Basket Representation Learning by Hypergraph Convolution on Repeated Items for Next-basket Recommendation

Yalin Yu[1*], **Enneng Yang**[1*], **Guibing Guo**[1†], **Linying Jiang**[1] and **Xingwei Wang**[2]
[1]Software College, Northeastern University, China
[2]School of Computer Science and Engineering, Northeastern University, China
{yalinyu,ennengyang}@stumail.neu.edu.cn, {guogb,jiangly}@swc.neu.edu.cn
wangxw@mail.neu.edu.cn

Abstract

Basket representation plays an important role in the task of next-basket recommendation. However, existing methods generally adopt pooling operations to learn a basket's representation, from which two critical issues can be identified. First, they treat a basket as a set of items independent and identically distributed. We find that items occurring in the same basket have much higher correlations than those randomly selected by conducting data analysis on two real datasets. Second, although some works have recognized the importance of items repeatedly purchased in multiple baskets, they ignore the correlations among the repeated items in the same basket, whose importance is shown by our data analysis. In this paper, we propose a novel Basket Representation Learning (BRL) model by leveraging the correlations among intra-basket items. Specifically, we first connect all the items (in a basket) as a hyperedge, where the correlations among different items can be well exploited by hypergraph convolution operations. Meanwhile, we also connect all the repeated items in the same basket as a hyperedge, whereby their correlations can be further strengthened. We generate a negative (positive) view of the basket by data augmentation on repeated (non-repeated) items, and apply contrastive learning to force more agreements on repeated items. Finally, experimental results on three real datasets show that our approach performs better than eight baselines in ranking accuracy.

1 Introduction

Next basket recommendation (NBR) has been widely used in a variety of recommendation scenarios [Yu *et al.*, 2016; Bai *et al.*, 2018], including community group purchases, grocery stores, and so on. It aims to learn accurate user preference from a user's historical basket sequence, and then to recommend a basket of items that the user may purchase next. Most of the existing NBR methods focus on modeling user preference from the transition of sequential baskets via typical neural networks, such as RNN [Le *et al.*, 2019; Schulz *et al.*, 2020; Qin *et al.*, 2021]. Although baskets are taken as the essential input of RNN, their representation learning has not attracted much attention in the literature. As a matter of fact, existing methods generally adopt pooling operations (max pooling, average pooling) [Hu and He, 2019; Qin *et al.*, 2021; Shen *et al.*, 2022] to learn a basket's representation by combining all the basket members together.

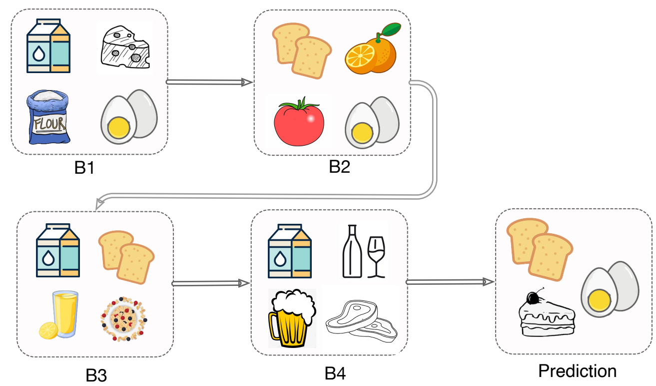

Figure 1: A toy example to explain our assumptions, where a user purchased four baskets and we make recommendations accordingly.

Two critical issues can be identified from the literature. Firstly, the underlying assumption behind pooling operations is that all the items in a basket are independent and identically distributed, which may not hold in real applications. We assume that it is not random that a number of items appear in the same basket, and there are positive correlations among these items, which are valuable for basket representation learning. Secondly, although some works [Qin *et al.*, 2021; Ariannezhad *et al.*, 2022; Katz *et al.*, 2022] have recognized the importance of items that are repeatedly purchased in multiple baskets, they ignore the correlations of repeated items[1] in the same basket. We assume that such kind of item correlations is even more significant than that of overall intra-basket items for basket representation learning.

A toy example to explain our assumptions is given in Fig. 1. Suppose a user purchased 4 baskets in the past, labeled as `B1`, `B2`, `B3` and `B4` respectively. We note that:

[1]We use term 'repeated items' exchangeable with term 'repurchased items' throughout this paper.

