FineRec: Exploring Fine-grained Sequential Recommendation

Xiaokun Zhang*
Dalian University of Technology
dawnkun1993@gmail.com

Bo Xu†
Dalian University of Technology
xubo@dlut.edu.cn

Youlin Wu
Dalian University of Technology
wuyoulin@mail.dlut.edu.cn

Yuan Zhong
Pennsylvania State University
yfz5556@psu.edu

Hongfei Lin
Dalian University of Technology
hflin@dlut.edu.cn

Fenglong Ma
Pennsylvania State University
fenglong@psu.edu

ABSTRACT

Sequential recommendation is dedicated to offering items of interest for users based on their history behaviors. The attribute-opinion pairs, expressed by users in their reviews for items, provide the potentials to capture user preferences and item characteristics at a fine-grained level. To this end, we propose a novel framework FineRec that explores the attribute-opinion pairs of reviews to finely handle sequential recommendation. Specifically, we utilize a large language model to extract attribute-opinion pairs from reviews. For each attribute, a unique attribute-specific user-opinion-item graph is created, where corresponding opinions serve as the edges linking heterogeneous user and item nodes. To tackle the diversity of opinions, we devise a diversity-aware convolution operation to aggregate information within the graphs, enabling attribute-specific user and item representation learning. Ultimately, we present an interaction-driven fusion mechanism to integrate attribute-specific user/item representations across all attributes for generating recommendations. Extensive experiments conducted on several real-world datasets demonstrate the superiority of our FineRec over existing state-of-the-art methods. Further analysis also verifies the effectiveness of our fine-grained manner in handling the task.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Sequential Recommendation, User-item Reviews, Attribute-Opinions, Fine-grained User and Item Representation.

ACM Reference Format:
Xiaokun Zhang, Bo Xu, Youlin Wu, Yuan Zhong, Hongfei Lin, and Fenglong Ma. 2024. FineRec: Exploring Fine-grained Sequential Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July*

---

*This work was done during Xiaokun visiting Prof.Fenglong Ma at Peen State.
†Corresponding Author.

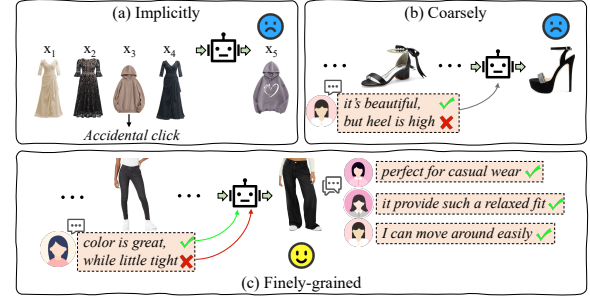**Figure 1: The limitations of existing implicit (a) and coarse (b) recommendations. Our proposed fine-grained manner(c).**

1 INTRODUCTION

Sequential Recommendation (SR) aims to provide items of interest for users based on their history behaviors [12, 16]. SR concentrates on addressing the inherent feature of user behaviors, *i.e.,* chronologically evolving preferences. This focus equips SR with the ability to grasp users' shifting interests, deliver personalized services timely, and adapt to the fast-paced trends of the modern information age. Owing to such merits, SR has received widespread attention from both academia and industry in recent years [10, 20, 36, 40].

Existing methods typically rely on neural networks to obtain item and user embeddings, basing recommendations on the similarity between these embeddings. Most efforts attend to model **implicit** activities of users, like clicks, to infer their preferences, employing various neural architectures including Recurrent Neural Networks (RNN) [12, 19], attention mechanism [16, 34] and Graph Neural Networks (GNN) [35, 37]. As shown in Figure 1 (a), unfortunately, a significant drawback of this implicit approach is that implicit actions, like accidental clicks, can sometimes misrepresent user interests, thereby introducing noise into the models [37, 45]. Another line of research [9, 18, 32] utilizes user-item reviews to explicitly capture user preferences. However, they all **coarsely** treat a review as a whole, overlooking the fact that users may have distinct opinions about different attributes within a single review. As presented in Figure 1 (b), this failure to discern such fine-grained distinctions limits the effectiveness of these methods.

Actually, in a review, a user expresses her specific feelings for an item in the form of **attribute-opinion** pairs. These pairs provide an opportunity to capture user preferences and item characteristics at an **explicit and fine-grained** manner, offering a new perspective

to improve recommendation. As depicted in Figure 1 (c), this manner facilitates satisfactory recommendations, such as suggesting loose clothing to a user who previously expressed a dislike for tight fits, and the new item is favorably described as loose by other users. Despite holding encouraging prospects, effectively implementing the fine-grained manner in SR faces several obstacles:

Firstly, it is difficult to *extract informative attribute-opinion pairs* from reviews. Users often express their opinions using informal and implicit language, which complicates the extraction. For instance, consider the review, "I was robbed by this thing!"–neither the attribute (Price) nor opinion (Expensive) is delivered in a straightforward way. Lacking rich language knowledge, current methods, either based on hand-craft rules [15] or well-designed models [3] tailored for certain fields, can not accurately extract informative attribute-opinion pairs from such reviews.

Secondly, a major challenge lies in *finely representing users and items* under each attribute with diverse opinions. On one hand, users and items exhibit unique preferences and characteristics under different attributes. However, current methods typically represent them in a overall way, overlooking the distinctions across attributes. On the other hand, even in a certain attribute, the opinion diversity impedes the learning of fine-grained user and item representations. For users, their attitudes would be changing along with distinct items on an attribute. For example, a user may prefer televisions with large size for enhanced viewing, while favoring small-sized phones due to portability. As to items, they may receive various ratings from different users under an attribute. It is common that some users favor brightly colored clothes, while others may hate them. Thus, it is imperative to consider opinion diversity to achieve fine-grained user and item representations at attribute-level.

Thirdly, it remains a problem to *generate recommendations* in the fine-grained manner. User behaviors are determined by user preferences and item characteristics in various attributes jointly. For instance, a loyal fan may buy an item of her favored brand despite disliking its high price. Unfortunately, we can only observe the overall user-item interactions, while there are no explicit indicators to show how each attribute affects user decisions. This makes it challenging to infer user actions in the fine-grained manner.

To tackle these issues, we propose a novel framework to incorporate attribute-opinion pairs for <u>Fine</u>ly handling sequential <u>Rec</u>ommendation (FineRec). Firstly, trained on massive text data, the Large Language Model (LLM) encapsulates a wealth of language knowledge [1, 11, 28]. In light of this, we utilize LLM to extract attribute-opinion pairs from reviews. To relieve hallucinations of LLMs in tackling complex tasks, we obtain attributes based on their importance on websites and subsequently extract corresponding opinions via LLM. Secondly, to represent users/items under distinct attributes, we create a unique attribute-specific user-opinion-item graph for each attribute. Within the graph, heterogeneous user and item nodes are connected by opinion edges which represent the specific opinions users have expressed about items under the attribute. Afterwards, a diversity-aware convolution operation is devised to achieve fine-grained user and item representation on the graphs, with an emphasis on opinion diversity during the information aggregation. Lastly, we present an interaction-driven fusion mechanism that leverages user-item interaction information to guide the fusion of attribute-specific item/user representations,

achieving final recommendations. In summary, the contributions of this study are outlined as follows:

- We present a new perspective to handle sequential recommendation, where the attribute-opinions of reviews are explored to finely reveal user preferences and item characteristics. As far as we know, this is the pioneering attempt to handle the task in such a fine-grained manner.
- We proposed a novel framework FineRec comprising several innovative techniques, including LLM-based attribute-opinion extraction, attribute-specific user-opinion-item graph with a diversity-aware convolution operation, and an interaction-driven fusion mechanism, to achieve fine-grained sequential recommendation.
- Extensive experiments on several public benchmarks validate the superiority of our FineRec over state-of-the-art methods. Further analysis also justifies the effectiveness of our fine-grained manner in handling the task.

