Towards Hierarchical Intent Disentanglement for Bundle Recommendation

Ding Zou*, Sen Zhao*, Wei Wei† , Xian-ling Mao, Ruixuan Li, Dangyang Chen, Rui Fang, and Yuanyuan Fu

Abstract—Bundle recommendation aims to recommend a bundle of items for the user to purchase together, for which two scenarios (*i.e.*, Next-bundle recommendation and Within-bundle recommendation) are explored to recommend a specific bundle of items for the user and a specific item to fill the user's current bundle, respectively. Previous works largely model the user's preference with a uniform intent, without considering the diversity of intents when adopting the items within the bundle. In the real scenario of bundle recommendation, user intents modeling actually needs to be considered from three hierarchical levels, for that: a user's intents may be naturally distributed in different bundles (user level), one bundle may contain multiple intents of a user (bundle level), and an item in different bundles may also present different user intents (item level). To this end, we develop a novel model, Hierarchical Intent Disentangle Graph Networks (HIDGN) for bundle recommendation. HIDGN is capable of capturing the diversity of the user's intent precisely and comprehensively from the hierarchical structure with an cross-task intent contrastive learning, which is unified with the supervised next-/within-bundle recommendation sub-tasks as a multi-task framework. Extensive experiments on three benchmark datasets demonstrate that HIDGN outperforms the state-of-the-art methods by 43.0%, 13.2%, and 73.3%, respectively.

Index Terms—Bundle Recommendation, Disentangled Representation Learning, Contrastive Learning.

1 Introduction

Generally, traditional recommender systems [2], [3], [4], [5], [6], [7], [8], [9], [10] are mainly devoted to recommending individual items to users. Nevertheless, in a variety of real-world scenarios such as music or shopping platforms, providing bundles instead of items could satisfy user's interest more comprehensively. Consequently, the problem of bundle recommendation (BR) [11], [12], [13] has attracted increasing attention and emerged as an crucial recommendation scenario.

In terms of the way to construct bundles, existing works on bundle recommendation could be generally categorized into two sub-tasks: 1) Next-bundle recommendation, which recommends the user a pre-built **bundle** consisting of items to consume as a whole. 2) Within-bundle recommendation, which recommends a specific **item** to fill the user's current bundle. And they both have broad applications in real-world scenarios. For example, Apple Music can recommend pre-built music lists to the user; Amazon can recommend an individual item to put into the user's shopping cart.

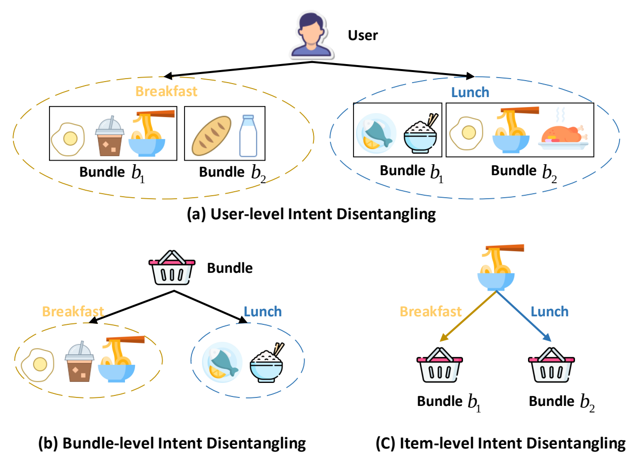

Fig. 1: An example of the user's multi-intent pattern for bundle recommendation.

Indeed, there are already lots of researches on these two distinct sub-tasks of bundle recommendation. For **next-bundle recommendation**, existing methods have made a lot of trials aiming at capturing the user's preference from both the bundle itself (*i.e.*, user-bundle interactions) and the content of the bundle (*i.e.*, bundle-item interactions). Earlier works [11], [14] simultaneously utilize both the user-bundle and user-item interactions to model user's preference for bundles, and some studies [15] model user-bundle and user-item interactions jointly under a multi-task framework. A more recent work [16] further unifies user-bundle-item affiliations into one heterogeneous graph and utilizes graph convolution networks (GCN) to learn the user and bundle's

*Ding Zou and Sen Zhao contributed equally to this work.
†Corresponding author: weiw@hust.edu.cn.

- *Ding Zou, Sen Zhao, Wei Wei are with Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430070, China. E-mail: {m202173662, senzhao, weiw}@hust.edu.cn*
- *Xian-ling Mao is with Beijing Institute of Technology, Beijing 100081, China. E-mail: maoxl@bit.edu.cn*
- *Ruixuan Li is with School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430070, China. E-mail: rxli@hust.edu.cn*
- *Dangyang Chen, Rui Fang, Yuanyuan Fu are with Ping An Property & Casualty Insurance company of China, Ltd, Shenzhen 518040, China. E-mail: {chendangyang273, fangrui051, fuyuanyuan723}@pingan.com.cn*

The preliminary conference version of this work has been published in AAAI-2022, titled: " Multi-view Intent Disentangle Graph Networks for Bundle Recommendation" [1].

representations, capturing user's preference from the item associations in bundle-item affiliations. For **Within-bundle recommendation**, existing methods make bundle-item prediction based on the user's preference modeled from the historically user-item interactions and bundle contents. Early models [17], [18], [19] jointly consider the user-item and item-item interactions within the bundle to model the user's preference and learn bundle/item representation. Follow-on studies [20], [21] apply GCN to explore the user-bundle-item graph, aiming to model the ignored heterogeneous high-order connectivity information.

Despite effectiveness, we argue that current methods ignore the diversity of the user's intents in next-/within-bundle recommendation. DGCF [22] stresses the importance of intent disentangling, which disentangles user-item interactions according to the user's intents for the item-based recommendation. Nevertheless, there exists a more complicated scenario in bundle recommendation, due to the special user-bundle-item structure which actually contains a hierarchical intent structure. To have a thorough analysis of user's intent distribution in bundle recommendation, we consider intents from the following three levels: 1) user-level, a user's intents may be naturally distributed in the different bundles; 2) bundle-level, a bundle may satisfy multiple intents of a user; 3) item-level, an item could be added into different bundles for satisfying various intents. A common fact is that all of the three levels are of vital importance for intents modeling, where the user level treats items as basic units for user intents and stresses that items in different bundles could present the same user intent, the bundle level reveals the item associations under each user intent, and the item level stresses that not only item associations but also one single item contain multiple user intents. However, a natural challenge occurs, for that the two sub-tasks have a distinct emphasis on the above three-level intent. Next-bundle recommendation models the possibility between user and bundle, hence focusing more on user- and item-level intents; while within-bundle recommendation models the bundle-item affiliation relation, hence emphasizes bundle- and item-level more. Consequently, it motivates us to design a hierarchical intent disentanglement framework for solving both next- and within-bundle intent modeling problems.

Actually, it's still non-trial to sufficiently disentangle user intents for the two distinct bundle recommendation sub-tasks, for that there are no labels for user intents. Inspired by the contrastive learning, one successful self-supervised learning paradigm, which could learn discriminative representations from unlabeled data, we propose to incorporate it into intent modeling for optimizing the intents in a self-supervised manner. Towards a better intent representation learning, previous methods [22] consider the independence between the intents, which is far from enough for effective intent representations. We hence contrast the learned intents of user-, bundle-, and item-level across next- and within-bundle tasks, for supervising intents learning of each level in a self-supervised manner.

In this paper, we proposed a novel model named HIDGN (*i.e.,* Hierarchical Intent Disentangle Graph Networks) to capture the diversity of the user's intents from the hierarchical structure with a cross-task intent contrastive learning. Specifically, we disentangle the representations of the user, bundle and item into chunks, where each represents a latent intent. For **next-bundle recommendation**, a graph neural network equipped with neighbor routing mechanism (graph disentangling module) is applied simultaneously to disentangle the user-item and the bundle-item graph, where user-level and bundle-level intents are disentangled correspondingly. For **within-bundle recommendation**, the graph disentangling module is applied to disentangle the user-level and item-level intents with the user's preference assisted.

The information from different levels presents user intents from different views, and the three intents learned from different levels are aligned with each other. To propagate information of intents across different levels, the intent propagating model is proposed.

Moreover, a cross-task intent contrastive learning is proposed to facilitate the user-/bundle-/item-level intent modeling in a self-supervised manner, which contrasts the disentangled intents of each level across the next- and within-bundle recommendation tasks. A multi-task framework is hence formed to unify the supervised next-/within-bundle recommendation and self-supervised intent contrastive learning. Extensive experiments conducted on NetEase, Youshu and Instacart demonstrate that our proposed model HIDGN outperforms the state-of-the-art methods, such as average relative increases of 43.0%, 17.0%, and 57.8% respectively in terms of Recall@20.

