a collaborative-guided semantic alignment strategy with an anchor-guided cross-view contrastive learning objective, which effectively calibrates the dense semantic space using sparse collaborative signals and preserves semantic richness without overwhelming behavioral patterns.
- Extensive experiments on multiple real-world datasets are conducted to validate the effectiveness of SENSE in both rating prediction and Top-N recommendation tasks. Experimental results show that SENSE significantly outperforms state-of-the-art methods by up to 23.08%.

2 Related Work

Review-aware Graph Neural Networks. Utilizing review texts to assist user and item representation learning has become a prevalent paradigm in recommender systems [McAuley and Leskovec, 2013]. Early works primarily focus on using deep neural networks to extract semantic features from texts. For instance, DeepCoNN [Zheng *et al.*, 2017] employs parallel CNNs to extract review features for users and items respectively; TransNet [Catherine and Cohen, 2017] introduces a transformational layer to model latent relationships; NARRE [Chen *et al.*, 2018] introduces an attention mechanism to distinguish the usefulness of different reviews. With the rise of GNNs [Kipf, 2016; He *et al.*, 2020], recent research has begun to explore integrating review information into graph structures. SSG [Gao *et al.*, 2020] and HGNR [Liu *et al.*, 2020] construct heterogeneous graphs by associating review texts with edge weights to achieve deeper fusion and employ message passing mechanisms to capture high-order connectivity. *However, by converting complex review semantics into a simple scalar weight, these approaches fail to preserve the fine-grained preference information inherent in the reviews.*

Graph Contrastive Learning for Recommendation. To alleviate data sparsity and noise issues, Contrastive Learning has been widely adopted in graph-based recommendation. SGL [Wu *et al.*, 2021] and SimGCL [Yu *et al.*, 2022] establish the effectiveness of self-supervised learning in recommendation by constructing augmented views via random perturbations on graph structures. NCL [Lin *et al.*, 2022] proposes to incorporate structural neighbors and semantic prototypes into the contrastive objective. Subsequently, researchers start to leverage review data to construct more meaningful augmented views [Shi *et al.*, 2023; Fang *et al.*, 2025]; RMCL [Yang *et al.*, 2023] and RGCL [Shuai *et al.*, 2022] integrate review semantics into the contrastive framework, generating review-based augmented graphs and maximizing their mutual information with the original interaction graph. ReCAFR [V. Dong *et al.*, 2025] propose a review-centric contrastive framework to align user, item, and review representations in a unified space. *However, these works mainly focus on the consistency of node representations regarding global semantics. This coarse-grained alignment fails to effectively distinguish the entangled latent intents and fine-grained sentiment tendencies within the reviews.*

Disentangled and Sentiment-aware Recommendation. To address the aforementioned entanglement issue, another frontier trend is to disentangle complex user interests [Fu *et al.*, 2023; Zhang *et al.*, 2024a; Zhang *et al.*, 2024b]. DGCF [Wang *et al.*, 2020] mainly mines latent intent factors based on interaction graphs; SGDN [Ren *et al.*, 2023] proposes utilizing review semantics to guide graph disentanglement, decomposing the user-item graph into multiple latent factor graphs to capture diverse interaction motivations. DCMGNN [Li *et al.*, 2025a] constructs a multiplex graph to represent various interaction types and introduces an explicit behavior pattern learner to capture complex behavior patterns. To further capture user preferences, LETTER [Son *et al.*, 2025] explicitly splits user representations into "Likes" and "Dislikes" components and models their effects on ratings separately. *Despite significant progress, these works typically treat latent factors as parallel interaction channels, lacking an explicit mechanism to model the positive and negative counteraction. Furthermore, existing methods predominantly rely on hard thresholds to classify sentiment polarity, overlooking the subjectivity of ratings as well as the mixed sentiments within reviews.*

3 Problem Definition

Definition 1 (Notation and Input)**.** *Let $\mathcal{U}$ and $\mathcal{I}$ denote the sets of users and items, respectively. The model leverages both the explicit rating matrix $R \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}|}$ and the associated textual review set $\mathcal{D}$. The set of observed interaction pairs is denoted as $\mathcal{O} = \{(u, i) \mid r_{ui} \neq 0\}$.*

Problem (Task Formulation)**.** *Given $\mathcal{U}$, $\mathcal{I}$, R, and $\mathcal{D}$, the goal is to learn a predictive function $f_\Theta : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}$, parameterized by Θ. We formally define two distinct sub-tasks with their respective optimization objectives:*