2 RELATED WORK

2.1 Sequential Recommendation

With the powerful ability in representing users and items, neural networks have come to dominate the field of sequential recommendation [10, 27]. Various neural structures are employed to model user behavior sequences like RNN [12, 19], attention mechanism [16, 38, 39], Transformers [34, 45, 47], GNN [35, 37], MLP [21, 22] and contrastive learning [5, 30]. Some recent methods incorporate extra item information to mine user preferences from multiple views including categories [2, 4, 26], brands [10, 33], price [41, 43], text [13, 20, 28] as well as images [14, 42]. However, all of them focus on modeling user implicit behaviors which can not represent genuine intents of users in some cases. Thus, this limitation significantly hampers their effectiveness.

2.2 Review-driven Recommendation

Reviews, posted by users to explicitly convey their opinions on items, are utilized by recent works to improve recommendation task [9, 18, 32]. These works typically employ reviews to build user and item embeddings, and then rely on certain neural structures to offer suggestions like attention mechanism [9, 25] and GNN [32]. RNS [18] is a representative work incorporating reviews into sequential recommendation, where review contents are utilized to obtain users' long-term and short-term preferences. However, these efforts ignore the significant differences in a user review regarding different attributes, and instead coarsely model a review as a whole. Moreover, some methods attempt to capture user distinct interest on varying aspects [6, 7, 31, 44]. Unfortunately, they only focus on discerning various aspects while do not capture specific user opinions towards these aspects. Such paradigms fall short of finely disclosing user preferences or item characteristics either.

3 PROBLEM STATEMENT

Let $\mathcal{U}$ and $\mathcal{X}$ denote the unique user and item set, respectively. The $\mathcal{R}$ is a review set consisting of all reviews posted by users on items. Each user $u_i \in \mathcal{U}$ has interacted with a sequence of items chronologically, producing a behavior sequence denoted as: $\mathcal{S}^{u_i} =$
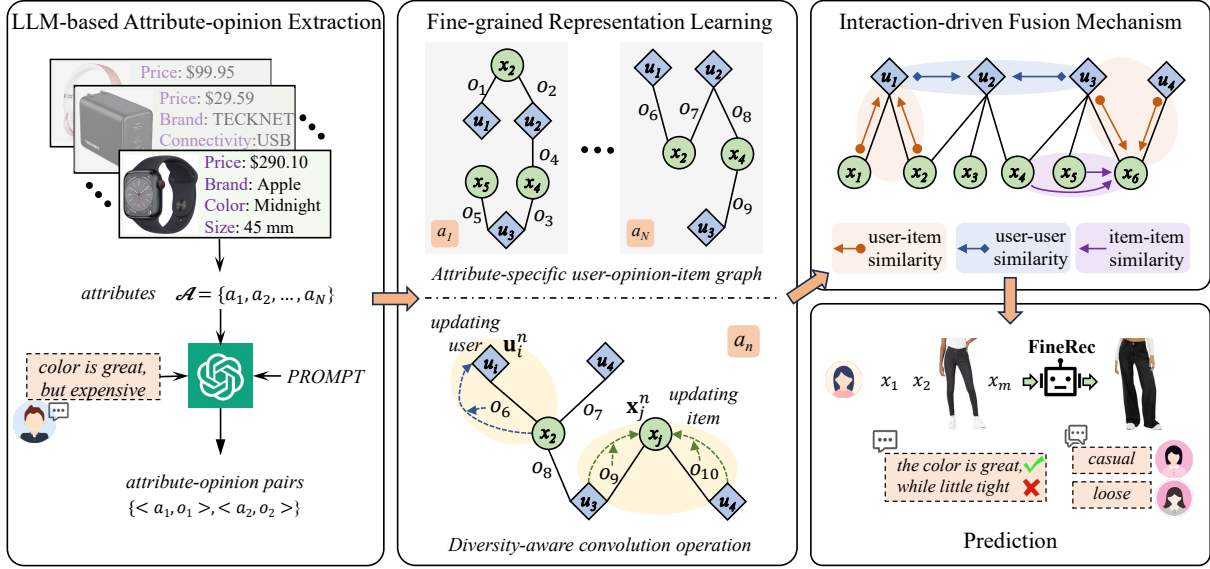
Figure 2: The workflow of FineRec. We employ LLM to extract attribute-opinion pairs from user-item reviews. Afterwards, an attribute-specific user-opinion-item graph is created in each attribute, where a diversity-aware convolution operation is devised to conduct information aggregating. Interaction-driven fusion mechanism leverages user-item interactions to integrate attribute-specific item/user representations. Finally, FineRec generates recommendations in the fine-grained manner.

$[x_1, x_2, \cdots, x_m]$, where $x_j \in \mathcal{X}$ and m is the sequence length. The review posted by user u_i on item x_j is denoted as $r_{ij} \in \mathcal{R}$. In a review, the attribute-opinion pairs depict user preferences and item characteristics in a fine-grained manner. In this study, therefore, we propose to finely represent users and items via exploring the attribute-opinion pairs from reviews. Accordingly, we can forecast the user next interacted item x_{m+1} based on $\mathcal{S}^{u_i}$.

4 THE PROPOSED APPROACH

4.1 Overview of the FineRec

As illustrated in Figure 2, the proposed FineRec mainly consists of the following interconnected components: (1) **LLM-based attribute-opinion extraction** identifies informative attributes based on their importance and subsequently extracts the attribute-opinion pairs from reviews via LLM; (2) **Fine-grained representation learning** obtains fine-grained user and item embeddings via exploring attribute-opinion pairs, where *an attribute-specific user-opinion-item graph* is created under each attribute, with *a diversity-aware convolution operation* conducting information aggregating; (3) **Interaction-driven fusion mechanism** leverages user-item interaction information to guide the fusion of attribute-specific item/user representations; and (4) **Prediction** provides recommendations in the fine-grained manner.

4.2 LLM-based Attribute-opinion Extraction

Large Language Models (LLMs), benefiting from their training on massive text corpora, have become reservoirs of vast language knowledge. This foundation empowers LLMs with remarkable proficiency in various tasks of Natural Language Processing (NLP) [1, 11, 28]. Leveraging this capability, we utilize LLMs for

extracting attribute-opinion pairs from reviews. In the context of recommendation, the item attributes that users care about are heavily overlapping, that is, a small number of important attributes significantly influence user choice like item price [43]. Besides, LLMs can sometimes suffer from "hallucinations" where they perform unsatisfactorily in addressing complex issues [29]. In light of these facts, we present to obtain informative attributes based on their importance, and subsequently extract corresponding opinions via LLMs. Such a paradigm enables FineRec to focus on important attributes and greatly simplifies the extraction task.

Specifically, we gather item attributes directly from E-commerce websites, with an assumption that the displayed attributes are important. Adhering to the Pareto principle, we focus on the most prevalent attributes, selecting the top N attributes based on their frequency of occurrence to build the attribute set $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$, where N is the number of extracted attributes. Note that, each a_i represents an informative attribute like 'Price', 'Color' and so on.