In summary, this work makes the following contributions:

- We emphasize the importance of disentangling the user's intent in two categories of bundle recommendation problems, and explore **hierarchical** intent disentanglement under the user-bundle-item heterogeneous structure.
- We propose a model named HIDGN, which builds a multi-task hierarchical intent modeling framework for next- and within-bundle recommendation. HIDGN develops a hierarchical intent disentanglement module for disentangling the user-/bundle-/item-level intents. HIDGN then performs intent contrastive learning across next-/within-bundle sub-tasks to optimize the intent representation learning in a self-supervised manner.
- We conduct extensive experiments on three benchmark datasets and achieve over 10 percent improvement of HIDGN over the state-of-art methods in both next-bundle recommendation and within-bundle recommendation.

2 Related Work

2.1 Bundle Recommendation

2.1.1 Next-Bundle Recommendation

Next-bundle recommendation(*a.k.a.,* bundle list recommendation [23], [24]) recommends an existing set of items to the user for consuming as a whole. Lire [11] simultaneously utilizes the user-bundle interactions and user-item interactions under the Bayesian Personalized Ranking(BPR) [25] framework, building a linear combination between the bundle and items. Cao *et al.* [14] treats bundle and item as sentence and word respectively, to capture the bundle-item co-occurrence information with a word embedding algorithm. He *et al.* [26] propose a self-attentive aggregation layer to capture the user-list-item structure. Chen *et al.* [15] develop a deep attentive model under the multi-task framework for jointly capturing the user's interactions with bundles and items. Later works focus on modeling the long-range connectivity in user-bundle-item interaction with Graph convolution networks (GCN). Chang *et al.* [16] apply GCN to learn user and bundle representations from the heterogeneous user-bundle-item graph. Li *et al.* [27] develop an attention network to capture information from the object level and bundle level, meanwhile propose fine- and coarse-grained aggregation network to learn the user's preference from the two-level information. However, previous works ignore the fact that the user's intents is diverse when adopting items, which making the learned user representation only express a coarse-grained preference of the user.

2.1.2 Within-Bundle Recommendation

The goal of within-bundle recommendation is to recommend a specific item to the user's current bundle. Early works [28], [29] model such bundle-item possibility simply according to historical user-item interactions. For example, [30], [31] model the travel package recommendation task as a *(linear) knapsack problem* and ignore the cross-item dependencies. [32] formalize the curriculum recommendation as a *constraint satisfaction problem* and utilize the item-to-item dependencies as hard-constraints for recommendation. Le *et al.* [17] develop a factorization-based method which considers multiple user-item associations. Latter work [13], [33] applies the skip-gram [34] framework to bundles to capture the no-linear interactions between bundles and items. Wan *et al.* [19] further explore the association of (item, item, user) triples that are linked by the same bundle and propose a model named triple2vec to learn representation from such associations. Recent studies start to capture the high-order connectivity information for within-bundle recommendation. BasConv [20] design three types of GCN aggregators for users, bundles, and items, respectively, which learn the heterogeneous high-order connectivity information in the user-bundle-item graph. Liu *et al.* [21] further explore the multiple representations of the bundle with a translation-based model and GNN. However, they are not able to disentangle the user's multiple intents within the bundle, which makes a suboptimal bundle representation learning.

2.2 Disentangled Representation Learning

Disentangled representation learning focuses on identifying and disentangling underlying factors hidden in the observed data [35]. In graph-structured areas, there are couples of works to utilize GCN for disentangled representation learning and explore disentangled GCN networks [36], [36], [37], [38]. Ma *et al.* [36] equip GCN with a neighbor routing mechanism and propose DisenGCN, which is capable of dynamically identifying the latent factor that may have caused the edge between a node and one of its neighbors. Liu *et al.* [37] improve DisenGCN by enhancing the independence of the different parts of the embedded feature. Yang *et al.* [38] further consider the multiple relations of the graph and develop a multi-relation disentanglement for the GCN. Recently, disentangled representation learning is applied to item-based recommendation. Wang *et al.* [22] considers high-order user-item relationships at the finer granularity of user's intents with the neighbor routing mechanism [36]. However, in bundle recommendation, little effort has been made towards disentangling user/bundle representation coupled with user's intents.

2.3 Contrastive Learning

Contrastive Learning [39], [40] optimizes node representations through contrasting positive pairs against negative pairs. DGI [40] utilizes Infomax [41] to contrast the node embeddings with graph embeddings in graph representation learning. GMI [42] then contrasts the node with its neighbors in aspects of feature and structure. Latter, MVGRL [43] contrasts the nodes of two structural graph views, where node representations are learned from node and graph levels. And in the item-based recommendation area, SGL [39] contrasts the original user-item graph with a corrupted one. SimGCL [44] throughly discusses the effectiveness of contrastive learning in improving representation uniformity for recommendation. However, there is little effort to bring the superiority of contrastive learning into bundle recommendation, especially for the intent modeling.

3 Preliminaries

Here we introduce the problem formulation of next- and within-bundle recommendations. Let $\mathcal{U} = \{u_1, u_2, \cdots, u_M\}$ be the user set, and $\mathcal{B} = \{b_1, b_2, \cdots, b_O\}$ and $\mathcal{I} = \{i_1, i_2, \cdots, i_N\}$ be the associated bundle set and item set, where M, O, N denote the number of users, bundles and items. According to the history of bundles user consumed, we can define the user-bundle interaction matrix, bundle-item affiliation matrix and user-item interaction matrix as $\mathbf{Y}_{M\times O} = \{y_{ub} | u \in \mathcal{U}, b \in \mathbf{B}\}$, $\mathbf{H}_{O\times N} = \{h_{bi} | b \in \mathcal{B}, i \in \mathbf{I}\}$ and $\mathbf{R}_{M\times N} = \{r_{ui} | u \in \mathcal{U}, i \in \mathcal{I}\}$, in which $y_{ub} = 1$, $h_{bi} = 1$ and $r_{ui} = 1$ means user u once interacted bundle b, bundle b contains item i, and user u once bought item i, respectively.

Basing on the above definition, the task of **next-bundle recommendation** is to predict the probability of user u potentially interacting with a given bundle that the user has never seen before. Specifically, our goal is to learn a predict function $\hat{\mathbf{y}}_{ub} = \mathcal{F}(u, b|\theta, \mathbf{H})$, where $\hat{\mathbf{y}}$ is the estimation probability and θ implies the parameters of function $\mathcal{F}$.

Unlike the next-bundle recommendation, within-bundle recommendation is to fill items i into the user's bundle b to maximize the user's preference over the bundle. Specifically the goal of **within-bundle recommendation** is to learn a predict function $\hat{\mathbf{y}}_{bi} = \mathcal{F}(b, i|u, \theta, \mathbf{H})$, where $\hat{\mathbf{y}}$ is the estimation probability and θ implies the parameters of function $\mathcal{F}$.

4 Hierarchical Intent Disentangle Graph Networks

We present the proposed Hierarchical Intent Disentangle Graph Networks in detail (shown in Figure 2), which comprises three main components: 1) Hierarchical graph disentangling module, which first proposes a graph disentangling

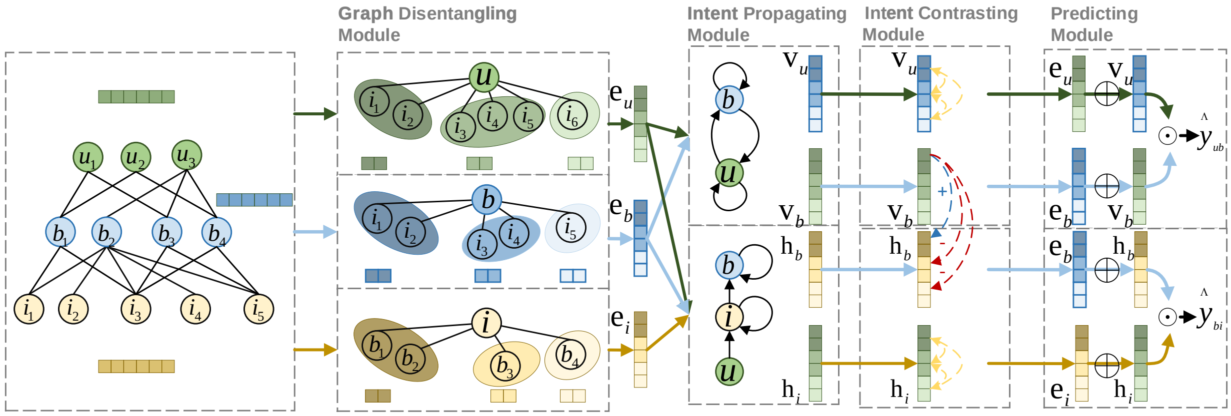

Fig. 2: The framework of the proposed HIDGN, which contains hierarchical graph disentangling module, intent propagating module, cross-task intent contrasting module and predicting module. (Best view in color)

module and then utilizes it to disentangle the interaction graph and learn the representations coupled with the user's intents from *user-level*, *bundle-level* and *item-level*, respectively. 2) Intent propagating module, which propagates the user-/bundle-/item-level intent disentangled information through user-bundle-item graph to acquire sufficient representations for next- and within-bundle recommendation. 3) Cross-task intent contrastive learning, which contrasts the learned intents of different levels across next- and within-bundle recommendation sub-tasks, to improve the intent representation learning in a self-supervised manner.