(1) *Milk* and *Bread* are repurchased in basket B3, and such a co-occurrence inspires us that the two repeated items are somehow correlated in meeting user preference as a bundle. Hence, when we find *Milk* is purchased in basket B4, it is reasonable to predict that the user will buy *Bread* in the next time step. (2) *Egg* is purchased more frequently than many other items, leaving us a hint for the future suggestion. (3) Since *Milk* has been predicted in case (1), we make *Cake* as a recommendation since items in baskets B1 and B3 can be broadly classified into the same category 'dessert/snacks', indicating the correlations for all the items in the same basket.

To further validate our assumptions, we conduct data analysis on two real datasets in Sec. 2. The results show that: (1) items within a basket have a higher correlation than those randomly selected, validating our first assumption; (2) repeated items in a basket have a higher correlation than all the items in the same basket, demonstrating our second assumption.

Hence, this paper proposes a novel Basket Representation Learning (BRL) model by making use of item correlations within a basket, especially those of repeated items. Specifically, we first connect all the items (in a basket) as a hyperedge, where each item serves as a node. Multiple hyperedges construct a hypergraph. Then, the correlations among items can be well exploited by hypergraph convolution operations [Yi and Park, 2020]. Meanwhile, we also connect all the repeated items in the same basket as a hyperedge, whereby their correlations can be further strengthened by propagating messages between nodes. Lastly, we generate a negative view of the basket by randomly replacing repeated items with non-repeated ones, and a positive view by removing weakly correlated items from the basket. Contrastive learning [Zhang *et al.*, 2021] is then applied to force the agreements more on repeated items when representing the basket.

The main contributions are summarized as follows:

- We propose two assumptions to focus more on the correlations of intra-baseket items, especially those of repeated items. These assumptions are validated by our data analysis on two real datasets.
- We propose a novel Basket Representation Learning (BRL) model to realize the two assumptions by applying hypergraph convolution operations on the hyperedges, which are constructed by connecting all items and repeated items in the same basket. We also designed a contrastive learning task to highlight more on repeated items for better basket representation.
- Experimental results on three real datasets demonstrate that our proposed method can achieve better performance in comparison with eight competing models.

2 Assumption Validation

In this section, we will conduct data analysis on two real datasets to validate our assumptions. Specifically, we first choose the dataset Instacart(Dunnhumby)[2] since it contains 30(20) categories of items. Then, we define the correlation as the percentage of items within a basket belonging to the same category, and take the greatest percentage as the correlation if items belong to multiple categories. We denote it as 'basket'. As a comparison, we generate the same number of baskets, each of which has the same average number (i.e., 8(7) as given in Tab. 2) of items as that in two real datasets. All the items in a generated basket are randomly selected from the categories users purchased before. After that, we calculate the correlation of the generated baskets. We denote this baseline as 'random'. Lastly, we construct a new basket by retaining only the repeated items from the original basket, and compute the percentage of repurchased items that belong to the same category. For easy discussion, we denote this variant as 'repeated items'.

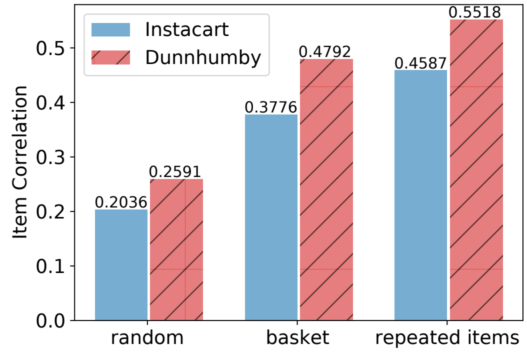

Figure 2: Item correlation within a basket on the Dunnhumby and Instacart datasets.

The results are illustrated in Fig. 2. It shows that randomly selected items have the smallest item correlation, i.e., two random items have $20\%(25\%)$ percentage located in the same category, given the distribution of categories purchased by users. Our first assumption is valid since 'basket' has a higher correlation than 'random', indicating the importance of items appearing in the same basket. Furthermore, 'repeated items' has an even higher correlation (i.e., $45\%(55\%)$) than the case of 'basket' (i.e., $37\%(47\%)$), implying the validation of our second assumption. Therefore, both of our assumptions hold in real applications, which inspires us to make better use of those correlations for basket representation learning.