A well-crafted prompt is the key to unleash the potential of LLMs. Building upon the extracted attributes, thus, we develop a specialized prompt template to facilitate the LLM in extracting attribute-opinion pairs from reviews as follows,

PROMPT. *Please extract user opinion words towards the attribute "a_n" from the review: "review contents". Only return the opinion words! Your answer should be short.*

where $a_n \in \mathcal{A}$ and *"review contents"* denotes a specific review. For a review r_{ij}, the LLM outputs a sentence that captures the opinion of a user (u_i) about an item (x_j) under the attribute (a_n), denoted as $o_{ij}^n \in O$, where O is opinion set. If the review does not mention any attribute in $\mathcal{A}$, the result is disregarded. Note that, we apply LLM to conduct the extraction across all reviews for every attribute,

forming the attribute-opinion pairs. For instance, consider a review $r_{ij} \in \mathcal{R}$ from user u_i about item x_j stating, "*it smells nice, but too expensive*", with the attribute set $\mathcal{A} = \{Scent, Price, Brand\}$, the extracted attribute-opinion pairs would be { <Scent, nice>, <Price, expensive> }. These extracted pairs allow us to finely represent users and items based on their specific attributes and opinions.

4.3 Fine-grained Representation Learning

4.3.1 Attribute-specific user-opinion-item graph.
Generally, a/an user/item holds distinct preferences/characteristics under different attributes. Therefore, we create an attribute-specific user-opinion-item graph to finely-grained encode users/items under each attribute. Specifically, for an attribute a_n, a unique user-opinion-item graph $\mathcal{G}^n = (\mathcal{U}^n \bigcup \mathcal{X}^n, O^n)$ is built. In this graph, $\mathcal{U}^n \bigcup \mathcal{X}^n$ is the node set consisting of users and items, where $\mathcal{U}^n \subseteq \mathcal{U}$ and $\mathcal{X}^n \subseteq \mathcal{X}$. The $O^n \subseteq O$ is the edge set encapsulating opinions expressed by users about items regarding the specific attribute. As shown in Figure 2, within the graph $\mathcal{G}^n$ of the attribute a_n, a user $u_i \in \mathcal{U}^n$ and an item $x_j \in \mathcal{X}^n$ are connected by an opinion edge $o^n_{ij} \in O^n$ if the user expressed her opinion o^n_{ij} to the item in relation to the attribute. Note that, the $\mathcal{G}^n$ will not record the user and item if there is not an opinion between them under the attribute. Moreover, the attribute-specific user-opinion-item graph captures interaction and opinion information between users and items, providing the potentials to discern fine-grained user preferences and item characteristics specific to each attribute.

To represent various user/item preferences/characteristics under different attributes, we employ distinct embeddings to represent users/items in each attribute. To elaborate, a user u_i is separately represented by N unique embeddings as $\{\mathbf{u}^1_i, \mathbf{u}^2_i, \cdots, \mathbf{u}^N_i\}$, where $\mathbf{u}^n_i \in \mathbb{R}^d$ indicates the user preferences about attribute $a_n \in \mathcal{A}$. The same applies to items, *i.e.*, each item x_j is represented as $\{\mathbf{x}^1_j, \mathbf{x}^2_j, \cdots, \mathbf{x}^N_j\}$, where $\mathbf{x}^n_j \in \mathbb{R}^d$ denotes the item characteristics in attribute a_n. Besides, we employ pre-trained BERT [8] to represent an attribute a_n as $\mathbf{a}^n \in \mathbb{R}^d$ and an opinion text o^n_{ij} as $\mathbf{o}^n_{ij} \in \mathbb{R}^d$.

4.3.2 Diversity-aware convolution operation.
The extensive diversity of opinions impedes fine-grained user and item representation learning through attribute-opinion manner. Specifically, for a given attribute, a user may have varying opinions about different items, and similarly, an item might receive a range of opinions from different users. To address this diversity, we devise a diversity-aware convolution operation to conduct information aggregating on each attribute-specific user-opinion-item graph. Formally, we update a user's embedding $\mathbf{u}^n_i$ on the $\mathcal{G}^n$ of attribute $\mathbf{a}^n$ via,

$$\mathbf{u}^n_i := \mathbf{u}^n_i + \sum_{x^n_j \in \mathcal{X}^n_{u_i}} \alpha_j (\mathbf{x}^n_j + \mathbf{o}^n_{ij}) \quad (1)$$

where $\mathcal{X}^n_{u_i} \subseteq \mathcal{X}^n$ consists of items adjacent with the user in the graph $\mathcal{G}^n$, while $\mathbf{o}^n_{ij}$ is embedding of corresponding opinion. It's important to highlight that our method updates user embeddings by jointly considering both the item and its associated opinion (instantiated with an additive operation). This approach allows our FineRec to capture the diversity of a user's opinions towards different items. The importance of different item-opinion pairs is determined by,

$$\alpha_j = \frac{sim(\mathbf{u}^n_i, (\mathbf{a}^n + \mathbf{x}^n_j))}{\sum_{x^n_k \in \mathcal{X}^n_{u_i}} sim(\mathbf{u}^n_i, (\mathbf{a}^n + \mathbf{x}^n_k))}, \quad (2)$$

where the $sim()$ is cosine similarity. To highlight the attribute's influence on user behaviors, we employ the attribute embedding $\mathbf{a}^n$ to calculate the importance. Similarly, we update an item embedding $\mathbf{x}^n_j$ on the attribute-specific graph $\mathcal{G}^n$ as follows,

$$\mathbf{x}^n_j := \mathbf{x}^n_j + \sum_{u^n_i \in \mathcal{U}^n_{x_j}} \beta_i (\mathbf{u}^n_i + \mathbf{o}^n_{ij}) \quad (3)$$

$$\beta_i = \frac{sim(\mathbf{x}^n_j, (\mathbf{a}^n + \mathbf{u}^n_i))}{\sum_{u^n_k \in \mathcal{U}^n_{x_j}} sim(\mathbf{x}^n_j, (\mathbf{a}^n + \mathbf{u}^n_k))} \quad (4)$$

where $\mathcal{U}^n_{x_j} \subseteq \mathcal{U}^n$ contains users expressing opinions on the item under the attribute a_n.

4.4 Interaction-driven Fusion Mechanism

The future actions of users are jointly influenced by their preferences and items' characteristics across various attributes. However, capturing the intricate patterns of the influence is challenging, primarily because there are no explicit signals indicating how each attribute impacts user decisions. Fortunately, the user-item interactions provide insights into these complex influencing patterns. For example, when a user selects certain items from a vast choices, it reflects a match between her overall preferences and the characteristics of those items. In fact, these interactions hint at the similarity at the embedding level, considering neural models operate based on the similarity of representations. Thus, we propose to employ the similarity implied in the user-item interactions to integrate attribute-specific item/user representations.

To intuitively illustrate user-item interaction relations, we present a global user-item interaction graph in the upper right part of Figure 2. As shown in the figure, user-item interactions manifest three types of similarities: (1) user-item similarity, indicating the commonality between a user and her interacted items; (2) user-user similarity, where users buying the same item can be considered similar; (3) item-item similarity, where items interacted with the same user are deemed similar. We start by concatenating attribute-specific item and user embeddings as,

$$\hat{\mathbf{u}}_i = [\mathbf{u}^1_i; \mathbf{u}^2_i; \cdots; \mathbf{u}^N_i] \quad (5)$$

$$\hat{\mathbf{x}}_j = [\mathbf{x}^1_j; \mathbf{x}^2_j; \cdots; \mathbf{x}^N_j] \quad (6)$$

where $[;]$ denotes concatenation operation, and $\hat{\mathbf{u}}_i, \hat{\mathbf{x}}_j \in \mathbb{R}^{Nd}$.

We find that, interestingly, three types of similarity can be conceptualized as adjacent relationships within the global user-item graph. Specifically, the user-item similarity represents direct, one-hop connections in the graph, while user-user and item-item similarities correspond to two-hop relationships. In light of aggregating algorithms in GNN, we integrate various preferences of the user (u_i) on different attributes as,

$$\mathbf{u}_i = \hat{\mathbf{u}}_i + \frac{1}{|\mathcal{X}_{u_i}|} \sum_{x_j \in \mathcal{X}_{u_i}} W_1 \hat{\mathbf{x}}_j + \frac{1}{|\mathcal{U}_{u_i}|} \sum_{u_k \in \mathcal{U}_{u_i}} W_2 \hat{\mathbf{u}}_k, \quad (7)$$

where $W_1, W_2 \in \mathbb{R}^{Nd \times Nd}$ are trainable matrices enabling attribute-specific embeddings to interact with each other, $\mathcal{X}_{u_i} \subseteq \mathcal{X}$ are items

Table 1: Statistics of datasets.