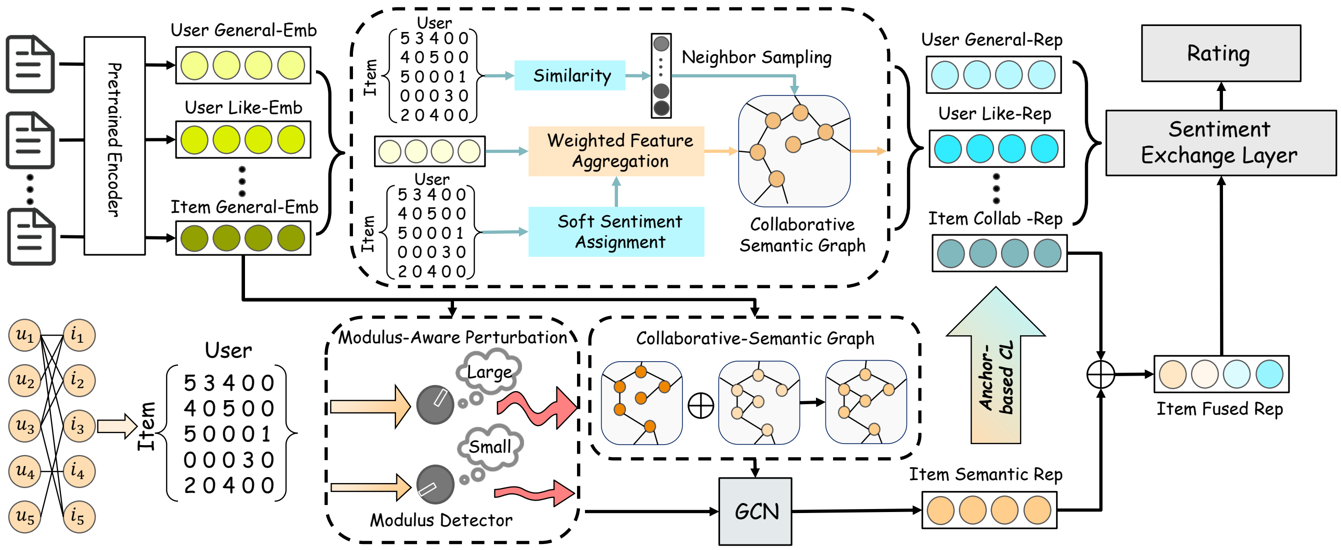


Figure 1: Overall framework of the proposed SENSE.

(1) Rating Prediction. This task aims to predict the exact rating value. The objective is to learn Θ such that the predicted rating $\hat{r}_{ui}$ approximates the ground-truth rating r_{ui} by minimizing the expected error:

$$\mathcal{L}_{rating} = \sum_{(u,i)\in\mathcal{O}} \left(r_{ui} - f_\Theta(u,i)\right)^2. \tag{1}$$

(2) Top-N Recommendation. This task aims to generate a ranked list of items. The objective is to learn Θ that ranks an observed item i higher than an unobserved item j, maximizing the pairwise ranking condition:

$$\mathcal{L}_{rank} = \sum_{((u,i),(u,j))\in\mathcal{D}_{rank}} -\ln\sigma\left(f_\Theta(u,i) - f_\Theta(u,j)\right), \tag{2}$$

where $\mathcal{D}_{rank} = \{((u,i),(u,j)) \mid (u,i) \in \mathcal{O} \wedge (u,j) \notin \mathcal{O}\}$ denotes the set of pairwise training samples.

4 Proposed Model

4.1 Sentiment-aware Behavior Modeling

Soft Sentiment Assignment and Feature Initialization. Given a user u and an item i associated with an interaction review d_{ui}, we first employ a pre-trained BERT encoder to extract the semantic embedding $e_{ui} \in \mathbb{R}^d$. To avoid discarding valuable information from mixed-sentiment reviews or introducing label noise, we bypass hard thresholds that categorize reviews into mutually exclusive positive or negative buckets. Instead, we adopt a soft assignment strategy based on Empirical Bayes Shrinkage. Specifically, we first estimate the global rating mean μ, along with two key variance parameters: the pooled within-user variance σ^2, representing the volatility of ratings within users, and the between-user variance δ^2, representing the diversity of rating biases across users.

Based on these statistics, we calculate a personalized shrinkage factor w_u for each user u. This balances their observed mean $\bar{r}_u$ with the global mean μ:

$$w_u = \frac{n_u\delta^2}{n_u\delta^2 + \sigma^2}, \tag{3}$$

where n_u denotes the number of interactions. Subsequently, we derive the user's personalized threshold T_u (i.e., the estimated posterior mean of user preference):

$$T_u = \mu + w_u(\bar{r}_u - \mu). \tag{4}$$

To measure the uncertainty of this estimation, we define a scaling factor $s_u = \sqrt{\sigma^2/n_u}$. This means that as user data grows, the uncertainty decreases. This makes the soft label boundary steeper. Finally, we transform the rating r_{ui} to a positivity confidence score l_{ui} using the sigmoid function:

$$l_{ui} = \sigma\left(\frac{r_{ui} - T_u}{s_u}\right). \tag{5}$$

This mechanism ensures that thresholds for cold-start users move to the global average, while thresholds for active users depend on their own past behaviors. Based on this score, we assign soft weights for the positive and negative views:

$$w_{ui}^{pos} = l_{ui}, \quad w_{ui}^{neg} = 1 - l_{ui}, \tag{6}$$

These weights function as confidence gates, allowing a single review to contribute to both views with varying intensities. Subsequently, within the weighted feature aggregation module, we initialize the user's specific sentiment features by aggregating their associated review embeddings via a weighted average. This ensures that the scale of the representation remains stable regardless of the number of reviews:

$$h_u^s = \frac{\sum_{i\in\mathcal{N}_u} w_{ui}^s \cdot e_{ui}}{\sum_{i\in\mathcal{N}_u} w_{ui}^s}, \quad s \in \{pos, neg\}, \tag{7}$$

where $\mathcal{N}_u$ denotes the set of items interacted by user u. A similar weighted aggregation is applied to items to obtain their respective sentiment representations h_i^s. For the general view which captures overall characteristics, we aggregate all reviews using uniform weights.