3 Our BRL Model

We introduce a number of notations for the sake of discussion. Let $I = \{i_1, i_2, i_3, \ldots, i_{|I|}\}$ denote the set of items and $U = \{u_1, u_2, u_3, \ldots, u_{|U|}\}$ denote all users, where $|I|$ and $|U|$ represent the number of items and users, respectively. We define $\mathbf{X} \in \mathbb{R}^{|I| \times d}$ as the embedding matrix corresponding to items, where d represents the embedding dimension. For each user $u \in U$, we use $B^u = \{b_1^u, b_2^u, b_3^u, \ldots, b_m^u\}$ to denote the historical purchased baskets with length m, where $b_k^u = \{i_{k,1}, i_{k,2}, \ldots, i_{k,j}, \ldots, i_{k,n}\}$ represents the k-th basket purchased by the user u and $i_{k,j} \in I$. We denote the basket data of all users as $B = [B^1, B^2, \ldots, B^{|U|}]$. The task of this paper is to predict the next basket b_{m+1}^u for a given user u.

3.1 Model Overview

This paper propose a novel Basket Representation Learning (BRL) model by leveraging the correlations among intra-

[2] We defer the dataset description to Sec. 4.1. The TaFeng dataset is not used since it does not have category information.

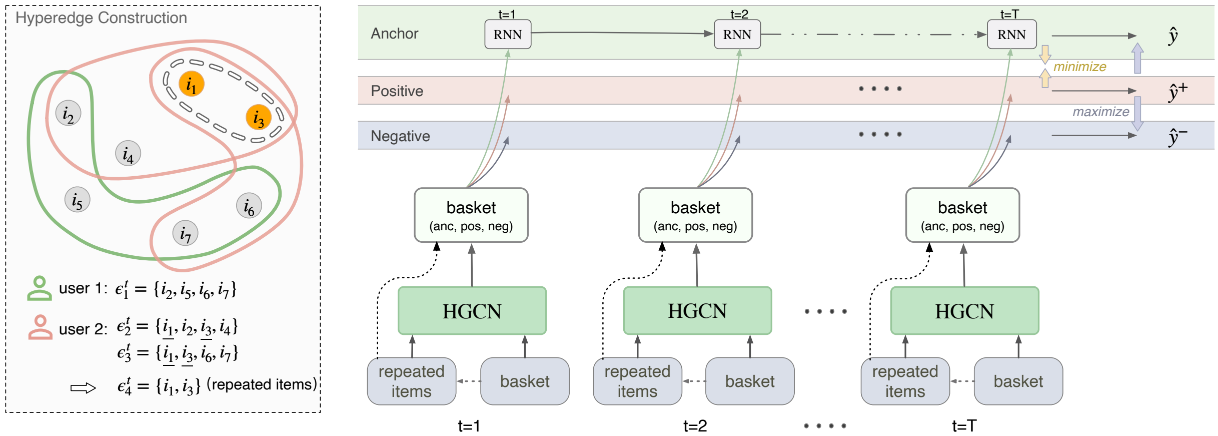

Figure 3: Examples of hypergraph construction and the overall framework of the proposed Basket Representation Learning (BRL) model.

basket items for next-basket recommendation. First, as shown on the left side of Fig. 3, we connect each basket with a hyperedge, and at the same time connect the repeated items in a basket with a hyperedge. Then construct all the basket data in the same period into a hypergraph (Sec. 3.2). Next, as shown on the right side of Fig. 3, we perform multi-layer hypergraph convolutions networks (HGCN) on the hypergraph to explore the correlation between items (Sec. 3.3). Finally, we perform data augmentation on repeated items to construct positive and negative views of the anchor basket, and then use contrastive learning techniques to make anchor basket representations more concerned about repeated items (Sec. 3.4). In addition, the representation of the basket is fed into the Recurrent Neural Network (RNN) to extract user preferences and recommend items for the next basket (Sec. 3.4).

3.2 Hypergraph Construction and Representation

Here, we first illustrate how to transform the basket data into a hypergraph, and how to represent a hypergraph.