Datasets	Cellphones	Beauty	Sports	Yelp
#item	6,208	10,176	11,017	12,391
#user	7,598	15,152	11,817	12,373
#interaction	50,140	123,148	87,594	110,313
avg.length	6.60	8.13	7.41	8.92

Table 2: Used attributes in each dataset.

Datasets	Attribute list
Cellphones	Battery, Brand, Color, Connectivity, Performance, Price, Size
Beauty	Brand, Color, Effectiveness, Ingredients, Price, Scent, Size
Sports	Brand, Comfort, Functionality, Material, Price, Quality, Size
Yelp	Ambience, Cleanliness, Food, Location, Parking, Price, Service

interacted by the user u_i, $\mathcal{U}_{u_i} \subseteq \mathcal{U}$ contains users buying the same items as the user u_i, and $\mathbf{u}_i$ is the final user representation. Note that, $\mathcal{X}_{u_i}$ indicates user-item similarity, while $\mathcal{U}_{u_i}$ denotes user-user similarity. In Eq.(7), for a user, we inject the information of its similar items and users into its embedding. This method highlights the similarities revealed through user-item interactions, urging similar users and items closer in the embedding space. By doing so, it guides the fusion for representations from various attributes. Similarly, we conduct the fusion for an item (x_j) as,

$$\mathbf{x}_j = \hat{\mathbf{x}}_j + \frac{1}{|\mathcal{U}_{x_j}|} \sum_{u_i \in \mathcal{U}_{x_j}} W_3 \hat{\mathbf{u}}_i + \frac{1}{|\mathcal{X}_{x_j}|} \sum_{x_k \in \mathcal{X}_{x_j}} W_4 \hat{\mathbf{x}}_k, \qquad (8)$$

where $\mathcal{U}_{x_j} \subseteq \mathcal{U}$ denotes users interacting with the item x_j, $\mathcal{X}_{x_j} \subseteq \mathcal{X}$ consists of items co-interacted by users with item x_j, and $\mathbf{x}_j$ is the final item representation. Besides, $\mathcal{U}_{x_j}$ indicates user-item similarity, while $\mathcal{X}_{x_j}$ denotes item-item similarity.

4.5 Prediction

Recent actions of a user can reflect her recent interest. Following [18, 35, 38], thus, we represent a user's recent interest $\bar{\mathbf{u}}_i$ by conducting average-pooling on her last l interacted items as,

$$\bar{\mathbf{u}}_i = \frac{1}{l} \sum_{k=0}^{l-1} \mathbf{x}_{m-k}. \qquad (9)$$

Based on user embedding $\mathbf{u}_i$, her recent interest $\bar{\mathbf{u}}_i$ and item embedding $\mathbf{x}_j$, we can forecast next behavior of the user as,

$$y_j = (\mathbf{u}_i + \bar{\mathbf{u}}_i)^\top \mathbf{x}_j, \qquad (10)$$

where y_j is the interacted score predicted for a candidate item x_j. To ensure a fair comparison, we follow the training paradigm of full ranking on the entail item set as in [46, 47]. Formally, the FineRec is trained via cross-entropy loss as follows,

$$\mathcal{L}(\mathbf{p}, \mathbf{y}) = -\sum_{j=1}^{|\mathcal{X}|} p_j \log(y_j) + (1 - p_j) \log(1 - y_j) \qquad (11)$$

where p_j is the ground truth indicating whether the user purchases item x_j.

5 EXPERIMENTAL SETUP

5.1 Research Questions

We conduct extensive experiments on several real-world datasets to evaluate the proposed FineRec and all baselines, with the focus on answering the following research questions:

- **RQ1**: How does our FineRec perform compared with existing state-of-the-art methods? (ref. Section 6.1)

- **RQ2**: What is the effect of the fine-grained attribute-opinion manner in handling the task? (ref. Section 6.2)
- **RQ3**: Does each proposed technique contribute positively to FineRec's performance? (ref. Section 6.3-6.4)
- **RQ4**: What is the influence of the key hyper-parameters on FineRec? (ref. Section 6.5)
- **RQ5**: How well dose the proposed FineRec work in the real-world instance? (ref. Section 6.6)

5.2 Datasets and Preprocessing

To scrutinize the effectiveness of our FineRec, we employ the following four popular public datasets in our experiments:

- **Cellphones**, **Beauty** and **Sports** are three datasets covering different domains in Amazon[1]. As widely used benchmarks for the sequential recommendation [7, 18], these datasets contain users' purchasing behavior sequences and corresponding user-item reviews.
- **Yelp** [2] containing users' reviews for restaurants is a popular dataset used in the task [7, 23]. As in [23, 24], we retain the transaction records of the year 2019 in this dataset.

Following [13, 16, 18, 30], we apply the 5-core method to preprocess these datasets, where items and users with less than 5 interactions are filtered out. Besides, to fairly examine the impact of finely-grained modeling, we eliminate reviews that fail to mention the attributes included in the attribute list $\mathcal{A}$. In line with common practices [17, 47], *leave-one-out* operation is used to split these datasets, where the last item in a sequence is used for testing, the penultimate item for validation, and the remaining for training. We present the statistical details of all four datasets in Table 1.

5.3 Evaluation Metrics

Following existing works [10, 18, 30, 47], we evaluate the performance of all methods with following two metrics:

- **Prec@k**: Precision measures the proportion of cases where the ground-truth item is within the recommendation list.
- **NDCG@k**: Normalized Discounted Cumulative Gain considers the rank of the ground truth item among the recommendation list.

It is worth noting that the Prec@k metric does not take into account the ranking of an item, as long as it appears within the top-k recommendations. Conversely, NDCG@k emphasizes the item rank, which is crucial in scenarios where the order of recommendations matters. In this study, we report the results with $k = 10$ and 20.

---

[1] http://jmcauley.ucsd.edu/data/amazon/
[2] https://www.yelp.com/dataset

Table 3: Experimental results (%) of different methods on four datasets. Bold scores are the best performance among all methods, while underlined scores are the second best. Improvements of FineRec over the best baseline (*) are statistically significant with t-test ($p < 0.01$).