4.1 Hierarchical Graph Disentangling Module

4.1.1 Graph Disentangling Module

The interactions between users, bundles and items are coupled with diverse intents of users to different degrees. The graph disentangling module is to disentangle the user's intents coupled with the interaction graph and refine the representation of user/bundle/item under different intents.

Initialization of embeddings and graphs. To obtain representations under different intents, we divide the user's representations as K chunks:

$$\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_K), \tag{1}$$

where K is the pre-defined number of the intents and each chunk $\mathbf{u}_k \in R^{\frac{d}{K}}$ indicates the user representation the k-th intent. Analogously, the representation of the bundle $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_K)$ and item $\mathbf{i} = (\mathbf{i}_1, \mathbf{i}_2, \cdots, \mathbf{i}_K)$ are accordingly build. Each chunk of the representation is independently initialized.

Each interaction in the interaction graph $\mathcal{G}$ associated to K intents to different degrees. To model the interactions under different intents, we build a set of intent-aware graphs $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_K\}$, each of which is associated with a weighted adjacent matrix $\mathbf{A}_k$. Each entry $\mathbf{A}_k(a, c)$ of the adjacent matrix indicates the degree that the interaction between nodes a and c associates to the k-th intent, where a and c[1] are placeholders for the user, bundle and item. Furthermore, for each interaction within the graph, we construct a score vector $\mathbf{A}(a, c) = (\mathbf{A}_1(a, c), \mathbf{A}_2(a, c), \cdots, \mathbf{A}_K(a, c))$ that indicates the degrees that the interaction is associated with the user's K intents. The score vector is uniformly initialized as:

$$\mathbf{A}(a, c) = (1, 1, \cdots, 1). \tag{2}$$

Disentangling of the intent-aware graphs. Each chunk of the representation $\mathbf{a_k}$ is specialized with an interaction graph $\mathcal{G}_k$ under the same intent. We should disentangle the interaction graph coupled with the intent and learn the representation with the intent-aware graph as follows:

$$\mathbf{e}_{ak}^{(1)} = g(\mathbf{a}_k, \mathbf{c}_k, \mathbf{A}_k), \tag{3}$$

where $g(\cdot)$ indicates the graph disentangling layer and $\mathbf{A}_k$ denotes the adjacent matrix of the interaction graph $\mathcal{G}_k$. $\mathbf{e}_{ak}^{(1)}$ is expected to encode the collaborative signal associated with the k-th intent of the user. The super-index (1) indicates the first-order graph disentangling layer.

As shown in Figure 2, we employ the neighbour routing mechanism to iteratively update the adjacent matrix $\mathbf{A}_k$ and chunk of the representation $\mathbf{a}_k$ under each intent. $\mathbf{a}_k^t$ and $\mathbf{A}_k^t$ are utilized to memorize the intermediate values of $\mathbf{a}_k$ and $\mathbf{A}_k$.

Update within each iteration. With the score vector $\{\mathbf{A}_k^t(a, c) | \forall k \in \{1, 2, \cdots, K\}\}$ for each interaction (a, c), we calculate its distribution over intents with the softmax function as:

$$\tilde{\mathbf{A}}_k^t(a, c) = \frac{exp \mathbf{A}_k^t(a, c)}{\sum_{k'=1}^{K} exp \mathbf{A}_{k'}^t(a, c)}, \tag{4}$$

which denotes the confidence of the interaction (a, c) under each intent.

Each chunk of the representation $\mathbf{a}_k$ is updated by weighted aggregation over the intent-aware graph $\mathcal{G}_k$ under the same intent:

$$\mathbf{a}_k^t = \sum_{c \in \mathcal{N}_a} \frac{\tilde{\mathbf{A}}_k^t(a, c)}{\sqrt{D_t^k(a) \cdot D_t^k(c)}} \cdot \mathbf{c}_k, \tag{5}$$

where $D_t^k(a) = \sum_{c \in \mathcal{N}_a} \tilde{\mathbf{A}}_k^t(a, c)$ and $D_t^k(c) = \sum_{a \in \mathcal{N}_c} \tilde{\mathbf{A}}_k^t(a, c)$ are degrees of node a and c respectively.

1. a and c are utilized as a uniform place holders for the user u, bundle b and item i in this paper

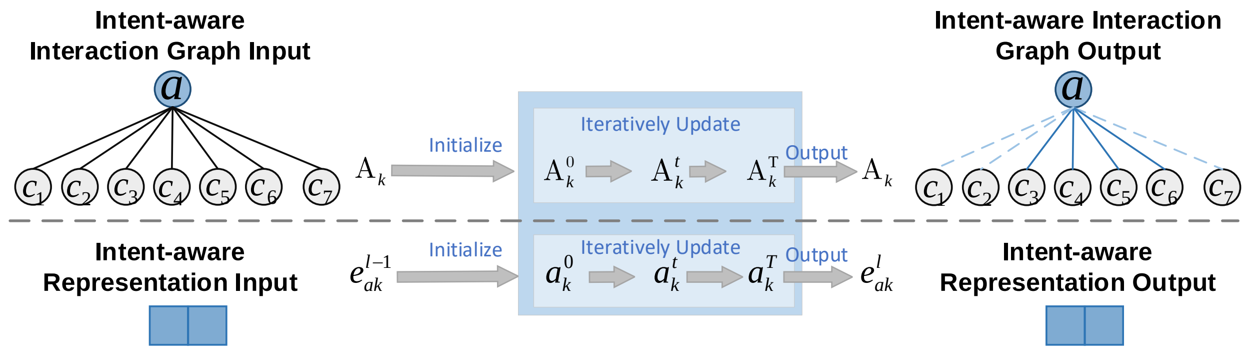

Fig. 3: The graph disentangling module. (Best view in color)

Then the confidence $\mathbf{A}_k^t(a,c)$ of interaction (a,c) under each intent is updated through:

$$\mathbf{A}_k^{t+1}(a,c) = \mathbf{A}_k^t(a,c) + {\mathbf{a}_k^t}^{\mathrm{T}} \cdot \mathbf{c}_k, \tag{6}$$

where the the connection $\mathbf{A}_k(a,c)$ between nodes with similar embedding will be enhanced.

Layer combination. $\mathbf{e}_{ak}^1$ involve first-order collaborative signal within the intent-aware graph $\mathcal{G}_k$. To incorporate higher-order information, more graph disentangling layers are employed and combined as:

$$\mathbf{e}_{ak}^l = g(\mathbf{e}_{ak}^{l-1}, \mathbf{c}_k, \mathbf{A}_k). \tag{7}$$

$$\mathbf{e}_{ak} = \sum_l \mathbf{e}_{ak}^l. \tag{8}$$

To sum up, the presentation of the node a is disentangled into chunks $\mathbf{e}_a = (\mathbf{e}_{a1}, \mathbf{e}_{a2}, \cdots, \mathbf{e}_{aK})$ coupled with the user's intents by the graph disentangling module as:

$$\mathbf{e}_a = g^*(\mathbf{a}, \mathbf{c}, \mathbf{A}), \tag{9}$$

where $g^*(\cdot)$ demonstrates the graph disentangling module formulated from eq. (1) to eq. (8).