Rating-aware Homogeneous Graph Propagation. The initialized features h_u^s capture only local information. To capture high-order collaborative signals, we dynamically construct homogeneous graphs based on rating similarity. We then refine the representation via weighted neighbor aggregation and a residual connection:

$$e_{u,s}^{beh} = h_u^s + (\sum_{v \in \mathcal{N}_u} A_{uv} h_v^s)\mathbf{W}, \tag{8}$$

where $\mathbf{W}$ is a trainable weight matrix, A_{uv} represents the normalized edge weight based on rating similarity, and $\mathcal{N}_u$ is the set of sampled neighbors. The item behavior representations $e_{i,s}^{beh}$ are obtained analogously. These representations form the set E_{beh}, serving as the reliable backbone for our subsequent dual-view learning.

4.2 Collaborative-Guided Semantic Modeling

Relying solely on behavior view is often limited by data sparsity. It also cannot capture semantic connections between items. To solve this, we construct a parallel semantic view. To ensure incorporated semantic information is both rich and reliable, we design the collaborative-semantic graph, modulus-aware perturbation, and anchor-guided cross-view alignment.

Collaborative-Semantic Graph Construction

Constructing a high-quality item association graph is fundamental to semantic learning. Graphs built directly from review text similarities are prone to introducing irrelevant noise (e.g., items with similar phrasing but distinct categories), whereas pure interaction graphs suffer from sparsity in cold-start scenarios. To resolve this, we propose a collaborative-semantic graph, which aims to establish a unified graph structure calibrated by collaborative signals.

Specifically, we first calculate similarities from two independent views. We get the collaborative similarity $\mathbf{S}^{col}$ based on normalized rating vectors. We also get the semantic similarity $\mathbf{S}^{sem}$ based on BERT embeddings. To remove unneeded connections and reduce calculation costs, we apply a top-K filtering strategy to each matrix. For each item i, we keep only the top-K neighbors with the highest similarity. We mask the others to negative infinity:

$$\hat{\mathbf{S}}_{ij}^{\phi} = \begin{cases} \mathbf{S}_{ij}^{\phi}, & \text{if } j \in \mathcal{N}_K^{\phi}(i) \\ -\infty, & \text{otherwise} \end{cases}, \quad \phi \in \{\text{col}, \text{sem}\}. \tag{9}$$

Subsequently, we fuse them using a weighted union strategy. Unlike simple structural superposition, we explicitly control the contribution of collaborative and semantic signals via a balance coefficient β, generating the final fused adjacency matrix $\mathbf{A}_{ij}^{fused}$:

$$\mathbf{A}_{ij}^{fused} = \text{Softmax}\left(\beta \cdot \hat{\mathbf{S}}_{ij}^{col} + (1-\beta) \cdot \hat{\mathbf{S}}_{ij}^{sem}\right). \tag{10}$$

Through this mechanism, collaborative-semantic graph achieves a "soft calibration": when collaborative signals are sufficient (larger β), the graph structure is dominated by user behavioral patterns; when interactions are sparse, semantic similarity serves as an effective complement to bridge structural gaps. Finally, we employ a GCN layer to aggregate neighbor information on this fused graph, obtaining the item semantic representation $\mathbf{e}_i^{sem}$.

Modulus-aware Perturbation

To enhance robustness against semantic noise, we introduce modulus-aware perturbation. Unlike uniform noise injection, modulus-aware perturbation dynamically adapts perturbation intensity based on the feature modulus. This mechanism imposes sufficient regularization on dominant features with larger moduli to prevent overfitting, while applying minimal noise to fragile features with smaller moduli to preserve their semantic integrity [Zhang *et al.*, 2025b]. The adaptive injection is formulated as:

$$p_i = \frac{|\mathbf{h}_i|_2}{\max_{j \in \mathcal{I}} |\mathbf{h}_j|_2 + \xi}, \quad \mathbf{h}_i' = \mathbf{h}_i + \epsilon \cdot p_i \cdot \text{sign}(\mathbf{h}_i) \odot \boldsymbol{\delta}, \tag{11}$$

where $\mathbf{h}_i$ represents the semantic feature, $\boldsymbol{\delta} \sim \mathcal{U}(0,1)$ denotes random noise, and ϵ controls the magnitude. This strategy achieves scale-adaptive regularization, balancing overfitting suppression and semantic stability.

Anchor-Guided Cross-View Alignment

Through the preceding processes, we derive sparse yet accurate item collaborative representations $\mathbf{e}_i^{\text{beh}}$ from the interaction graph, while simultaneously generating rich but noise-prone semantic representations $\mathbf{e}_i^{\text{sem}}$ via the collaborative-semantic graph. To map these two distinct modalities into a unified feature manifold and leverage semantic information to augment behavior representations, we adopt an anchor-guided contrastive learning strategy.

Crucially, distinct from conventional contrastive learning that optimizes views symmetrically, we establish the behavior view which faithfully reflects the true user preference topology as the fixed anchor. Consequently, we enforce the semantic view to unidirectionally align with this anchor in the feature space. This asymmetric alignment objective is implemented via a one-way InfoNCE loss [Oord *et al.*, 2018; Chen *et al.*, 2020]:

$$\mathcal{L}_{Align} = -\sum_{i \in \mathcal{B}} \log \frac{\exp(\text{sim}(\mathbf{e}_i^{beh}, \mathbf{e}_i^{sem})/\tau)}{\sum_{j \in \mathcal{B}} \exp(\text{sim}(\mathbf{e}_i^{beh}, \mathbf{e}_j^{sem})/\tau)}, \tag{12}$$

where τ is the temperature coefficient and $\mathcal{B}$ denotes the item set in the current batch. Specifically, this objective establishes the behavior representation $\mathbf{e}_i^{beh}$ as the reference anchor, optimizing the proximity of the corresponding semantic representation $\mathbf{e}_i^{sem}$ within the feature space while distancing irrelevant semantic samples $\mathbf{e}_j^{sem}$. This mechanism essentially imposes a distribution constraint, enforcing the manifold structure of the semantic space to maintain consistency with the interaction topology. Thereby, it implicitly injects high-confidence collaborative signals into semantic representations, effectively filtering out textual noise irrelevant to user preferences.