Datasets	Metrics	SKNN	NARM	SASRec	SR-GNN	RNS	ICLRec	UniSRec	A-Mixer	MCLRec	ACTSR	FineRec	*impro.*
Cellphones	Prec@10	2.54	3.23	3.95	4.21	4.88	3.59	4.35	4.04	5.83	6.12	**7.66***	25.16%
	NDCG@10	1.45	1.74	1.80	2.02	3.18	1.54	2.28	2.03	2.78	2.91	**3.87***	21.70%
	Prec@20	4.01	4.50	6.11	6.03	7.09	5.27	6.29	5.46	8.47	8.84	**11.89***	34.50%
	NDCG@20	1.88	2.06	2.42	2.53	3.79	2.05	2.66	2.34	3.38	3.55	**4.75***	25.33%
Beauty	Prec@10	1.73	2.61	3.97	2.25	2.87	1.97	3.71	3.47	4.41	4.16	**5.71***	29.48%
	NDCG@10	1.01	1.51	1.77	1.11	1.52	1.22	1.87	1.91	2.08	2.05	**2.92***	40.38%
	Prec@20	2.68	3.89	6.64	3.69	4.63	3.04	6.23	5.17	7.93	7.23	**9.25***	16.65%
	NDCG@20	1.76	1.83	2.40	1.52	2.01	1.99	2.14	2.29	2.80	2.62	**3.77***	34.64%
Sports	Prec@10	1.13	1.48	1.69	1.88	2.44	1.59	1.85	1.44	2.85	2.57	**3.50***	22.81%
	NDCG@10	0.73	0.81	0.94	0.91	1.27	0.90	0.93	0.88	1.17	1.02	**1.88***	48.03%
	Prec@20	1.68	1.84	2.63	2.90	3.11	2.31	2.49	1.87	4.44	4.09	**5.45***	22.75%
	NDCG@20	0.91	0.95	1.20	1.12	1.69	1.03	1.09	1.17	1.58	1.37	**2.37***	40.24%
Yelp	Prec@10	1.45	1.95	3.79	2.94	2.22	1.51	1.89	1.80	4.54	4.18	**5.63***	24.01%
	NDCG@10	0.82	1.26	1.84	1.54	1.78	1.02	1.22	1.11	2.15	1.76	**3.09***	43.72%
	Prec@20	2.17	2.65	4.95	3.71	3.65	2.29	2.62	2.53	7.94	7.59	**9.16***	15.37%
	NDCG@20	1.30	1.56	2.40	1.73	1.93	1.28	1.60	1.30	3.02	2.51	**3.69***	22.19%

5.4 Baselines

To evaluate the performance of the proposed FineRec, we use following ten competitive methods in this study as baselines:

- **SKNN** recommends items based on the similarity between the current session and the other sessions.
- **NARM** [19] utilizes RNNs with an attention mechanism to capture the user's main purpose.
- **SASRec** [16] employs the self-attention mechanism to capture sequential patterns of user behaviors.
- **SR-GNN** [35] builds session graphs and applies graph neural networks to capture item transitions.
- **RNS** [18] is a representative method incorporating reviews into the task to represent users and items.
- **ICLRec** [5] mines user latent intents with contrastive learning for enhancing user behavior understanding.
- **UniSRec** [13] represents items with description text for obtaining universal sequence representations.
- **A-Mixer** [38] (*i.e.,* Atten-Mixer) leverages multi-level user intents to achieve multi-level reasoning on item transitions.
- **MCLRec** [30] presents meta-optimized contrastive learning to achieve informative data and model augmentations.
- **ACTSR** [47] enhances Transformer-based methods by calibrating unreliable attention weights in transformer layers.

5.5 Implementation Details

We employ ChatGPT-3.5[3] to conduct attribute-opinion extraction. Based on attribute popularity, the attributes utilized in each dataset are outlined in Table 2. To strike a balance between efficiency and

effectiveness, we focus on seven most popular attributes for each dataset. In order to mitigate the randomness of LLM, we execute the opinion extraction process five times, considering the union of these outputs as the final attribute-opinion pairs.

To ensure a fair comparison, the hyper-parameters of FineRec and all baselines are determined via grid search according to their performance on Prec@20 in the validation set. In our FineRec, we investigate the dimension of attribute-specific user/item embedding d in {4, 8, 16, 32}. As in [18], the averaged embeddings of last 5 items ($l=5$) is used to represent a user's recent interest. We set the mini-batch size to 512 and use the Adam optimizer with an initial learning rate of 0.001 to optimize FineRec. We have released all source codes for broader accessibility and reproducibility[4].

6 RESULTS AND ANALYSIS

6.1 Overall Performance (RQ1)

The performance of all baselines and FineRec on four datasets is detailed in Table 3, where we can obtain the following insights:

Firstly, the performance of baselines varies greatly across different datasets. Taking the ACTSR as an example, it shows best results among all baselines in Prec metric on Cellphones, while obtaining inferior performance in other contexts. Similarly, SASRec excels on Beauty dataset but performs poorly on others. These facts signify the complexity and difficulty of the sequential recommendation task, which requires capturing users' evolving interests from their historical behaviors. Moreover, most existing methods handle the task in an implicit way, primarily focusing on users' implicit clicking behaviors. Such a manner, unfortunately, often

---

[3]https://chat.openai.com

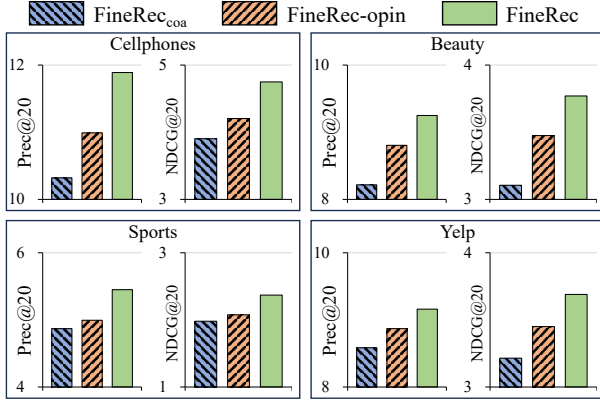[4]https://github.com/Zhang-xiaokun/FineRec

Figure 3: The effect of fine-grained manner on SR.

fails to grasp user genuine intents, thereby significantly limiting their effectiveness.

Secondly, RNS, as a representative work incorporating reviews into SR, exhibits competitive performance in certain scenarios, *i.e.,* with the NDCG metric in Cellphones and Sports. Instead of modeling implicit user actions, RNS explicitly mines user interests from their reviews, enhancing its grasp of user intents. Nevertheless, its performance is less satisfactory on the Beauty and Yelp datasets. The primary drawback of RNS lies in its coarse paradigm for modeling a review as a whole. Indiscriminately blending distinct attitudes towards various attributes within a single review, RNS struggles to capture a user's specific preferences for each attribute. This leads to its inconsistent performance across different datasets.

Thirdly, the most recent approaches, namely MCLRec and ACTSR, generally outperform other baselines. Their impressive performance could be largely attributed to their utilization of cutting-edge techniques. Specifically, MCLRec leverages meta-optimized contrastive learning to enrich user behavior data. As to ACTSR, it calibrates attention weights within Transformer architectures to better adapt to the task. However, despite equipping with advanced techniques, these methods are unable to obtain consistent performance across various contexts. This once again suggests the limitations of implicit manner for modeling user behaviors in the task of sequential recommendation.

Finally, the proposed FineRec achieves consistent improvements over all baselines in terms of all metrics on all datasets, which demonstrates its effectiveness for sequential recommendation. In particular, FineRec outperforms the best baselines in Prec@20 and NDCG@20 by 34.50% and 25.33% on Cellphones, 16.65% and 34.64% on Beauty, 22.75% and 40.24% on Sports, and 15.37% and 22.19% on Yelp. We believe that the consistent superiority of FineRec over current start-of-the-art methods comes from its fine-grained manner in handling the task. Benefiting from finely representing users and items via attribute-opinions, FineRec can identify fine-grained user preferences and item characteristics on various attributes. This manner enhances the prediction accuracy, significantly contributing to the overall effectiveness of the proposed FineRec.

6.2 The effect of fine-grained manner for handling SR (RQ2)

The key innovation of our FineRec lies in its fine-grained manner in handling SR, achieved via exploring attribute-opinions from user-item reviews. In order to validate the effectiveness of this novel paradigm, the following variants of FineRec are designed: "FineRec$_{coa}$" builds a single user-review-item graph based on the entire user-item interactions. In this graph, users and items form the nodes, and their reviews serve as the edges. That is, FineRec$_{coa}$ coarsely models a review as a whole, without distinguishing users' various opinions on different attributes. Besides, "FineRec-opin" creates an attribute-specific user-item graph for each attribute, without considering the related opinions between users and items. It relies on the conventional GCN to update user and item embeddings under each attribute.