Hypergraph Construction. Considering that the characteristics of an item have a certain timeliness, for example, when a mobile phone is first released, its user attention is relatively high, but after one year, the item becomes unpopular. Therefore, we group the baskets according to the timestamp, assuming that the characteristics of an item will not change within a certain period. To capture the correlations among intra-basket items, this paper converts all the basket data for each period into a hypergraph. Specifically, to establish the connection between the items in the basket and emphasize the relationship between the repeatedly purchased items, we first use hyperedges to connect each basket and the repeated items. An example is shown on the left side of Fig. 3, the ϵ_1^t (solid green line) is a hyperedge, which represents a basket data (i.e., items $\{i_2, i_5, i_6, i_7\}$) of user 1 within the period t. And the ϵ_4^t (white dotted line) is also a hyperedge, indicating that items i_1 and i_3 in user 2's latest basket are previously purchased items. Then, we construct all hyperedges (baskets and subsets of repeated items) $E^t = \{\epsilon_1^t, \epsilon_2^t, \ldots, \epsilon_{|S|}^t\}$ and related vertices (item subset) $V^t \subset I$ in period t into a hypergraph G^t, where $|S|$ represents the number of hyperedges. Let $G = \{G^1, G^2, \ldots, G^t, \ldots, G^T\}$ represent a series of hypergraphs, where $G^t = \{V^t, E^t\}$ is constructed from all the baskets purchased during time period t, and T represents the total number of time periods.

Hypergraph Representation. To facilitate storage and computation of hypergraphs, each hypergraph $G^t \in G$ can be represented by an incidence matrix $\mathbf{H}^t$ of size $|V^t| \times |E^t|$, where $|V^t|$ and $|E^t|$ denote the number of vertices and hyperedges in the G^t, respectively. When a hyperedge (basket) $\epsilon \in E^t$ is incident with a vertex (item) i, in other words, when item i is connected by ϵ, we set $H_{i\epsilon}^t = 1$, otherwise $H_{i\epsilon}^t = 0$. Each hyperedge $\epsilon \in E^t$ is assigned a positive weight $W_{\epsilon\epsilon}^t$, and all the weights formulate a diagonal matrix $\mathbf{W}^t \in \mathbb{R}^{|S| \times |S|}$. $W_{\epsilon\epsilon}$ is set to 1 when $\epsilon \in E^t$ is composed of a basket (e.g., ϵ_1^t in Fig. 3) and is set to 2 if composed of a repeated sub-basket (e.g., ϵ_4^t in Fig. 3). Degree of vertices and hyperedges $\mathbf{D}^t \in \mathbb{R}^{|V^t| \times |V^t|}$ and $\mathbf{F}^t \in \mathbb{R}^{|S| \times |S|}$ are defined as $D_{ii}^t = \sum_{\epsilon=1}^{|S|} W_{\epsilon\epsilon}^t H_{i\epsilon}^t$ and $F_{\epsilon\epsilon}^t = \sum_{i=1}^{|V^t|} H_{i\epsilon}^t$, respectively.

3.3 Hypergraph Convolutional Network (HGCN)

The core idea of hypergraph convolution [Feng *et al.*, 2019] is to propagate and aggregate information between neighbor nodes on the hypergraph structure so that similar nodes can obtain similar representations (higher correlation). Therefore, to better explore the correlations among intra-basket items in this paper, we perform a multi-layer hypergraph convolution network (HGCN) on the obtained hypergraph $G^t \in G$.

We first obtain the initial representation of the whole item set $\mathbf{X}$. For hypergraph G^t, given the hyperedges E^t and item's initial latent representation (or embedding) $\mathbf{X}^{t,(0)}$, we design L convolutional layers for each period t. Then we define a convolution operation over the hypergraph, the representation update rule for item embeddings at layer $(l+1)$ is given by:

$$x_i^{t,(l+1)} = \sum_{j=1}^{|V^t|} \sum_{\epsilon=1}^{|E^t|} \tilde{H}_{i\epsilon}^t W_{\epsilon\epsilon}^t \tilde{H}_{j\epsilon}^t x_j^{t,(l)}, \qquad (1)$$

where $\mathbf{x}_j^{t,(l)}$ is the representation of the j-th vertex in the (l)-