4.3 Fine-grained Sentiment Interaction and Multi-Task Optimization

In the preceding sections, we extract user and item representations across three sentiment dimensions (general, positive, negative) from the behavior view, and enhance the item representation via the semantic view. This section details how to fuse these dual-view information sources and achieve precise prediction through a sentiment exchange mechanism, followed by the joint optimization objective.

View Fusion and Sentiment Exchange Prediction. Since the semantic view has been mapped to the same feature manifold as the behavior view via the alignment operation in Section 4.2, we employ an additive fusion strategy to enhance the item general representation:

$$\mathbf{e}_i^{\text{fused}} = \mathbf{e}_i^{\text{beh}} + \mathbf{e}_i^{\text{sem}}, \tag{13}$$

where $\mathbf{e}_i^{\text{beh}}$ denotes the item general representation from the behavior view, and $\mathbf{e}_i^{\text{sem}}$ is the semantic representation aggregated via collaborative-semantic graph. It is worth noting that semantic augmentation is applied exclusively to the item general dimension, while the sentiment-specific dimensions (positive/negative) retain pure behavioral signals to maintain the discriminative power of sentiment polarity.

Based on the fused representations, we design a sentiment exchange mechanism for prediction. The core idea is to explicitly match consistency between users and items at a fine-grained sentiment level. Specifically, a user's overall attitude depends not only on the match between their preferences and item merits, but also on the conflict between their dislikes and item flaws, as well as the consistency of general features. Therefore, we decompose the final prediction score $\hat{r}_{ui}$ into the sum of interactions across multiple dimensions:

$$\begin{aligned}\hat{r}_{ui} = & \underbrace{\mathbf{e}_u^{\text{beh}\top}\mathbf{e}_i^{\text{fused}}}_{\text{General Match}} + \underbrace{\mathbf{e}_u^{\text{pos}\top}\mathbf{e}_i^{\text{pos}}}_{\text{Positive Match}} - \underbrace{\mathbf{e}_u^{\text{neg}\top}\mathbf{e}_i^{\text{neg}}}_{\text{Negative Match}} \\ & + \underbrace{\mathbf{e}_u^{\text{pos}\top}\mathbf{e}_i^{\text{fused}} - \mathbf{e}_u^{\text{neg}\top}\mathbf{e}_i^{\text{fused}}}_{\text{Cross-Dimension Interaction}} + b_u + b_i + \mu,\end{aligned} \tag{14}$$

where b_u and b_i are user and item bias terms, and μ is the global average. This decomposition allows the model to evaluate the user's attitude towards the item from both positive incentives and negative resistance.

Multi-Task Joint Training. To accommodate different recommendation scenarios and optimize representation quality, we adopt a multi-task learning framework incorporating a main task loss, a contrastive alignment loss, and regularization. Depending on the specific task (Rating Prediction or Top-N Recommendation), the main task loss $\mathcal{L}_{\text{Main}}$ is defined as follows:

1) Rating Prediction Task. For explicit feedback, we employ the Mean Squared Error (MSE) loss.

2) Top-N Recommendation Task. For implicit ranking, we employ the Bayesian Personalized Ranking (BPR) loss, which aims to maximize the score margin between positive items i and negative items j.

Joint Optimization Objective. Combining the cross-view alignment loss $\mathcal{L}_{\text{Align}}$ proposed in Section 4.2, we define the total loss function as:

$$\mathcal{L} = \mathcal{L}_{\text{Main}} + \lambda_1 \mathcal{L}_{\text{Align}} + \lambda_2 \|\Theta\|_2^2, \tag{15}$$

where $\mathcal{L}_{\text{Main}}$ is either $\mathcal{L}_{\text{MSE}}$ or $\mathcal{L}_{\text{BPR}}$ depending on the task. $\mathcal{L}_{\text{Align}}$ constrains the semantic representation to align with the behavioral topological structure, $\|\Theta\|_2^2$ is the L_2 regularization term to prevent overfitting, and λ_1, λ_2 are hyperparameters. Through this joint optimization, the model continuously calibrates semantic representations via $\mathcal{L}_{\text{Align}}$ while learning precise interaction patterns, achieving mutual reinforcement between the dual views.

Datasets	#Users	#Items	#Interactions	Density
VG	24,303	10,672	231,780	0.089%
TG	19,412	11,924	167,597	0.072%
SO	35,598	18,357	296,337	0.045%
CL	39,387	23,033	278,677	0.031%

Table 1: Statistics of the datasets.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate our model on four Amazon 5-core datasets [McAuley *et al.*, 2015] covering different domains: Toys and Games (TG), Video Games (VG), Clothing, Shoes and Jewelry (CL), and Sports and Outdoors (SO). Table 1 summarizes the statistics of datasets.