4.1.2 Hierarchical Graph Disentangling

The hierarchical graph disentangling module learns the user's intents from different levels under the hierarchical framework (user, bundle, and item levels). From the user level, the hierarchical graph disentangling module disentangles the user-item interaction graph and learns the user's intents distributed in different bundles. From the bundle level, the hierarchical graph disentangling module learns the diverse intents of the user within each bundle. And from the item level, the hierarchical graph disentangling module learns different intents that each item may satisfy when been added into different bundles. Specifically, the module divides the representation of users, bundles and items into chunks coupling with diverse intents of the user, and learns the chunks of representation from disentangled interaction graphs using the graph disentangling module eq. (9) as:

$$\begin{aligned} \mathbf{e}_u &= g^*(\mathbf{u}, \mathbf{i}, \mathbf{A}), \\ \mathbf{e}_b &= g^*(\mathbf{b}, \mathbf{i}, \mathbf{A}), \\ \mathbf{e}_i &= g^*(\mathbf{i}, \mathbf{b}, \mathbf{A}), \end{aligned} \tag{10}$$

where $\mathbf{e}_u = (\mathbf{e}_{u1}, \mathbf{e}_{u2}, \cdots, \mathbf{e}_{uK})$, $\mathbf{e}_b = (\mathbf{e}_{b1}, \mathbf{e}_{b2}, \cdots, \mathbf{e}_{bK})$ and $\mathbf{e}_i = (\mathbf{e}_{i1}, \mathbf{e}_{i2}, \cdots, \mathbf{e}_{iK})$ are couple with the user's intents learned from the user, bundle and item levels, respectively.

4.2 Intent Propagating Module

The representations of the user $\mathbf{e}_u$, the bundle $\mathbf{e}_b$ and the item $\mathbf{e}_i$ learned from the hierarchical graph disentangling module are coupled with intents learned from different levels. To propagate information of intents cross different levels, intent propagating model is proposed.

4.2.1 Intent Propagating Module for Next-Bundle Recommendation

Intent propagating module for the next-bundle recommendation exchanges information of the user's intents from the user level and the bundle level with the graph aggregation mechanism:

$$\begin{aligned} \mathbf{v}_u &= \sum_{b \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u|}\sqrt{|\mathcal{N}_b|}} \mathbf{e}_b + \mathbf{e}_u, \\ \mathbf{v}_b &= \sum_{u \in \mathcal{N}_b} \frac{1}{\sqrt{|\mathcal{N}_b|}\sqrt{|\mathcal{N}_u|}} \mathbf{e}_u + \mathbf{e}_b. \end{aligned} \tag{11}$$

The representation $\mathbf{v}_u$ and $\mathbf{v}_b$ incorporate information of intents from the bundle level and the user level, respectively.

4.2.2 Intent Propagating Module for Within-Bundle Recommendation

The intent propagating module propagates the user level information to both the representation of the item and the bundle, which can help better learn the user's preference for within-bundle recommendation. Since the user first chooses the item and then adds it into the bundle in the within-bundle recommendation, the information from the user level is propagated through the item to the bundle:

$$\begin{aligned} \mathbf{h}_i &= \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i|}\sqrt{|\mathcal{N}_u|}} \mathbf{e}_u + \mathbf{e}_i, \\ \mathbf{h}_b &= \sum_{i \in \mathcal{N}_b} \frac{1}{\sqrt{|\mathcal{N}_b|}\sqrt{|\mathcal{N}_u|}} \mathbf{h}_i + \mathbf{e}_b. \end{aligned} \tag{12}$$

where $\mathbf{h}_b = (\mathbf{h}_{b1}, \mathbf{h}_{b2}, \cdots, \mathbf{h}_{bK})$ and $\mathbf{h}_i = (\mathbf{h}_{i1}, \mathbf{h}_{i2}, \cdots, \mathbf{h}_{iK})$ are representations of the bundle and item which are coupled with the user's intents. Learned from the users within their neighborhood, each chunk of these representations can present users' preference under each intent.

4.3 Cross-task Intent Contrastive Learning

After disentangling latent intents from user/bundle/item level in next-/within-bundle recommendation sub-task, a natural challenge occurs, which is the fact of no labels for intents. Previous methods usually optimize the intents by making certain constraints, such as independence between intents [22]. However, it's far from enough for hierarchical intents modeling, for that three-level intents are disentangled independently, which requires the intents of different levels to align with each other. Recently, contrastive learning has become a mainstream solution to train without explicit labels, due to its effective performance in mining labels from data itself, by minimizing the distance between positive samples meanwhile maximizing the distance between negative samples. As a result, a contrastive learning based intent optimizing module is proposed for a more effective intent representation learning. It stresses the independence among intents and aligning intents of hierarchical levels across two sub-tasks, by contrasting the disentangled bundle representations $\mathbf{h}_b$ and $\mathbf{v}_b$ learned from supervised next- and within-bundle sub-tasks.

Firstly, a space mapping is performed according previous contrastive learning methods, with a MLP projection layer:

$$\begin{aligned} \mathbf{h}_b^p &= W^{(2)}\sigma\left(W^{(1)}\mathbf{h}_b + b^{(1)}\right) + b^{(2)}, \\ \mathbf{v}_b^p &= W^{(2)}\sigma\left(W^{(1)}\mathbf{v}_b + b^{(1)}\right) + b^{(2)}, \end{aligned} \tag{13}$$

where σ is the sigmoid function, $W^{(\cdot)} \in \mathbb{R}^{d\times d}$ and $b^{(\cdot)} \in \mathbb{R}^{d\times 1}$ are trainable parameters,. Here the positive and negative pairs are defined as follows: for an intent in a certain sub-task (next- or within-bundle recommendation sub-tasks), the same intent's representations learned from two sub-tasks form the positive pairs, while other intent representations in two sub-tasks are naturally regarded as negative pairs, which hence include both intra-task and inter-task negative pairs.

Then the contrastive loss for cross-task intent contrasting is naturally obtained as follows:

$$\begin{aligned} &\mathcal{L}^{SSL} = \\ &-\sum_{b=1}^{O} \log \frac{e^{\mathrm{s}\left(\mathbf{v}_b^p, \mathbf{h}_b{}^p\right)/\tau}}{\underbrace{e^{\mathrm{s}\left(\mathbf{v}_b^p, \mathbf{h}_b^p\right)/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k\neq b} e^{\mathrm{s}\left(\mathbf{v}_b^p, \mathbf{v}_k^p\right)/\tau}}_{\text{intra-task negative pairs}} + \underbrace{\sum_{k\neq b} e^{\mathrm{s}\left(\mathbf{v}_b^p, \mathbf{h}_k^p\right)/\tau}}_{\text{inter-task negative pairs}}}, \end{aligned} \tag{14}$$

where $\mathrm{s}(\cdot)$ is the cosine similarity function, and τ is the parameter of temperature. In this way, a self-supervised learning task is introduced into intents modeling and optimization.

4.4 Predicting Module

After obtaining the presentation of the user, bundle and item from hierarchical intent modeling and propagating, we get the final representations by concatenating, and we estimate the likelihood of their interactions $\hat{\mathbf{y}}_{ub}$ and $\hat{\mathbf{y}}_{bi}$ through inner product for next- and within-bundle recommendation, as follows:

$$\begin{aligned} \hat{\mathbf{y}}_{ub} &= (\mathbf{e}_u \oplus \mathbf{v}_u) \odot (\mathbf{e}_b \oplus \mathbf{v}_b), \\ \hat{\mathbf{y}}_{bi} &= (\mathbf{e}_b \oplus \mathbf{h}_b) \odot (\mathbf{e}_i \oplus \mathbf{h}_i), \end{aligned} \tag{15}$$

where $\odot$ denotes dot product and $\oplus$ denotes the concatenation.

Towards combining the self-supervised task with the two supervised bundle recommendation tasks, a unified multi-task training strategy is hence proposed to optimize the whole model. For the supervised next- and within-bundle recommendation, the Bayesian Personalized Ranking loss [45] is employed as:

$$\begin{aligned} \mathcal{L}_{pred}^n &= \sum_{(u,b^+,b^-)\in Q} -\ln\sigma(\hat{\mathbf{y}}_{ub^+} - \hat{\mathbf{y}}_{ub^-}) + \lambda_1 \cdot ||\theta||^2, \\ \mathcal{L}_{pred}^w &= \sum_{(b,i^+,i^-)\in O} -\ln\sigma(\hat{\mathbf{y}}_{bi^+} - \hat{\mathbf{y}}_{ui^-}) + \lambda_1 \cdot ||\theta||^2, \end{aligned} \tag{16}$$

where $Q = \{(u,b^+,b^-)|(u,b^+) \in y^+, (u,b^-) \in y^-\}$, $O = \{(b,i^+,i^-)|(b,i^+) \in y^+, (b,i^-) \in y^-\}$ are training data for next- and within-bundle recommendation, respectively. y^+ denotes the observed interactions and y^- indicates the sampled negative interactions. L_2 regularization controlled by the coefficient λ_1 is applied on the model's parameter θ to avoid over-fitting. We combine the intent contrastive loss with BPR loss and optimize the model by minimizing the following objective function:

$$\mathcal{L}_{HIDGN} = \mathcal{L}_{pred}^n + \mathcal{L}_{pred}^w + \mathcal{L}^{SSL} \tag{17}$$

5 Experiments

Extensive experiments are conducted on three public datasets to answer the following questions.