th layer, $\mathbf{x}_j^{t,(0)} \in \mathbb{R}^d$ represents the j-th row of $\mathbf{X}^{t,(0)}$. $\tilde{\mathbf{H}}^t$ is a variant of incidence mathrix $\mathbf{H}^t$ and $\tilde{H}_{j\epsilon}^t$ reflects the importance of item j in the hyperedge ϵ. In other words, each vertex has a different importance to a hyperedge ϵ. Thus, we draw on the idea of term frequency–inverse document frequency (TF-IDF) [Havrlant and Kreinovich, 2017] to evaluate the importance of items within a basket: when item i appears more frequently in user u's historical baskets and less frequently in other baskets, it is considered more important in the current basket. The formal expression is as follows:

$$\tilde{H}_{i\epsilon}^t = H_{i\epsilon}^t \cdot \frac{\alpha_i^u}{\alpha_i}, \tag{2}$$

where u is the creator of hyperedge ϵ, α_i^u denotes the user-wise frequency for a given item i and α_i refers to the global frequency of item i. $H_{i\epsilon}^t = 0$ if item i is not in a hyperedge $\epsilon \in E^t$. The detailed calculation rules of α_i and α_i^u are as:

$$\alpha_i = \frac{\sum_{B_j \in B} \mathbb{I}(B_j)}{|B|},\ \alpha_i^u = \frac{\sum_{B_j^u \in B^u} \mathbb{I}(B_j^u)}{|B^u|}, \tag{3}$$

where $|B|$ and $|B^u|$ represent the total number of baskets of all users and the total number of baskets of user u, respectively. $\mathbb{I}(B_j)$ is an indicative function, if item i belongs to a basket B_j, then $\mathbb{I}(B_j)$ is equal to 1, otherwise it is equal to 0. It can be seen that the global popularity of items is used as a penalty indicator. When an item is more popular, its purchase can less reflect the connection with other items in the basket and reflect the user's personalized preferences.

The hypergraph convolution rule for all vertices in the hypergraph G^t, that is, the matrix form of Eq. 1 can be written as: $\mathbf{X}^{t,(l+1)} = \tilde{\mathbf{H}}^t \mathbf{W}^t \tilde{\mathbf{H}}_t^\top \mathbf{X}^{t,(l)}$. However, the spectral radius of $\tilde{\mathbf{H}}^t \mathbf{W}^t \tilde{\mathbf{H}}_t^\top$ is not constrained in this formula, which means that the scale of $\mathbf{X}^{t,(l+1)}$ may change and leads to numerical instability in the optimization. Similar to [Bai *et al.*, 2021], we apply a symmetric normalization, and the matrix form of the final vertex hypergraph convolution rule is expressed as follows:

$$\mathbf{X}^{t,(l+1)} = (\mathbf{D}^t)^{-1} \tilde{\mathbf{H}}^t \mathbf{W}^t (\mathbf{F}^t)^{-1} \tilde{\mathbf{H}}_t^\top \mathbf{X}^{t,(l)}, \tag{4}$$

where $\mathbf{D}^t$ and $\mathbf{F}^t$ are degrees of vertices and hyperedges, respectively, which are defined in Sec. 3.2.

After passing through the L hypergraph convolution layers on hypergraph G^t, vertex (i.e., item) embedding can stack more higher-order messages propagated from multi-hop neighbors, thus effectively establishing correlations among intra-basket items. We obtain the final item embeddings $\mathbf{X}^t$ in Eq. 5 based on the mean pooling over all layers, i.e.,:

$$\mathbf{X}^t = \frac{1}{L+1} \sum_{l=0}^{L} \mathbf{X}^{t,(l)} \tag{5}$$

The representation of items in other periods $t \in \{1, 2, \ldots, T\}$ can be obtained in the manner of Eq. 5. So far, each item representation we have obtained has fully considered the correlation with other items in the basket. In particular, our proposed method adds a new hyperedge connection to the repeated items in the basket, further strengthening the relationship between items.

3.4 Basket Representation Learning

As shown in Sec. 2, repeated items in the basket are usually more closely related than other items, and play a more important role in basket representation learning. In this section, we generate both positive and negative views for the anchor basket through data augmentation, and emphasize the effect of removing repeated items on the basket representation through contrastive learning.