Evaluation Metrics. To evaluate the performance of rating prediction and Top-N recommendation, we employ widely used metrics: Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) for rating prediction. Recall@10 and Normalized Discounted Cumulative Gain (NDCG@10) [Järvelin and Kekäläinen, 2002] for recommendation performance.

Baselines. We compare SENSE with three categories of state-of-the-art methods: (1) **Review-aware Representation Learning Methods**, including DeepCoNN [Zheng *et al.*, 2017] , NARRE [Chen *et al.*, 2018] and RMG [Wu *et al.*, 2019]; (2) **Graph Contrastive Learning Methods**, covering LightGCN [He *et al.*, 2020], RGCL [Shuai *et al.*, 2022], RMCL [Yang *et al.*, 2023], MAGCL [Wang *et al.*, 2023], ReHCL [Wang *et al.*, 2024], HGCL [Shui *et al.*, 2024], and MHCL [Li *et al.*, 2025b]; and (3) **Disentangled Representation Learning Methods**, represented by SGDN [Ren *et al.*, 2023], APH [Liu *et al.*, 2025], and LETTER [Son *et al.*, 2025].

5.2 Performance Comparison

We comprehensively compare the performance of SENSE with all baseline models across four datasets. Table 2 reports the main experimental results for the rating prediction task (RMSE, MSE, MAE), while we evaluate the performance of each model on the Top-N recommendation task (Recall@10, NDCG@10) using a full ranking setting, with the results shown in Table 3.

Through a deep comparison of the performance of different categories of models, we derive the following three key observations:

- *Observation 1*: SENSE consistently outperforms all baseline models across all datasets and metrics. Compared to

Model	Toys and Games (TG)			Video Games (VG)			Clothing (CL)			Sports (SO)		
	MAE	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**
DeepCoNN	0.6458	0.8026	0.8890	0.8831	1.1410	1.0682	0.8656	1.1184	1.0575	0.7194	0.8856	0.9412
NARRE	0.6226	0.7961	0.8769	0.8449	1.0916	1.0448	0.8564	1.1064	1.0519	0.6902	0.8821	0.9493
RMG	0.6190	0.7901	0.8438	0.8621	1.1138	1.0554	0.8521	1.1064	1.0519	0.6613	0.9032	0.9411
AHN	0.6362	0.8220	0.9066	0.8611	1.1125	1.0548	0.8856	1.1442	1.0697	0.6406	0.9123	0.9323
RGCL	0.6015	0.7771	0.8815	0.8216	1.0615	1.0303	0.8404	1.0858	1.0420	0.6564	0.8446	0.9190
RMCL	0.5904	0.7491	0.8655	0.7952	1.0274	1.0136	0.7961	1.0286	1.0142	0.6359	0.7685	0.9163
ReHCL	0.6111	0.7755	0.8806	0.7816	1.0693	1.0341	0.7678	1.0668	1.0329	0.6704	0.8491	0.9215
SGDN	0.5885	0.7603	0.8720	0.7827	1.0112	1.0056	0.8101	1.0466	1.0230	0.6416	0.7791	0.9188
HGCL	0.5916	0.7125	0.8441	0.7582	1.0292	1.0145	0.7137	0.9614	0.9805	0.6380	0.7703	0.8777
MHCL	0.5862	0.7616	0.8727	0.7965	1.0432	1.0214	0.8131	1.0582	1.0287	0.6424	0.8329	0.9126
APH	0.6051	0.7859	0.8865	0.8256	1.0829	1.0406	0.8514	1.0819	1.0401	0.6935	0.8869	0.9418
LETTER	0.5529	0.6364	0.7888	0.7260	0.9623	0.9760	0.6488	0.7771	0.8775	0.5834	0.6675	0.8125
SENSE	**0.4665**	**0.4895**	**0.6910**	**0.6573**	**0.7962**	**0.8876**	**0.5642**	**0.6331**	**0.7923**	**0.4800**	**0.5463**	**0.7327**
Improv.	15.63%	23.08%	12.40%	9.46%	17.26%	9.06%	13.04%	18.53%	9.71%	17.72%	18.16%	9.82%

Table 2: Performance comparison on Rating Prediction task. The best results are highlighted in **bold**, and the runner-ups are underlined.

Model	Toys and Games		Video Games		Clothing		Sports	
	Recall@10	**NDCG@10**	**Recall@10**	**NDCG@10**	**Recall@10**	**NDCG@10**	**Recall@10**	**NDCG@10**
LightGCN	0.0725	0.0391	0.1138	0.0711	0.0316	0.0194	0.0552	0.0341
RGCL	0.0683	0.0412	0.0955	0.0585	0.0348	0.0200	0.0537	0.0313
MAGCL	0.0728	0.0441	0.1008	0.0616	0.0367	0.0214	0.0572	0.0332
ReHCL	0.0726	0.0448	0.1046	0.0639	0.0409	0.0234	0.0678	0.0412
HGCL	0.0792	0.0492	0.1157	0.0700	0.0446	0.0259	0.0740	0.0457
MHCL	0.0919	0.0545	0.1275	0.0785	0.0521	0.0312	0.0851	0.0515
LETTER	0.1062	0.0637	0.0990	0.0571	0.0699	0.0411	0.1049	0.0632
SENSE	**0.1306**	**0.0774**	**0.1487**	**0.0911**	**0.0779**	**0.0448**	**0.1117**	**0.0664**
Improv.	22.98%	21.51%	16.65%	16.07%	11.44%	9.00%	6.48%	5.06%

Table 3: Performance comparison on Top-N Recommendation. The best results are highlighted in **bold**, and the runner-ups are underlined.

the strongest baseline, SENSE achieves a significant improvement of up to 23.08% in MSE for rating prediction, and improves Recall@10 and NDCG@10 by 22.98% and 21.51% respectively for ranking. This success is attributed to the synergy of three core designs: the soft sentiment assignment eliminates label noise from rigid thresholds to optimize rating accuracy; the sentiment exchange mechanism improves ranking by modeling the interplay between positive incentives and negative resistance; and the anchor-guided alignment filters semantic noise to ensure effective multi-modal fusion.