From Figure 3, we can obtain the following key points: (1) FineRec-opin achieves better performance than FineRec$_{coa}$. It underscores the rationality of building distinct attribute-specific subgraphs for various attributes. This distinction aligns with the fact that users exhibit varied preferences and items display different characteristics across various attributes. By recognizing these differences, FineRec-opin obtains improvements over FineRec$_{coa}$; (2) FineRec-opin is defeated by FineRec with a large margin, which highlights the significance of jointly exploring attributes and opinions. For an attribute, users may express completely different attitudes. Thus, only considering attributes while ignoring corresponding opinions, FineRec-opin fails to make accurate predictions; and (3) The FineRec surpasses both its variants in all cases, validating the efficacy of our fine-grained manner in handling the task. By delving into the attribute-opinions within reviews, FineRec is able to uncover user preferences and item characteristics, significantly improving the recommendation performance.

6.3 The effect of diversity-aware convolution operation (RQ3)

To tackle the opinion diversity issue in fine-grained user and item representation learning, we introduce a diversity-aware convolution operation. The variant "w/o diver" discards the diversity-aware convolution operation. Instead, it employs a straightforward approach by summing all adjacent opinion embeddings and item/user embeddings to update the user/item embeddings. That is, it omits the diversity of opinions. Additionally, we include the results of the best-performing baselines for each metric to provide a comprehensive comparison.

Table 4 clearly shows that FineRec consistently outperforms the "w/o diver" variant in all scenarios. It demonstrates the effectiveness of our specially devised diversity-aware convolution operation. By focusing on the diversity of opinions in user/item representation learning, this operation achieves two critical objectives: (1) grasping user varying opinions on distinct items, and (2) providing comprehensive portrayal of item characteristics based on various user opinions. Consequently, this operation enables FineRec to obtain robust user/item representations under each attribute. In addition, the underperformance of the best baselines compared to the "w/o diver" variant further indicates the efficiency of our fine-grained manner in tackling sequential recommendation.

Table 4: The effect of diversity-aware convolution operation.

Method	Cellphones		Beauty		Sports		Yelp	
	Prec@20	NDCG@20	Prec@20	NDCG@20	Prec@20	NDCG@20	Prec@20	NDCG@20
Best baselines	8.84	3.79	7.93	2.80	4.44	1.69	7.94	3.02
w/o diver	11.04	4.29	8.23	3.18	5.05	2.04	8.73	3.47
FineRec	**11.89***	**4.75***	**9.25***	**3.77***	**5.45***	**2.37***	**9.16***	**3.69***
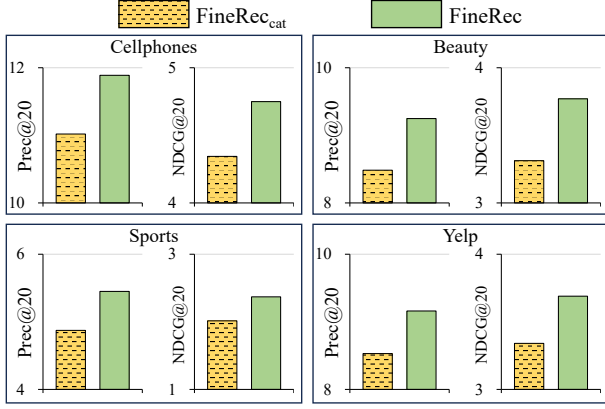
Figure 4: The effect of interaction-driven fusion mechanism.
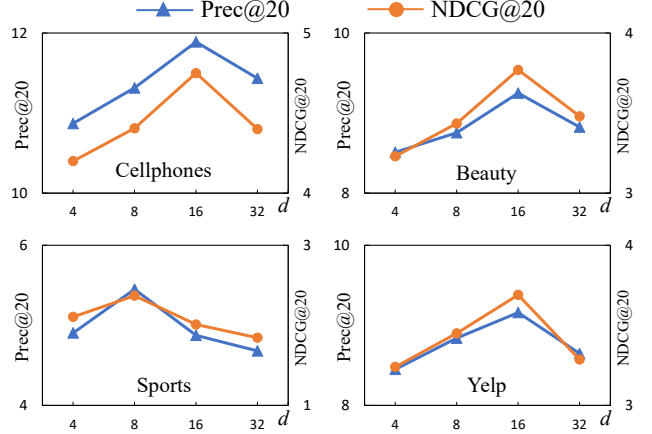

Figure 5: The impact of hyper-parameter d.

6.4 The effect of interaction-driven fusion mechanism (RQ3)

To comprehensively understand user behaviors, we present an interaction-driven fusion mechanism that integrates user/item representations across all attributes. This mechanism employs informative user-item interactions to guide the fusion of attribute-specific user/item embeddings. In contrast, the FineRec$_{cat}$ variant employs a conventional approach for the fusion. It concatenates the embeddings from different attributes and then merges them using a Multi-Layer Perception (MLP), bypassing our specialized fusion mechanism.

As can be observed from Figure 4, FineRec achieves better performance over FineRec$_{cat}$. It validates the effectiveness of our interaction-driven fusion mechanism in integrating attribute-specific embeddings. We believe that the user-item interactions reflect the relationships, notably the similarity at the embedding-level, between users and items. By harnessing these similarities, the designed fusion mechanism gains a deeper understanding of how various attributes influence user behaviors. Therefore, the proposed mechanism facilitates an effective fusion process, thereby establishing the superiority of FineRec over FineRec$_{cat}$.

6.5 Hyper-parameter Study (RQ4)

In this section, we investigate the influence of the main hyper-parameter d on FineRec's performance. Parameter d denotes the dimension of attribute embeddings, opinion embeddings, as well as attribute-specific user/item embeddings. The performance trends of FineRec across various d values, evaluated using Prec@20 and NDCG@20 metrics on all datasets, are illustrated in Figure 5. The

performance curve reveals that as d increases, FineRec initially shows improvements but eventually experiences a decline. This pattern suggests that while a larger d enhances representation capability, leading to better performance, an excessively high d might cause overfitting, thereby degrading performance. Interestingly, we find that a relatively small d value suffices for optimal performance in FineRec, with d set to 8 for Sports, and 16 for Cellphones, Beauty, as well as Yelp datasets. Our FineRec aims to capture specific user/item preferences/characteristics on each attribute, instead of indiscriminately modeling them as a whole like existing methods. Benefiting from such a fine-grained manner, FineRec can accurately represent users or items under each attribute at a low cost (*i.e.,* a small number of embedding dimension). It is both efficient and effective. We believe that this merit of FineRec significantly enhances its practical applicability.

6.6 Case Study (RQ5)

In order to intuitively scrutinize the effects of FineRec in generating recommendations, we randomly select an instance from the dataset Sports. This real-world case is presented in Figure 6, where we detail a user's historically interacted items, reviews he posted for each item, the attribute-opinion pairs extracted using LLM, and the ground truth item. Additionally, we display recommended items from FineRec at the low part of the figure. For these recommended items, we include three reviews written by other users, along with their respective attribute-opinion pairs. Due to space constraints, we present the top-3 recommended items of FineRec.