Data Augumentation. Suppose a basket $b_k^u = [\underline{i_1(\text{milk})}, i_2(\text{egg}), \underline{i_3(\text{bread})}, i_4(\text{toy})]$ is given by user u, we regard it as an anchor basket and re-denote it by b^{anc} for expressiveness. Then, we create a negative b^{neg} (positive b^{pos}) view from the anchor basket by making changes on repeated (non-repeated) items. Let's assume that items i_1 and i_3 are previously purchased items (i.e., repeated) in this basket. (i) repeated items in a basket are usually more closely related and representative of the basket and the user's preference. The basket semantics will greatly shift if we perturb this part of the repeated purchase items. Hence, we replace all the repeated items in the anchor basket with random items to get a **negative view** of anchor basket, e.g., $b^{neg} = [i_5(\text{rose}), i_2(\text{egg}), i_6(\text{pencil}), i_4(\text{toy})]$, where items i_5 and i_6 are randomly sampled from the item set. (ii) due to the diversity and randomness of user behaviors, a basket may contain some items that are less relevant to the basket as a whole, such as an item i_4(toy) in the basket b^{anc} mentioned above, which we call noise items. Therefore, removing these noise items does not affect the representation of basket semantics. The basket we get after removing noise items from the anchor basket is taken as the **positive view**, e.g., $b^{pos} = [i_1(\text{milk}), i_2(\text{egg}), i_3(\text{bread})]$. More specifically, after obtaining the whole representation $\mathbf{e}_{b^{anc}}$ of the anchor basket (from Eq. 6), we calculate and sort its correlation with the representation $\mathbf{X}_i^t$ (from Eq. 5) of each item i in the basket, and rank the least relevant top 20% items removed. In this way, we obtain the anchor basket b^{anc}, and its positive b^{pos} and negative b^{neg} basket views.

Basket Encoding. Now, we need to fuse the representations of the items in the basket to get an overall representation of the basket. Since we have emphasized the importance of each item in the basket to the hyperedge (that is, the basket) in the hypergraph convolution, see Eq. 2 for details. At the same time, the item representation obtained after hypergraph convolution (i.e., Eq. 5 in Sec. 3.3) has captured the relevance of repeated items between baskets, as well as emphasizing the closer connection between items purchased repeatedly. So we can use the simplest mean pooling to take the average representation of all items in the basket b^{anc} as the whole representation $\mathbf{e}_{b^{anc}}$ of the basket in hypergraph G^t. That is,

$$\mathbf{e}_{b^{anc}} = \frac{1}{|b^{anc}|} \sum_{i \in b^{anc}} \mathbf{X}_i^t \tag{6}$$

where $|b^{anc}|$ represents the number of items contained in the anchor basket b^{anc} in hypergraph G^t. Similarly, we can obtain representations $\mathbf{e}_{b^{pos}}$ and $\mathbf{e}_{b^{neg}}$ for the positive and negative baskets, respectively.

Contrastive Learning. To better capture the characteristics of dynamic changes in user u's preferences, we use RNN to model the conversion process of basket sequences B^u. Specifically, we input the representation $\mathbf{e}_{b^{anc}}$ of the user's basket $b^{anc} \in B^u$ at time t into the RNN unit, that is,

$$\mathbf{h}_t^{anc} = \mathrm{RNN}(\mathbf{e}_{b^{anc}}, \mathbf{h}_{t-1}^{anc}), \tag{7}$$

where RNN unit can be gated recurrent unit (GRU) [Chung *et al.*, 2014] and Long short-term memory (LSTM) [Hochreiter and Schmidhuber, 1997], $\mathbf{h}_{t-1}^{anc}$ is the hidden state of the RNN unit at time $t-1$. The last hidden state $\mathbf{h}_T^{anc}$ is employed as a composition of the user's preference. After that, a softmax function is applied to calculate the item scores:

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{h}_T^{anc}), \tag{8}$$

where $\hat{\mathbf{y}} \in \mathbb{R}^{|I|}$ represents the probability distribution that a user will be interested in the entire set of items. We can also calculate the probability distribution $\hat{\mathbf{y}}^+$ ($\hat{\mathbf{y}}^-$) of the user's preference on items from the positive (negative) basket view. Therefore, user representations learned from a positive view should be intuitively pulled closer to the original user representation. In contrast, representations learned from a negative view without repurchased concepts should be pushed away from the original result. We finally use triplet margin loss to measure the relative similarity between samples:

$$\mathcal{L}_{cl} = \sum_{i=1}^{|I|} \max\big(d(\hat{\mathbf{y}}_i, \hat{\mathbf{y}_i}^+) - d(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_i^-) + 1, 0\big), \tag{9}$$

where the distance function $d(\cdot,\cdot)$ is set to the L_2 norm since user representations generated by the same user preference representation are in the same embedding space.