- *Observation 2*: Although methods like RGCL capture high-order collaborative signals via message passing, their performance is limited on sparse datasets. SENSE addresses this by introducing a collaborative-guided semantic channel. It uses rich review semantics as external knowledge to supplement node representations and effectively compensates for the sparsity of pure interaction signals through the graph structure.
- *Observation 3*: Experiments show that some review-based models perform worse than pure interaction models in ranking tasks due to a lack of alignment, suggesting that unaligned semantic information can act as noise. SENSE employs an anchor-guided cross-view alignment mechanism to bridge the gap between semantic and behavioral features. This ensures that semantic information acts as a positive enhancement rather than interference.

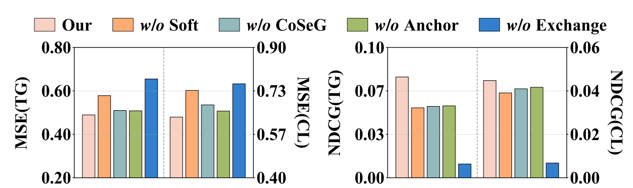
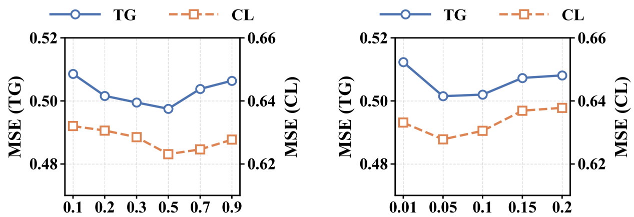


Figure 2: Ablation results. TG refers to the left y-axis, and CL refers to the right y-axis.

5.3 Ablation Studies

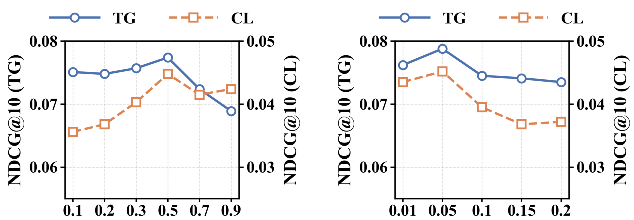
To investigate the effectiveness of the core components of SENSE, we compare the full model against four variants on the TG and CL datasets:

- *w/o Soft* uses the hard threshold of rating 4 to classify sentiment.
- *w/o CoSeG* removes the collaborative guidance in graph construction.



(a) Fusion Ratio β (b) Alignment Weight λ_1

Figure 3: Sensitivity analysis of key hyperparameters β and λ_1 on rating prediction performance (MSE).



(a) Fusion Ratio β (b) Alignment Weight λ_1

Figure 4: Sensitivity analysis of key hyperparameters β and λ_1 on Top-N recommendation performance (NDCG@10).

- *w/o Anchor* replaces anchor-guided alignment with standard mutual contrastive learning.
- *w/o Exchange* uses MLP concatenation instead of explicit sentiment interaction.

Figure 2 illustrates the performance of these variants in terms of MSE and NDCG@10.

Critical Role of Explicit Interaction Mechanism. The most notable observation arises from the *w/o Exchange* variant. This variant not only suffers significant degradation in MSE but also exhibits a performance collapse in the NDCG metric. This asymmetric degradation demonstrates that the proposed explicit interaction formula (Eq. 14), by constructing a counteraction between positive incentives and negative resistance within the vector space, empowers the model to precisely discriminate top-tier items within the dense semantic space. This capability to capture fine-grained preference rankings is indispensable for the model.

Impact of Uncertainty Modeling and Semantic Calibration. The significant inferiority of *w/o Soft* in MSE indicates that hard thresholding ignores subjective uncertainty of ratings and introduces label noise, whereas soft labels effectively smooth noise. Furthermore, the performance decline in *w/o CoSeG* and *w/o Anchor* confirms that graph structures constructed solely from textual semantics are prone to noise, and semantic features tend to drift away from true interaction manifold. The structural guidance from collaborative signals and anchor constraints successfully *calibrates* the geometry of semantic space, ensuring high-quality feature alignment.

5.4 Hyperparameter Sensitivity Analysis

We examine the impact of the graph fusion ratio β and the alignment weight λ_1 on model performance, as illustrated in Figure 3 and Figure 4.

Impact of Fusion Ratio β. We vary β within the range of $[0.1, 0.9]$. The results exhibit a consistent bell-shaped trend across all datasets and metrics. Performance initially improves as β increases, reaching an optimal peak at $\beta = 0.5$, before subsequently declining. This observation confirms that relying exclusively on either sparse collaborative signals or noisy semantic features yields suboptimal results, highlighting the necessity of dual-view integration. By assigning equal importance to both views, SENSE effectively leverages their complementary strengths to mitigate data sparsity and filter out text-induced noise, thereby enhancing representation robustness.