User history behaviors:

User: NQ3YTUI	Item: 9JZ92A	Item: TALNRS	Item: 2UNMERW	Item: H8ZO1U
Posted Reviews	I wear these under my cargo shorts and they are so comfortable. the only thing is it was more costly than I expected.	I can not say I love the products. They're cheap, but I'm disappointed with the quality of this product.	Always have bought these bottles for cycling. A cheap investment that's so worth it. They'll last you forever.	Condor's products are first rate, I will continue to order more of these.
Attribute-opinions	\<Comfort, Very comfortable\> \<Price, more costly than expected\>	\<Price, Cheap\> \< Quality, Disappointed \>	\<Price, Cheap investment\>	\<Brand, Condor's products are first rate\>

Ground truth:

Groud truth Item: RURGYQ
This case is way more than I expected. I love it. It's made with awesome quality.
\<Quality, Awesome\>

FineRec's Recommended Items:

Top-1: RURGYQ ✔	Reviews of other users	The case is perfect. Its quality is really impressive. Water resistant, price Is just right!	The price for this quality is unbeatable! Buy it! It fits incredibly well, much more than I expected.	Works well and is well constructed. Good case for its money.
	Attribute-opinions	\<Quality, Really impressive\> \<Functionality, Water-resistant\> \<Price, Just right\>	\<Price, Unbeatable\> \<Quality, High\> \<Size, Fits incredibly well\>	\<Price, Good value for its money\> \<Functionality, Works well\> \<Quality, Well constructed\>
Top-2: 9OU6NK	Reviews of other users	The hat is made from a very durable material, and it looks great. I will buy more products made by Condor in the future.	Perfect 5 star product! This Condor's hat is super light weight and breaths well. This helps allow heat to escape and keep cool. Fit me perfectly right out of the box with no adjustments.	Great hat, well made. Love this hat, you will too, it is a very well built hat. Id buy it again.
	Attribute-opinions	\<Brand, Positive towards Condor\> \<Material, Very durable\> \<Quality, Looks great\>	\<Brand, Condor - 5 star product\> \<Comfort, Fits perfectly\> \<Size, Perfect fit\>	\<Quality, Well-made\>
Top-3: 6Z08YM	Reviews of other users	Very beautiful design, easy to use as well. If you need a tough knife, this is a can't-miss.	For the money, this is a great knife. steel holds a decent edge and sharpens easy. Good buy!	It's so affordable yet so good. The size is just right! It's like it was sized just for me!
	Attribute-opinions	\<Quality, Tough\> \<Functionality, easy to use\>	\<Price, Great value\> \<Quality, Steel holds a decent edge and sharpens easily\>	\<Price, Affordable\> \<Quality, Good\> \<Size, Just right sized personally\>

Figure 6: A case from the Sports dataset. We present user-item reviews (Blue), corresponding attribute-opinions (Orange), and the recommended items of FineRec with other users' reviews.

The following observations are noted in Figure 6: (1) The LLM-based attribute-opinion extraction can effectively derive attribute-opinion pairs from user-item reviews. As shown in the case, the extracted attribute-opinions are generally informative and accurate. We believe that this efficiency likely stems from the following facts: (a) the LLM's extensive training on massive text data encapsulates it with rich language knowledge; and (b) our approach of directing the LLM towards specific attributes instead of from scratch to generate attribute-opinions alleviates its inherent "hallucinations". (2) Our proposed fine-grained manner does facilitate effective recommendations. As shown in the case, the attributes the user cares about are hit by our FineRec. More specifically, the attributes mentioned by the user are mostly overlapped with those of items provided by FineRec, *i.e.,* {Price, Quality}. This success is reasonable since that our FineRec represents users and items with fine-grained attribute-opinions instead of conventional implicit ID number. More surprisingly, based on user specific attribute-opinions, our FineRec can provide highly personalized recommendations. Concretely, the user expressed "disappointed" to the "Quality" of the item "TAL-NERS". Our FineRec can correctly recommend ground truth item "RURGYQ" which has been rated by other users as of "really impressive", "high" and "well constructed" Quality. Evidently, such type of recommendation can greatly improve user satisfaction. (3) The proposed FineRec can effectively discern fine-grained user preferences and item characteristics, offering relevant recommendations accordingly. For example, for the user who showed a preference for the "Condor" brand in his reviews, FineRec is able to recommend an item associated with the same brand, where the item brand is identified from other reviews.

7 CONCLUSION AND FUTURE WORK

In this study, we propose a novel framework FineRec to explore fine-grained sequential recommendation via mining the attribute-opinions from reviews. Specifically, FineRec employs a Large Language Model to extract attribute-opinions from reviews. For each attribute, we create an attribute-specific user-opinion-item graph, encoding fine-grained user preferences and item characteristics. Within these graphs, a diversity-aware convolution operation is devised to achieve fine-grained user and item representations, particularly handling the diversity of opinions. Lastly,an interaction-driven fusion mechanism is employed, harnessing user-item interactions to integrate attribute-specific embeddings for generating final recommendations. Comprehensive experiments on four real-world datasets highlight FineRec's superiority over current state-of-the-art methods. Further analysis also verifies the effectiveness of our finely-grained manner for the task.

As to future work, our fine-grained approach opens up exciting avenues for enhancing the interpretability of recommendation tasks. The prospect of providing explanations rooted in the fine-grained user preferences and item characteristics is both attractive and promising. Such a perspective could offer a more fine-grained and personalized understanding of recommendations, significantly enhancing the user experience.

ACKNOWLEDGEMENT

REFERENCES

[1] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2023. TALLRec: An Effective and Efficient Tuning Framework to Align Large Language Model with Recommendation. In *RecSys*. ACM, 1007–1014.

[2] Renqin Cai, Jibang Wu, Aidan San, Chong Wang, and Hongning Wang. 2021. Category-aware Collaborative Sequential Recommendation. In *SIGIR*. 388–397.

[3] Mohna Chakraborty, Adithya Kulkarni, and Qi Li. 2022. Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction. In *SIGKDD*, Aidong Zhang and Huzefa Rangwala (Eds.). ACM, 66–75.

[4] Qian Chen, Zhiqiang Guo, Jianjun Li, and Guohui Li. 2023. Knowledge-enhanced Multi-View Graph Neural Networks for Session-based Recommendation. In *SIGIR*. ACM, 352–361.

[5] Yongjun Chen, Zhiwei Liu, Jia Li, Julian J. McAuley, and Caiming Xiong. 2022. Intent Contrastive Learning for Sequential Recommendation. In *WWW*. 2172–2182.

[6] Zhiyong Cheng, Ying Ding, Xiangnan He, Lei Zhu, Xuemeng Song, and Mohan S. Kankanhalli. 2018. A^3NCF: An Adaptive Aspect Attention Model for Rating Prediction. In *IJCAI*. ijcai.org, 3748–3754.

[7] Jin Yao Chin, Kaiqi Zhao, Shafiq R. Joty, and Gao Cong. 2018. ANR: Aspect-based Neural Recommender. In *CIKM*. ACM, 147–156.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*. 4171–4186.

[9] Xin Dong, Jingchao Ni, Wei Cheng, Zhengzhang Chen, Bo Zong, Dongjin Song, Yanchi Liu, Haifeng Chen, and Gerard de Melo. 2020. Asymmetrical Hierarchical Networks with Attentive Interactions for Interpretable Review-Based Recommendation. In *AAAI*. AAAI Press, 7667–7674.

[10] Xinyu Du, Huanhuan Yuan, Pengpeng Zhao, Jianfeng Qu, Fuzhen Zhuang, Guanfeng Liu, Yanchi Liu, and Victor S. Sheng. 2023. Frequency Enhanced Hybrid Attention Network for Sequential Recommendation. In *SIGIR*. ACM, 78–88.

[11] Shijie Geng, Juntao Tan, Shuchang Liu, Zuohui Fu, and Yongfeng Zhang. 2023. VIP5: Towards Multimodal Foundation Models for Recommendation. In *Findings of EMNLP*. Association for Computational Linguistics, 9606–9620.

[12] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *ICLR*.

[13] Yupeng Hou, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Bolin Ding, and Ji-Rong Wen. 2022. Towards Universal Sequence Representation Learning for Recommender Systems. In *KDD*. 585–593.

[14] Hengchang Hu, Wei Guo, Yong Liu, and Min-Yen Kan. 2023. Adaptive Multi-Modalities Fusion in Sequential Recommendation Systems. In *CIKM*. ACM, 843–853.