- **RQ1:** Compared to previous approaches, how does HIDGN perform in next-bundle and within-bundle recommendation scenarios?
- **RQ2:** How do different components (*e.g.*, Graph Disentangling Module in the user/bundle/item level, Cross-task Intent Contrastive Learning module) affect the performance of HIDGN?
- **RQ3:** How do parameters (layer number of Graph Disentangling module, intent number) influence the results of HIDGN?
- **RQ4:** Can HIDGN provide in-depth analyses of the disentangled representations?

5.1 Datasets and metrics

As shown in Table 1, we use three datasets to evaluate our proposed method.

- **NetEase:** This dataset is provided by the work [14] and constructed with the logs from Netease Cloud Music. It contains users' clicking histories of the user-generated playlists and individual songs.
- **Youshu:** It is provided by the work [15] and constructed with the logs from a book review site, Youshu. It contains the users' interactions with individual books and bundles of books.
- **Instacart:** It is published by instacart and is constructed with the user's grocery transaction records from online grocery shopping.

It's worth noting that NetEase and Youshu are used for Next-bundle Recommendation, and Instacart is for Within-bundle Recommendation.

Dataset	NetEase	Youshu	Instacart
User	18,528	8,039	1,628
Bundle	22,864	4,771	3,081
Item	123,628	32,770	20,387
User-Bundle	302,303	51,337	3,081
Bundle-item	1,778,838	176,667	180,409
User-item	1,128,065	138,515	180,409
User-bundle density	0.06%	0.13%	0.06%
Bundle-item density	0.06%	0.11%	0.29%
User-item density	0.05%	0.05%	0.54%

TABLE 1: Statistics of three utilized datasets

Next-bundle Recommendation Sub-task For the next-bundle recommendation sub-task, a method will produce a top-k bundle recommendation list R for each use. Following previous bundle recommendation methods [16], [24], we employ Recall@K and NDCG@K as evaluation metrics. Recall@K means the ratio of test bundles/items recommended within the top-K list. NDCG@K assigns higher scores to the bundles/items at a higher position on R.

In the formula below, K denotes the implementation of a top-K ranking. Subsequently, a bundle list B of length K is systematically generated for each user. b_k indicates that the bundle situated at the k-th position within the generated bundle list B, and d represents the bundle from bundle set D with which the user has indeed interacted.

$$\text{Recall } @K = \frac{|D \cap B|}{|D|} \tag{18}$$

$$DCG@K = \sum_{k=1}^{K} \frac{2^{\text{hit}_k} - 1}{\log_2(k+1)} \tag{19}$$

where hit_k = 1 when $b_k \in D$, otherwise hit_k = 0.

Within-bundle Recommendation Sub-task For the within-bundle recommendation sub-task, we also select Recall@K and NDCG@K as evaluation metrics to judge the recommendation list performance. We first compute the matching scores of all the potential items for the basket b and then sort items based on the matching scores. Finally, we evaluate the ranking results by comparing them with the true items within basket b.

5.2 Baselines

We choose state-of-art methods to compare with our proposed HIDGN. Specifically, two matrix factorization based methods, and several graph based methods are chosen for Next-bundle Recommendation.

- **MFBPR** [46]: This work proposes a matrix factorization method under the Bayesian Personalized Ranking learning framework.
- **DAM** [15]: This work applies the factorized attention mechanism to capture collaborate signals and bundle-level associations under the multi-task learning framework.
- **NGCF** [47]: This work develops a method with GCN to capture high-order connectivity information for prediction.
- **NGCF-UB** [47]: The NGCF method is applied to the user-bundle graph.
- **RGCN** [48]: RGCN is a method based on GCN which is developed for the multi-relational graph.
- **BundleNet [49]**: it builds a user-bundle-item tripartite graph, leverages GCN and applies multi-task learning to learn the representations.
- **BGCN** [16]: BGCN is a graph based model proposed to explicitly model the complex relations within the user-bundle-item graph.
- **BRUCE [50]**: it introduces transformer into bundle recommendation, capturing the context of each item in the bundle.

For within-bundle recommendation, we also apply MF-BPR, NGCF, RGCN to the within-bundle recommendation. Except for these three baselines, we also compare with the following methods:

- **NGCF-BI** [47]: The NGCF is applied on the bundle-item graph to predict the interactions between bundles and items.
- **BasConv** [20]: This work devises heterogeneous aggregators to learn the embedding of each kind of nodes within the graph between users, bundles and items.
- **MITGNN** [16]: MITGNN combines the graph neural network with a translation-based model to consider the multiple intents within a basket.

Parameter Settings For HIDGN, the embedding size is set to 64, the number of layers and the number of intents are selected from {1, 2, 3, 4} and {1, 2, 4, 8}, respectively. In the training procedure, the negative sampling rate is set to 1, and the learning rate is selected from {1e-5, 3e-5, 1e-4, 3e-4, 1e-3, 3e-3}. The batch size is 4096 and BPR loss is adopted with the Adam optimizer. Our experiments are conducted on a Nvidia RTX 3090 graphics card equipped with AMD r9-5900x CPU.

5.3 Performance Comparison(RQ1)

From the experiment results of next-bundle recommendation (reported in Table 2), we have following observations.

- **Our proposed HIDGN achieves the best results.** Specifically, HIDGN improves the performance over the best baseline by 43.0% and 37.2% in terms of Recall@20 and NDCG@20 on NetEase. And HIDGN outperforms the best result by 13.2% and 11.1% in terms of Recall@20 and NDCG@20 on Youshu. We attribute the improvements to the following aspects: 1) By disentangling the embedding of user/bundle/item according to the user's intents in a hierarchical structure, HIDGN learns user/bundle representation at a more granular level. 2) By contrasting the intents of different sub-tasks, the unified multi-task framework learns more discriminative bundle embeddings coupled with intents.
- **Graph models achieve better performance in next-bundle recommendation.** Graph models (NGCF, RGCN) achieve better performance than MFBPR, which can be attributed to their superiority in capturing graph structure and high-order connectivity information.

Method	NetEase Metrics@20		NetEase Metrics@40		NetEase Metrics@80		Youshu Metrics@20		Youshu Metrics@40		Youshu Metrics@80	
	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG
MFBPR	0.0335	0.0181	0.0600	0.0246	0.0948	0.0323	0.1959	0.1117	0.2735	0.1320	0.3710	0.1543
DAM	0.0411	0.0210	0.0690	0.0281	0.1090	0.0372	0.2082	0.1198	0.2890	0.1418	0.3915	0.1658
NGCF	0.0384	0.0198	0.0636	0.0266	0.1015	0.0350	0.2119	0.1165	0.2761	0.1343	0.3743	0.1561
NGCF-UB	0.0395	0.0207	0.0646	0.0274	0.1021	0.0359	0.1985	0.1143	0.2658	0.1324	0.3542	0.1524
RGCN	0.0407	0.0210	0.0670	0.0280	0.1112	0.0378	0.2040	0.1069	0.3017	0.1330	0.4169	0.1595
BundleNet	0.0391	0.0201	0.0661	0.0271	0.1141	0.0369	0.1895	0.1125	0.2675	0.1335	0.3988	0.1548
BGCN	<u>0.0491</u>	<u>0.0258</u>	<u>0.0829</u>	<u>0.0346</u>	<u>0.1304</u>	<u>0.0453</u>	0.2347	0.1345	0.3248	0.1593	0.4355	0.1851
BRUCE	0.0489	0.0251	0.0821	0.339	0.1287	0.0425	<u>0.2425</u>	<u>0.1428</u>	<u>0.3412</u>	<u>0.1696</u>	<u>0.4621</u>	<u>0.1974</u>
MIDGN	0.0678	0.0343	0.1085	0.0451	0.1654	0.0578	0.2682	0.1527	0.3712	0.1808	0.4817	0.2063
HIDGN	**0.0702***	**0.0353***	**0.1117***	**0.0464***	**0.1699***	**0.0594***	**0.2746***	**0.1586***	**0.3836***	**0.1880***	**0.4972***	**0.2144***
%Improv.	43.0%	37.2%	34.8%	34.2%	30.3%	31.3%	13.2%	11.1%	12.4%	10.8%	7.6%	8.6%

TABLE 2: Performance comparisons on two real-world datasets. * denotes the statistical significance for $p < 0.001$ based on a two-tailed paired t-test.

- **It is important to capture the user's preference on both user and bundle levels.** Neglecting the bundle-level associations, graph models (NGCF, RGCN) is surpassed by DAM, a method unifying user- and bundle-level information under the multi-task framework. Besides, BGCN significantly improve the performance by learning the bundle-level associations with GCN and performs the best among baselines.