Impact of Alignment Weight λ_1. We investigate λ_1 in the range of $\{0.01, \ldots, 0.2\}$. The model achieves its best performance at $\lambda_1 = 0.05$. Performance drops when λ_1 is too small, indicating insufficient constraint to align semantic features with the behavioral manifold. Conversely, a value larger than 0.05 leads to performance degradation, which we attribute to over-regularization, where the auxiliary alignment objective interferes with the primary recommendation task.

5.5 Complexity Analysis

We analyze the time and space complexity of SENSE. Let d be the embedding dimension, $|\mathcal{O}|$ be the interaction count, $|\mathcal{U}|$ be the user count, and $|\mathcal{I}|$ be the item count.

Time Complexity. Semantic extraction is offline ($\mathcal{O}(1)$). Soft sentiment assignment and rating-aware graph propagation take $\mathcal{O}(|\mathcal{O}|)$ and $\mathcal{O}(|\mathcal{O}|d)$ time, respectively. Constructing the collaborative-semantic graph and computing GCN aggregation require $\mathcal{O}(|\mathcal{I}|Kd)$ time, where K is the number of sampled neighbors. Thus, the overall training time complexity is $\mathcal{O}((|\mathcal{O}| + |\mathcal{I}|K)d)$, which scales linearly with data size.

Space Complexity. Memory overhead mainly comes from trainable embeddings and graph topology. User and item representations across multi-sentiment dimensions require $\mathcal{O}((|\mathcal{U}|+|\mathcal{I}|)d)$ space. Sparse adjacency matrices for interaction and fused graphs take $\mathcal{O}(|\mathcal{O}| + |\mathcal{I}|K)$ space. The overall space complexity is $\mathcal{O}((|\mathcal{U}| + |\mathcal{I}|)d + |\mathcal{O}| + |\mathcal{I}|K)$, proving that SENSE is highly scalable for real-world scenarios.

6 Conclusion

We propose SENSE to address label noise and cross-modal misalignment in review-enhanced recommendation. Specifically, it employs a soft sentiment assignment to reduce label noise, and an anchor-guided cross-view alignment to calibrate semantic features using collaborative signals. Furthermore, a sentiment exchange mechanism is designed to explicitly model the trade-off between positive and negative user preferences. Extensive experiments on four datasets demonstrate that SENSE significantly outperforms state-of-the-art methods in both rating prediction and Top-N recommendation. Our future work will focus on integrating large language models to further enhance the multimodal semantic representation learning process.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 62176243, 62276079, 62472263, and 62302120, China Postdoctoral Science Foundation under Grant No. 2025M773472, and the Postdoctoral Fellowship Program of CPSF under Grant No. GZC20252288.

References

[Catherine and Cohen, 2017] Rose Catherine and William Cohen. Transnets: Learning to transform for recommendation. In *RecSys*, pages 288–296, 2017.

[Chen *et al.*, 2018] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 world wide web conference*, pages 1583–1592, 2018.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607, 2020.

[Fang *et al.*, 2025] Shuyun Fang, Junling Wang, and Fukun Chen. A hyperbolic graph neural network model with contrastive learning for rating–review recommendation. *Entropy*, 27(8):886, 2025.

[Fu *et al.*, 2023] Chaofan Fu, Guanjie Zheng, Chao Huang, Yanwei Yu, and Junyu Dong. Multiplex heterogeneous graph neural network with behavior pattern modeling. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pages 482–494, 2023.

[Gao *et al.*, 2020] Jingyue Gao, Yang Lin, Yasha Wang, Xiting Wang, Zhao Yang, Yuanduo He, and Xu Chu. Set-sequence-graph: A multi-view approach towards exploiting reviews for recommendation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 395–404, 2020.

[Guo *et al.*, 2025] Lei Guo, Chenlong Song, Feng Guo, Xiaohui Han, Xiaojun Chang, and Lei Zhu. Semantic-enhanced co-attention prompt learning for non-overlapping cross-domain recommendation. *ACM Transactions on Information Systems*, 43(5):1–27, 2025.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and 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*, pages 639–648, 2020.

[Järvelin and Kekäläinen, 2002] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[Kipf, 2016] TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Li *et al.*, 2025a] Xiang Li, Chaofan Fu, Zhongying Zhao, Guangjie Zheng, Chao Huang, Yanwei Yu, and Junyu Dong. Dual-channel multiplex graph neural networks for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 37(6):3327–3341, 2025.

[Li *et al.*, 2025b] Xiang Li, Changsheng Shui, Zhongying Zhao, Junyu Dong, and Yanwei Yu. Multi-channel hypergraph contrastive learning for matrix completion. *ACM Transactions on Information Systems*, 44(1):1–27, 2025.

[Li *et al.*, 2026a] Xiang Li, Yuan Cao, Zhongying Zhao, Guoqing Chao, and Yanwei Yu. Multiplex heterogeneous graph neural networks with euclidean-riemannian mutual space synergy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15126–15134, 2026.

[Li *et al.*, 2026b] Xiang Li, Jianpeng Qi, Haobing Liu, Yuan Cao, Guoqing Chao, Zhongying Zhao, Junyu Dong, Xinwang Liu, and Yanwei Yu. Scalegnn: Towards scalable graph neural networks via adaptive high-order neighboring feature fusion. In *Proceedings of the ACM Web Conference 2026*, pages 980–991, 2026.

[Lin *et al.*, 2022] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. Improving graph contrastive learning for recommendation. In *WWW*, pages 2459–2468, 2022.