[15] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *SIGKDD*, Won Kim, Ron Kohavi, Johannes Gehrke, and William DuMouchel (Eds.). ACM, 168–177.

[16] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendation. In *ICDM*. 197–206.

[17] Kibum Kim, Dongmin Hyun, Sukwon Yun, and Chanyoung Park. 2023. MELT: Mutual Enhancement of Long-Tailed User and Item for Sequential Recommendation. In *SIGIR*. ACM, 68–77.

[18] Chenliang Li, Xichuan Niu, Xiangyang Luo, Zhenzhong Chen, and Cong Quan. 2019. A Review-Driven Neural Model for Sequential Recommendation. In *IJCAI*. 2866–2872.

[19] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-based Recommendation. In *CIKM*. 1419–1428.

[20] Jiacheng Li, Ming Wang, Jin Li, Jinmiao Fu, Xin Shen, Jingbo Shang, and Julian J. McAuley. 2023. Text Is All You Need: Learning Language Representations for Sequential Recommendation. In *SIGKDD*. ACM, 1258–1267.

[21] Muyang Li, Zijian Zhang, Xiangyu Zhao, Wanyu Wang, Minghao Zhao, Runze Wu, and Ruocheng Guo. 2023. AutoMLP: Automated MLP for Sequential Recommendations. In *WWW*. ACM, 1190–1198.

[22] Muyang Li, Xiangyu Zhao, Chuan Lyu, Minghao Zhao, Runze Wu, and Ruocheng Guo. 2022. MLP4Rec: A Pure MLP Architecture for Sequential Recommendations. In *IJCAI*. 2138–2144.

[23] Xuewei Li, Aitong Sun, Mankun Zhao, Jian Yu, Kun Zhu, Di Jin, Mei Yu, and Ruiguo Yu. 2023. Multi-Intention Oriented Contrastive Learning for Sequential Recommendation. In *WSDM*. 411–419.

[24] Yujie Lin, Chenyang Wang, Zhumin Chen, Zhaochun Ren, Xin Xin, Qiang Yan, Maarten de Rijke, Xiuzhen Cheng, and Pengjie Ren. 2023. A Self-Correcting Sequential Recommender. In *WWW*. ACM, 1283–1293.

[25] Donghua Liu, Jing Li, Bo Du, Jun Chang, and Rong Gao. 2019. DAML: Dual Attention Mutual Learning between Ratings and Reviews for Item Recommendation. In *KDD*. ACM, 344–352.

[26] Haibo Liu, Zhixiang Deng, Liang Wang, Jinjia Peng, and Shi Feng. 2023. Distribution-based Learnable Filters with Side Information for Sequential Recommendation. In *RecSys*. ACM, 78–88.

[27] Sijia Liu, Jiahao Liu, Hansu Gu, Dongsheng Li, Tun Lu, Peng Zhang, and Ning Gu. 2023. AutoSeqRec: Autoencoder for Efficient Sequential Recommendation. In *CIKM*. ACM, 1493–1502.

[28] Zhenghao Liu, Sen Mei, Chenyan Xiong, Xiaohua Li, Shi Yu, Zhiyuan Liu, Yu Gu, and Ge Yu. 2023. Text Matching Improves Sequential Recommendation by Reducing Popularity Biases. In *CIKM*. ACM, 1534–1544.

[29] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2024. Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. *ACM Comput. Surv.* 56, 2 (2024), 30:1–30:40.

[30] Xiuyuan Qin, Huanhuan Yuan, Pengpeng Zhao, Junhua Fang, Fuzhen Zhuang, Guanfeng Liu, Yanchi Liu, and Victor S. Sheng. 2023. Meta-optimized Contrastive Learning for Sequential Recommendation. In *SIGIR*. ACM, 89–98.

[31] Yuyang Ren, Haonan Zhang, Qi Li, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. 2023. Self-supervised Graph Disentangled Networks for Review-based Recommendation. In *IJCAI*. ijcai.org, 2288–2295.

[32] Jie Shuai, Kun Zhang, Le Wu, Peijie Sun, Richang Hong, Meng Wang, and Yong Li. 2022. A Review-aware Graph Contrastive Learning Framework for Recommendation. In *SIGIR*. ACM, 1283–1293.

[33] Jiayu Song, Jiajie Xu, Rui Zhou, Lu Chen, Jianxin Li, and Chengfei Liu. 2021. CBML: A Cluster-based Meta-learning Model for Session-based Recommendation. In *CIKM*. 1713–1722.

[34] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *CIKM*. 1441–1450.

[35] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-Based Recommendation with Graph Neural Networks. In *AAAI*. 346–353.

[36] Heeyoon Yang, YunSeok Choi, Gahyung Kim, and Jee-Hyong Lee. 2023. LOAM: Improving Long-tail Session-based Recommendation via Niche Walk Augmentation and Tail Session Mixup. In *SIGIR*. 527–536.

[37] Yaowen Ye, Lianghao Xia, and Chao Huang. 2023. Graph Masked Autoencoder for Sequential Recommendation. In *SIGIR*. ACM, 321–330.

[38] Peiyan Zhang, Jiayan Guo, Chaozhuo Li, Yueqi Xie, Jaeboum Kim, Yan Zhang, Xing Xie, Haohan Wang, and Sunghun Kim. 2023. Efficiently Leveraging Multi-level User Intent for Session-based Recommendation via Atten-Mixer Network. In *WSDM*. 168–176.

[39] Xiaokun Zhang, Hongfei Lin, Bo Xu, Chenliang Li, Yuan Lin, Haifeng Liu, and Fenglong Ma. 2022. Dynamic intent-aware iterative denoising network for session-based recommendation. *Inf. Process. Manag.* 59, 3 (2022), 102936.

[40] Xiaokun Zhang, Bo Xu, Chenliang Li, Yao Zhou, Liangyue Li, and Hongfei Lin. 2024. Side Information-Driven Session-based Recommendation: A Survey. *arXiv preprint arXiv:2402.17129* (2024).

[41] Xiaokun Zhang, Bo Xu, Fenglong Ma, Chenliang Li, Yuan Lin, and Hongfei Lin. 2023. Bi-preference Learning Heterogeneous Hypergraph Networks for Session-based Recommendation. *ACM Trans. Inf. Syst.* 42, 3, Article 68 (2023), 28 pages.

[42] Xiaokun Zhang, Bo Xu, Fenglong Ma, Chenliang Li, Liang Yang, and Hongfei Lin. 2024. Beyond Co-Occurrence: Multi-Modal Session-Based Recommendation. *IEEE Trans. Knowl. Data Eng.* 36, 4 (2024), 1450–1462.

[43] Xiaokun Zhang, Bo Xu, Liang Yang, Chenliang Li, Fenglong Ma, Haifeng Liu, and Hongfei Lin. 2022. Price DOES Matter!: Modeling Price and Interest Preferences in Session-based Recommendation. In *SIGIR*. 1684–1693.

[44] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *SIGIR*. 83–92.

[45] Yipeng Zhang, Xin Wang, Hong Chen, and Wenwu Zhu. 2023. Adaptive Disentangled Transformer for Sequential Recommendation. In *KDD*. ACM, 3434–3445.

[46] Yichi Zhang, Guisheng Yin, and Yuxin Dong. 2023. Contrastive Learning with Frequency-Domain Interest Trends for Sequential Recommendation. In *RecSys*. ACM, 141–150.

[47] Peilin Zhou, Qichen Ye, Yueqi Xie, Jingqi Gao, Shoujin Wang, Jae Boum Kim, Chenyu You, and Sunghun Kim. 2023. Attention Calibration for Transformer-based Sequential Recommendation. In *CIKM*. ACM, 3595–3605.