From experimental results of within-bundle recommendation reported in Table 3, we have following observations.

- **Our proposed HIDGN achieves the best results.** On Instacart, HIDGN achieves improvement over the best baseline by 57.8% and 50.3% in terms of Recall@20 and NDCG@20. We attribute the improvements to the following aspects: 1) By disentangling hierarchical intents from three levels, HIDGN makes bundle-item prediction from a more granular level. 2) Through intent contrasting with next-bundle recommendation sub-task, intent modeling of the two sub-tasks gets improvement.
- **Graph models perform better for within-bundle recommendation.** Similar to the next-bundle recommendation, graph models (NGCF, RGCN, BasConv) achieve better performance, which is mainly attributed to the great superiority of GNN in capturing graph structure and high-order neighboring information. This fact motivates us to design a proper GNN mechanism for within-bundle recommendation.
- **Explicitly considering multiple user intents is effective in within-bundle recommendation.** Ignoring the effectiveness of user's intents, most of the graph models fail to surpass MITGNN, which simply transfers bundle representation into multi embeddings as intents. But MITGNN performs worse than BasConv in a few scenarios, which motivates us to disentangle user's intents more precisely.

Method	Instacart Metrics@20		Instacart Metrics@40		Instacart Metrics@80	
	Recall	NDCG	Recall	NDCG	Recall	NDCG
MFBPR	0.0289	0.0253	0.0506	0.0353	0.0849	0.0487
NGCF	0.0329	0.0283	0.0571	0.0395	0.0963	0.0549
NGCF-BI	0.0471	0.0432	0.0727	0.0552	0.1072	0.0687
RGCN	0.0250	0.0226	0.0415	0.0303	0.0704	0.0416
BasConv	<u>0.0491</u>	<u>0.0463</u>	0.0872	0.0592	0.1502	0.0887
MITGNN	0.0465	0.0423	<u>0.0892</u>	<u>0.0615</u>	<u>0.1675</u>	<u>0.0975</u>
HIDGN	**0.0851***	**0.0744***	**0.1360***	**0.0982***	**0.2132***	**0.1284***
%Improv.	73.3%	60.7%	52.5%	59.7%	27.3%	31.7%

TABLE 3: Performance comparisons on Instacart for Within-bundle recommendation. * denotes the statistical significance for $p < 0.001$ based on a two-tailed paired t-test.

Method	NetEase Metrics@20		Youshu Metrics@20	
	Recall	NDCG	Recall	NDCG
HIDGN$_{w/o\,user}$	0.0547	0.0287	0.2475	0.1393
HDGN$_{w/o\,bundle}$	0.0583	0.0303	0.2464	0.1441
HIDGN$_{w/o\,item}$	0.0609	0.0312	0.2588	0.1471
HIDGN$_{w/o\,contra}$	0.0683	0.0345	0.2706	0.1548
HIDGN	0.0702	0.0353	0.2746	0.1586

TABLE 4: Ablated models analysis for next-bundle recommendation.

5.4 Study of HIDGN (RQ2&RQ3)

5.4.1 Ablation Study (RQ2)

We investigate the underlining mechanism of our HIDGN for both next-bundle recommendation and within-bundle recommendation, with the following three ablated models:

- **HIDGN$_{w/o\,user}$** : We replace the graph disentangling module in the user level with GCN for both next- and within-bundle recommendation.
- **HIDGN$_{w/o\,bundle}$** : We remove the graph disentangling module in the bundle level and uses GCN instead for both next- and within-bundle recommendation.
- **HIDGN$_{w/o\,item}$** : We remove the graph disentangling module in the item level and uses GCN instead for both next- and within-bundle recommendation.
- **HIDGN$_{w/o\,contra}$** : We remove cross-task intent contrastive learning for both next- and within-bundle recommendation.

From results in Table 4 and Table 5, we have the following observations for next- and within-bundle recommendation:

Method	Instacart			
	Metrics@20		Metrics@40	
	Recall	NDCG	Recall	NDCG
HIDGN$_{w/o\,user}$	0.0742	0.0672	0.1148	0.0864
HIDGN$_{w/o\,bundle}$	0.0688	0.0627	0.1112	0.0825
HIDGN$_{w/o\,item}$	0.0603	0.0546	0.0969	0.0717
HIDGN$_{w/o\,contra}$	0.0724	0.0653	0.1179	0.0866
HIDGN	0.0851	0.0744	0.1360	0.0982

TABLE 5: Ablated models analysis for within-bundle recommendation.

- HIDGN outperforms the model that ablates intent disentanglement from any level. This demonstrates the importance of the hierarchical structure in modeling the user's multiple intents, where the user level stresses that items in different bundles could present the same user intent, the bundle level reveals the item associations under each user intent, and the item level stresses that not only item associations but also one single item contain multiple user intents.
- HIDGN$_{w/o\,user}$ performs the worst for next-bundle recommendation, while HIDGN$_{w/o\,item}$ performs worst for within-bundle recommendation. This demonstrates that intent disentanglement from various levels have different importance for next- and within- bundle recommendation. Next-bundle recommendation focuses more on user-level and item-level intents; while within-bundle recommendation emphasises bundle-level and item-level intents.
- HIDGN outperforms HIDGN$_{w/o\,contra.}$ in both next- and within-bundle recommendation. This fact demonstrates the effectiveness of the cross-task intent contrasting module in help optimize the intent modeling in a self-supervised manner.
- All the ablated models significantly outperform the baselines in all of the situations. This demonstrates the effectiveness of hierarchical intent disentangle graph networks for bundle recommendation.

5.4.2 Parameter Study (RQ3)

We further conduct parameter studies to investigate how the parameters (*i.e.,* the layer number of hierarchical graph disentangling module, the intent number of the user) influence the HIDGN's performance in next- and within-bundle recommendation sub-tasks.

Impact of Layer Number. The hierarchical graph disentangling module disentangles user intents coupled with the interaction graph and refine the representation of user/bundle/item under different intents. High-order connectivity information is distilled with more layers. To investigate how the layer number influences HIDGN's performance, we conduct experiments with layer L varying in range of $\{1, 2, 3, 4\}$ for next- and bundle recommendation, where HIDGN-L is to indicate L layers in the hierarchical graph disentangling module. From the results shown in Table 6 and Table 7, We have some observations:

- In both next- and within-bundle recommendation sub-tasks, increasing the layer number boosts the performance of HIDGN. HIDGN-1 performs the worst, which is mainly because HIDGN-1 only gains information from one-hop neighbors and neglects high-order collaborative information.
- The performance does not always improve with the improvement of the layer number. HIDGN-3 outperforms HIDGN-4 on Youshu. We attribute this to the noise increasing along with the hop of neighbors.

Method-L	NetEase		Youshu	
	Metrics@20		Metrics@20	
	Recall	NDCG	Recall	NDCG
HIDGN-1	0.0442	0.0233	0.2348	0.1348
HIDGN-2	0.0559	0.0295	0.2503	0.1429
HIDGN-3	0.0642	0.0331	0.2746	0.1586
HIDGN-4	0.0702	0.0353	0.2617	0.1472

TABLE 6: Impact of Layer Number(L) for next-bundle recommendation.

Method-L	Instacart			
	Metrics@20		Metrics@40	
	Recall	NDCG	Recall	NDCG
HIDGN-1	0.0657	0.0599	0.1059	0.0787
HIDGN-2	0.0703	0.0634	0.1141	0.0839
HIDGN-3	0.0752	0.0679	0.1211	0.0894
HIDGN-4	0.0851	0.0744	0.1360	0.0982

TABLE 7: Impact of Layer number(L) for within-bundle recommendation.

Impact of Intent Number. To investigate how the intent number influence the HIDGN's performance, we conduct experiments with the Intent number K in range $\{1, 2, 4, 8\}$. From the results shown in Fig. 5 and Fig. 6, we have the following observations:

- In both next- and within-bundle recommendation, increasing the intent number properly promotes the model performance, which presumes the importance of user's multiple intents for next-bundle recommendation and within-bundle recommendation.
- In next-bundle recommendation, HIDGN with the intent number $K = 1$ performs the worst, which illustrates the fact that the user's intents are diverse and can not be presented by a unitary presentation.
- With the increasing of intent number K from 4 to 8, the performance of HIDGN drops sharply in two recommendation sub-tasks. And especially in within-bundle recommendation, HIDGN gets the worst performance when $K = 8$. This fact suggests the model actually suffers from too fine-grained intents.