[Ling *et al.*, 2014] Guang Ling, Michael R Lyu, and Irwin King. Ratings meet reviews, a combined approach to recommend. In *RecSys*, pages 105–112, 2014.

[Liu *et al.*, 2020] Siwei Liu, Iadh Ounis, Craig Macdonald, and Zaiqiao Meng. A heterogeneous graph neural model for cold-start recommendation. In *SIGIR*, 2020.

[Liu *et al.*, 2025] Junrui Liu, Tong Li, Di Wu, Zifang Tang, Yuan Fang, and Zhen Yang. An aspect performance-aware hypergraph neural network for review-based recommendation. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 503–511, 2025.

[Liu *et al.*, 2026] Yuchen Liu, Kunyu Ni, Zhongying Zhao, Guoqing Chao, and Yanwei Yu. S^2hyrec: Self-supervised hypergraph sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15394–15402, 2026.

[McAuley and Leskovec, 2013] Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172, 2013.

[McAuley *et al.*, 2015] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *SIGIR*, pages 43–52, 2015.

[Oord *et al.*, 2018] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[Ren *et al.*, 2023] Yuyang Ren, Haonan Zhang, Qi Li, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. Self-supervised graph disentangled networks for review-based recommendation. In *IJCAI*, 2023.

[Shi *et al.*, 2023] Jiacheng Shi, Yanmin Zhu, Ke Wang, Mengyuan Jing, Tianzi Zang, Jiadi Yu, and Feilong Tang. Rating-review graph contrastive learning for review-based recommendation. In *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 1522–1529. IEEE, 2023.

[Shuai *et al.*, 2022] Wei Shuai, Zhang Kun, et al. Review-graph contrastive learning for recommendation. In *SIGIR*, 2022.

[Shui *et al.*, 2024] Changsheng Shui, Xiang Li, Jianpeng Qi, Guiyuan Jiang, and Yanwei Yu. Hierarchical graph contrastive learning for review-enhanced recommendation. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 423–440. Springer, 2024.

[Son *et al.*, 2025] Jiwon Son, Hyunjoon Kim, and Sang-Wook Kim. Rating-aware homogeneous review graphs and user likes/dislikes differentiation for effective recommendations. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2070–2080, 2025.

[V. Dong *et al.*, 2025] Hoang V. Dong, Yuan Fang, and Hady W Lauw. A contrastive framework with user, item and review alignment for recommendation. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 117–126, 2025.

[Wang *et al.*, 2020] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. Disentangled graph collaborative filtering. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 1001–1010, 2020.

[Wang *et al.*, 2023] Ke Wang, Yanmin Zhu, Tianzi Zang, Chunyang Wang, Kuan Liu, and Peibo Ma. Multi-aspect graph contrastive learning for review-enhanced recommendation. *ACM Transactions on Information Systems*, 42(2):1–29, 2023.

[Wang *et al.*, 2024] Ke Wang, Yanmin Zhu, Tianzi Zang, Chunyang Wang, and Mengyuan Jing. Review-enhanced hierarchical contrastive learning for recommendation. In *AAAI*, 2024.

[Wang *et al.*, 2025] Zihan Wang, Jinghao Lin, Xiaocui Yang, Yongkang Liu, Shi Feng, Daling Wang, and Yifei Zhang. Enhancing llm-based recommendation through semantic-aligned collaborative knowledge. *arXiv preprint arXiv:2504.10107*, 2025.

[Wu *et al.*, 2019] Chuhan Wu, Fangzhao Wu, Tao Qi, Suyu Ge, Yongfeng Huang, and Xing Xie. Reviews meet graphs: Enhancing user and item representations for recommendation with hierarchical attentive graph neural network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4884–4893, 2019.

[Wu *et al.*, 2021] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 726–735, 2021.

[Xie *et al.*, 2021] Ruobing Xie, Cheng Ling, Yalong Wang, Rui Wang, Feng Xia, and Leyu Lin. Deep feedback network for recommendation. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 2519–2525, 2021.

[Yang *et al.*, 2023] Wei Yang, Tengfei Huo, Zhiqiang Liu, and Chi Lu. based multi-intention contrastive learning for recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2339–2343, 2023.

[Yu *et al.*, 2022] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 1294–1303, 2022.

[Zhang *et al.*, 2019] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1):1–38, 2019.

[Zhang *et al.*, 2024a] Xiaokun Zhang, Bo Xu, Zhaochun Ren, Xiaochen Wang, Hongfei Lin, and Fenglong Ma. Disentangling id and modality effects for session-based recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pages 1883–1892, 2024.

[Zhang *et al.*, 2024b] Xiaokun Zhang, Bo Xu, Youlin Wu, Yuan Zhong, Hongfei Lin, and Fenglong Ma. Finerec: Exploring fine-grained sequential recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pages 1599–1608, 2024.

[Zhang *et al.*, 2025a] Jinyu Zhang, Zhongying Zhao, Chao Li, and Yanwei Yu. Lightweight yet fine-grained: A graph capsule convolutional network with subspace alignment for shared-account sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13242–13250, 2025.

[Zhang *et al.*, 2025b] Xiaoxiong Zhang, Xin Zhou, Zhiwei Zeng, Dusit Niyato, and Zhiqi Shen. Semantic item graph enhancement for multimodal recommendation. *arXiv preprint arXiv:2508.06154*, 2025.

[Zheng *et al.*, 2017] Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth ACM international conference on web search and data mining*, pages 425–434, 2017.