5.5 Visualization (RQ4)

Visualization of Item Embedding. To further verify whether HIDGN has disentangled the user's intents under hierarchical structure, we perform the t-SNE visualization, as shown in Figure 7. In Figure 7, three users in Instacart are randomly chosen and the representations of their interacted

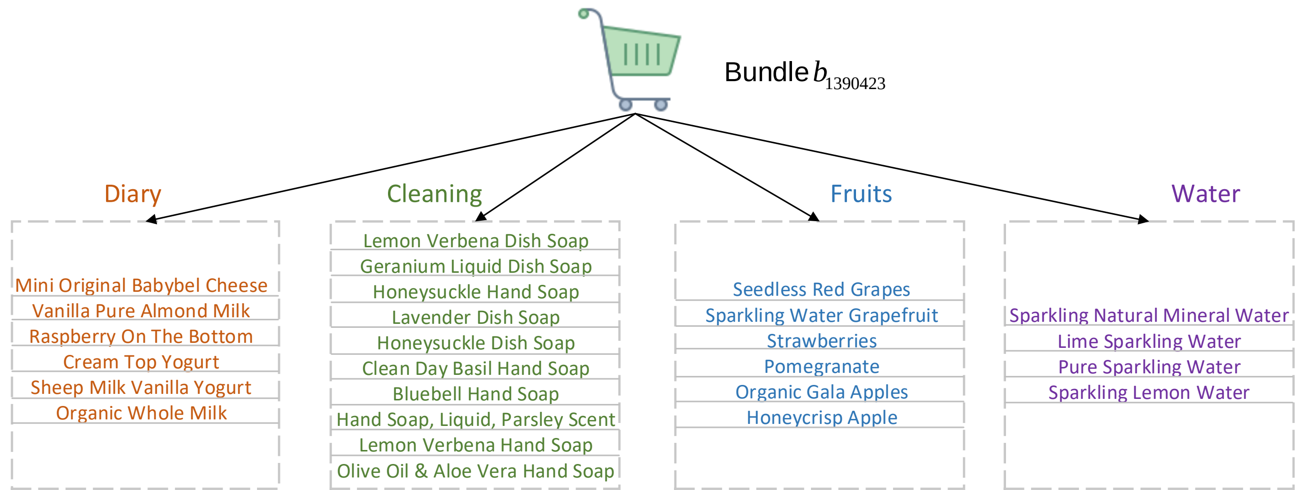

Fig. 4: Case study of the intent distribution within the bundle. (Best view in color)

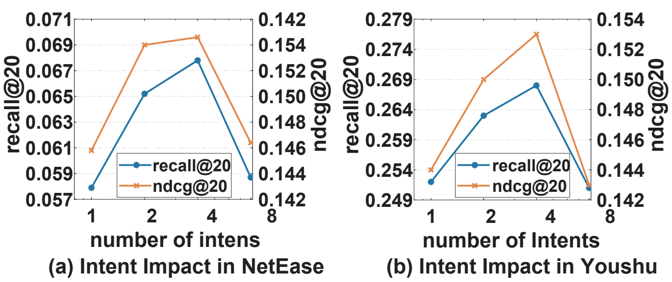

Fig. 5: Impact of Intent Number(K) for next-bundle recommendation.

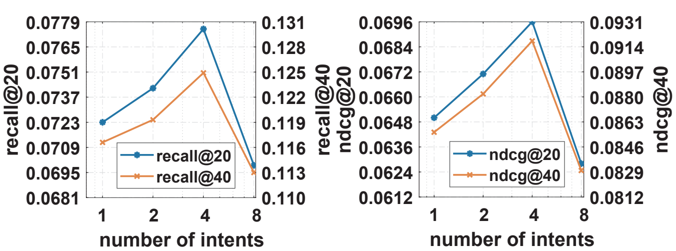

Fig. 6: Impact of Intent Number(K) for within-bundle recommendation.

items are presented with different colors to illustrate the user's multiple intents. The distribution of items under different intents are well clustered. This gives an intuitive visualization of the user's diverse intents and the effectiveness of our model in disentangling the user's intents.

Case Study. We further verify whether the intents disentangled by the HIDGN has meaning in the real-word scenarios as shown in Figure 4. We randomly select one bundle from Instacart and present names of items under different intents with different colors. From the figure, the intent within the bundle is disentangled into: "dairy", "cleaning", "fruits" and "water". This proves that the intents disentangled by HIDGN are meaningful in the real-word scenarios.

6 Conclusion

In this paper, we explore hierarchical intent disentanglement in the problem of personalized bundle recommendation

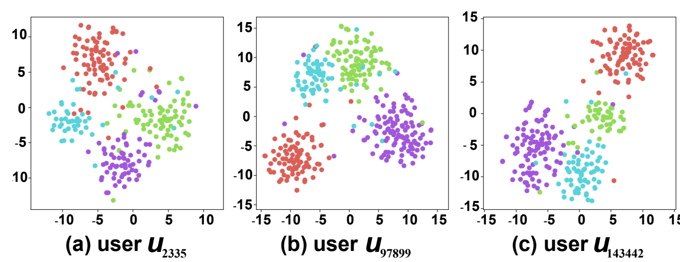

Fig. 7: t-SNE visualization of items bought by a user. Different colors present different intents. (Best view in color)

which contains two sub-tasks: next-bundle recommendation and within-bundle recommendation. We propose a novel model named HIDGN (*i.e.*, Hierarchical Intent Disentangle Graph Networks), which is capable of precisely and comprehensively capturing the diversity of the user's intent from hierarchical structure for next- and within-bundle recommendation. Specifically, 1) HIDGN disentangles user intents from user/bundle/item level and learns disentangled representations; 2) HIDGN contrasts the intents of different levels across next- and within-bundle recommendation sub-tasks, and hence optimizes the intent modeling in a self-supervised manner. A multi-task framework is further formed to unify the supervised next-bundle and within-bundle recommendation sub-tasks with self-supervised intent contrastive learning. Extensive experiments on Youshu, NetEase, and Instacart demonstrate the superiority of the proposed HIDGN.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No.62276110, Grant No.61772076, in part by CCF-AFSG Research Fund under Grant No.RF20210005, and in part by the fund of Joint Laboratory of HUST and Pingan Property & Casualty Research (HPL). The authors would also like to thank the anonymous reviewers for their comments on improving the quality of this paper.

REFERENCES

[1] S. Zhao, W. Wei, D. Zou, and X. Mao, "Multi-view intent disentangle graph networks for bundle recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 4379–4387.

[2] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 346–353.

[3] Y. Wang, S. Tang, Y. Lei, W. Song, S. Wang, and M. Zhang, "Disenhan: Disentangled heterogeneous graph attention network for recommendation," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 1605–1614.

[4] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, "Global context enhanced graph neural networks for session-based recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 169–178.

[5] Z. Wang, W. Wei, G. Cong, X.-L. Mao, X.-L. Li, and S. Feng, "Exploiting repeated behavior pattern and long-term item dependency for session-based recommendation," *arXiv preprint arXiv:2012.05422*, 2020.

[6] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, M. Qiu, and S. Feng, "Exploring global information for session-based recommendation," *arXiv preprint arXiv:2011.10173*, 2020.

[7] B. Chang, A. Karatzoglou, Y. Wang, C. Xu, E. H. Chi, and M. Chen, "Latent user intent modeling for sequential recommenders," in *Companion Proceedings of the ACM Web Conference 2023*, 2023, pp. 427–431.

[8] S. Zhao, W. Wei, X.-L. Mao, S. Zhu, M. Yang, Z. Wen, D. Chen, and F. Zhu, "Multi-view hypergraph contrastive policy learning for conversational recommendation," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 654–664.

[9] S. Zhao, W. Wei, Y. Liu, Z. Wang, W. Li, X.-L. Mao, S. Zhu, M. Yang, and Z. Wen, "Towards hierarchical policy learning for conversational recommendation with hypergraph-based reinforcement learning," *arXiv preprint arXiv:2305.02575*, 2023.

[10] L. Zhang and R. Li, "A large-scale friend suggestion architecture," in *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2022, pp. 1–9.

[11] Y. Liu, M. Xie, and L. V. Lakshmanan, "Recommending user generated item lists," in *Proceedings of the 8th ACM Conference on Recommender systems*, 2014, pp. 185–192.

[12] T. Zhu, P. Harrington, J. Li, and L. Tang, "Bundle recommendation in ecommerce," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 657–666.

[13] M. Grbovic, V. Radosavljevic, N. Djuric, N. Bhamidipati, J. Savla, V. Bhagwan, and D. Sharp, "E-commerce in your inbox: Product recommendations at scale," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, pp. 1809–1818.

[14] D. Cao, L. Nie, X. He, X. Wei, S. Zhu, and T.-S. Chua, "Embedding factorization models for jointly recommending items and user generated lists," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 585–594.

[15] L. Chen, Y. Liu, X. He, L. Gao, and Z. Zheng, "Matching user with item set: Collaborative bundle recommendation with deep attention network." in *IJCAI*, 2019, pp. 2095–2101.

[16] J. Chang, C. Gao, X. He, D. Jin, and Y. Li, "Bundle recommendation with graph convolutional networks," in *Proceedings of the 43rd international ACM SIGIR conference on Research and development in Information Retrieval*, 2020, pp. 1673–1676.

[17] D. T. Le, H. W. Lauw, and Y. Fang, "Basket-sensitive personalized item recommendation." IJCAI, 2017.

[18] G. Liu, Y. Fu, G. Chen, H. Xiong, and C. Chen, "Modeling buying motives for personalized product bundle recommendation," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 3, pp. 1–26, 2017.

[19] M. Wan, D. Wang, J. Liu, P. Bennett, and J. McAuley, "Representing and recommending shopping baskets with complementarity, compatibility and loyalty," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1133–1142.

[20] Z. Liu, M. Wan, S. Guo, K. Achan, and P. S. Yu, "Basconv: Aggregating heterogeneous interactions for basket recommendation with graph convolutional neural network," in *Proceedings of the 2020 SIAM International Conference on Data Mining*. SIAM, 2020, pp. 64–72.

[21] Z. Liu, X. Li, Z. Fan, S. Guo, K. Achan, and S. Y. Philip, "Basket recommendation with multi-intent translation graph neural network," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 728–737.

[22] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, "Disentangled graph collaborative filtering," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 1001–1010.

[23] O. Sar Shalom, N. Koenigstein, U. Paquet, and H. P. Vanchinathan, "Beyond collaborative filtering: The list recommendation problem," in *Proceedings of the 25th international conference on world wide web*, 2016, pp. 63–72.

[24] J. Bai, C. Zhou, J. Song, X. Qu, W. An, Z. Li, and J. Gao, "Personalized bundle list recommendation," in *The World Wide Web Conference*, 2019, pp. 60–71.

[25] S. Rendle, C. Freudenthaler, Z. Gantner, and L. B. Schmidt-Thieme, "Bayesian personalized ranking from implicit feedback," in *Proc. of Uncertainty in Artificial Intelligence*, 2014, pp. 452–461.

[26] Y. He, J. Wang, W. Niu, and J. Caverlee, "A hierarchical self-attentive model for recommending user-generated item lists," in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1481–1490.

[27] C. Li, Y. Lu, W. Wang, C. Shi, R. Xie, H. Yang, C. Yang, X. Zhang, and L. Lin, "Package recommendation with intra-and inter-package attention networks," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 595–604.

[28] Y.-L. Chen, K. Tang, R.-J. Shen, and Y.-H. Hu, "Market basket analysis in a multiple store environment," *Decision support systems*, vol. 40, no. 2, pp. 339–354, 2005.

[29] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, "Design of a shopbot and recommender system for bundle purchases," *Decision Support Systems*, vol. 42, no. 3, pp. 1974–1986, 2006.

[30] Q. Liu, Y. Ge, Z. Li, E. Chen, and H. Xiong, "Personalized travel package recommendation," in *2011 IEEE 11th International Conference on Data Mining*. IEEE, 2011, pp. 407–416.

[31] M. Xie, L. V. Lakshmanan, and P. T. Wood, "Breaking out of the box of recommendations: from items to packages," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 151–158.

[32] A. Parameswaran, P. Venetis, and H. Garcia-Molina, "Recommendation systems with complex constraints: A course recommendation perspective," *ACM Transactions on Information Systems (TOIS)*, vol. 29, no. 4, pp. 1–33, 2011.

[33] O. Barkan and N. Koenigstein, "Item2vec: neural item embedding for collaborative filtering," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2016, pp. 1–6.

[34] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013.

[35] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[36] J. Ma, P. Cui, K. Kuang, X. Wang, and W. Zhu, "Disentangled graph convolutional networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 4212–4221.

[37] Y. Liu, X. Wang, S. Wu, and Z. Xiao, "Independence promoted graph disentangled networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4916–4923.

[38] Y. Yang, Z. Feng, M. Song, and X. Wang, "Factorizable graph convolutional networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 20 286–20 296, 2020.

[39] J. Wu, X. Wang, F. Feng, X. He, L. Chen, J. Lian, and X. Xie, "Self-supervised graph learning for recommendation," in *SIGIR*, 2021, pp. 726–735.

[40] P. Velickovic, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, "Deep graph infomax." *ICLR (Poster)*, p. 4, 2019.

[41] R. Linsker, "Self-organization in a perceptual network," *Computer*, pp. 105–117, 1988.

[42] Z. Peng, W. Huang, M. Luo, Q. Zheng, Y. Rong, T. Xu, and J. Huang, "Graph representation learning via graphical mutual information maximization," in *WWW*, 2020, pp. 259–270.
[43] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *ICML*. PMLR, 2020, pp. 4116–4126.
[44] J. Yu, H. Yin, X. Xia, T. Chen, L. Cui, and Q. V. H. 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*, 2022, pp. 1294–1303.
[45] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," *arXiv preprint arXiv:1205.2618*, 2012.
[46] ——, "Bpr: Bayesian personalized ranking from implicit feedback," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 2009, pp. 452–461.
[47] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 2019, pp. 165–174.
[48] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *European semantic web conference*. Springer, 2018, pp. 593–607.
[49] Q. Deng, K. Wang, M. Zhao, Z. Zou, R. Wu, J. Tao, C. Fan, and L. Chen, "Personalized bundle recommendation in online games," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 2381–2388.
[50] T. Avny Brosh, A. Livne, O. Sar Shalom, B. Shapira, and M. Last, "Bruce: Bundle recommendation using contextualized item embeddings," in *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, pp. 237–245.

Ding Zou received the bachelor's degree at Huazhong University of Science and Technology, Wuhan, China, in 2021. He is currently pursuing the master's degree with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His main research interests include recommendation systems, knowledge graph and data mining techniques.

Sen Zhao received the bachelor's degree at Huazhong University of Science and Technology, Wuhan, China, in 2021. He is currently a Ph.D. student in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His main research interests include recommendation systems, natural language processing and reinforcement learning techniques.

Wei Wei received the PhD degree from Huazhong University of Science and Technology, Wuhan, China, in 2012. He is currently a professor with the School of Computer Science and Technology, Huazhong University of Science and Technology. He was a research fellow with Nanyang Technological University, Singapore, and Singapore Management University, Singapore. His current research interests include artificial intelligence, natural language processing, information retrieval, data mining, machine learning, and social computing and recommender system.

Xian-ling Mao received the PhD degree from Peking University, in 2013. He is currently an associate professor of computer science with the Beijing Institute of Technology. He works in the fields of machine learning and information retrieval. His current research interests include topic modeling, learning to hashing, and question answering. Dr. Mao is a member of the IEEE Computer Society and a member of the Association for Computing Machinery (ACM).

Ruixuan Li received the B.S., M.S., and Ph.D. degrees in computer science from Huazhong University of Science and Technology, Wuhan, China, in 1997, 2000, and 2004, respectively. He is a professor with the School of Computer Science and Technology, Huazhong University of Science and Technology. He was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, from 2009 to 2010. His research interests include cloud and edge computing, big data management, and distributed system security. Prof. Li is a member of ACM.

Dangyang Chen received the Ph.D. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2007. He is currently the CTO in Ping An Property Insurance(Group) Company. His current research interests include Artificial Intelligence, Computer Vision(Object detection, Optical Character Recognition, *etc.*), Natural language Processing, and Knowledge graphs (social network analysis, risk management, *etc.*).

Rui Fang received the Ph.D. degree from Michigan State University, Michigan State, USA, in 2014. He is currently in charge of AI fields in Ping An Property Insurance(Group) Company. His current research interests include Artificial Intelligence, Computer Vision(Object detection, Optical Character Recognition, *etc.*), Natural language Processing, and Knowledge graphs (social network analysis, risk management, *etc.*).

Yuanyuan Fu received the B.S and M.S degrees from Xidian University, Xi'an, China, in 2006 and 2010. He used to work in CEMNET laboratory of Nanyang Technology University, engaged in Pattern Recognition and Machine Learning related fields. Since 2018, he has been in charge of AI fields in Ping An Property Insurance(Group) Company. His current research interests include Artificial Intelligence, Computer Vision(Object detection, Optical Character Recognition, *etc.*), Natural language Processing, and Knowledge graphs (social network analysis, risk management, *etc.*).