3.5 Model Optimization

We train our BRL model by optimizing the main task on implicit feedback together with the contrastive learning loss (i.e., Eq. 9). Given a training set $B^u = \{b_1^u, b_2^u, b_3^u, \ldots, b_m^u\}$ of user u, $\mathbf{y}^u$ is the ground-truth next purchased basket for user u and $\hat{\mathbf{y}}^u$ represents the next basket predicted by our recommender. Inspired by [Hu and He, 2019], we define our optimized objective of the recommendation task as follows:

$$\mathcal{L}_{rec} = \sum_{u \in U} M(\mathbf{y}^u, \hat{\mathbf{y}}^u) + \gamma \cdot P(\mathbf{y}^u, \hat{\mathbf{y}}^u), \tag{10}$$

where γ is a hyperparameter, $M(\cdot,\cdot)$ represents the weighted mean square loss function and $P(\cdot,\cdot)$ is partitioned set margin constraint that maximizes the pair-wise margin between the predicted correct and error sets. $M(\mathbf{y}^u, \hat{\mathbf{y}}^u)$ is obtained by the following formula:

$$M(\mathbf{y}^u, \hat{\mathbf{y}}^u) = \sum_{m=1}^{|I|} \frac{(e_{y,m} - e_{\hat{y},m})^2}{\alpha_{i_m}}, \tag{11}$$

where $e_{y,m}$ is the m-th entry of $\mathbf{y}^u$ and $e_{\hat{y},m}$ is the m-th entry of $\hat{\mathbf{y}}^u$. α_{i_m} is the frequency of element i_m in the train set in Eq. 3, which is used to make the contributions of different elements to the loss equilibrium. Taking account of a pair-wise margin between the predicted correct and incorrect sets, we define the partitioned set margin in the next basket using the following function:

$$P(\mathbf{y}^u, \hat{\mathbf{y}}^u) = \frac{1}{|\mathbf{z}^u||\overline{\mathbf{z}}^u|} \sum_{i \in \mathbf{z}^u} \sum_{j \in \overline{\mathbf{z}}^u} \mathrm{marg}(e_{z^u,i}, e_{\overline{z}^u,j}) \tag{12}$$

in which $\mathbf{z}^u$ is the matched set of $\mathbf{y}^u \cap \hat{\mathbf{y}}^u$ and $\overline{\mathbf{z}}^u = \overline{\mathbf{y}}^u \cap \hat{\mathbf{y}}^u$, $\overline{\mathbf{y}}^u$ represents all the element not appearing in $\mathbf{y}^u$, $e_{z^u,i}$ and $e_{\overline{z}^u,j}$ is the i-th element in $\mathbf{z}^u$ and the j-th in $\overline{\mathbf{z}}^u$, separately. The marg$(\cdot)$ calculates the pair-wise margin. We adopt multi-task learning optimizing the recommendation loss and contrastive loss jointly. The joint loss is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda \cdot \mathcal{L}_{cl} \tag{13}$$

where λ controls the strength of contrastive learning loss $\mathcal{L}_{cl}$. We can use a gradient descent optimization algorithm to compute the loss according to Eq. 13 and backpropagate the gradients to update the parameters of our BRL model.

4 Experimentation

4.1 Experimental Settings

Datasets. We use three real-world datasets in our experiments: TaFeng[3], Dunnhumby[4] and Instacart[5]. TaFeng is a Chinese grocery store dataset, which contains shopping transactions in four months. Dunnhumby consists of household-level transactions from most frequent shoppers, and it is used in [Faggioli *et al.*, 2020a; Ariannezhad *et al.*, 2022]. Instacart records over 3 million online purchases on the Instacart grocery service. Note that all users with less than 3 baskets are removed to ensure temporal sequential signals exist in the records (applied in [Hu and He, 2019; Ariannezhad *et al.*, 2022]). For TaFeng, we save the top 5,000 most frequent items that cover 83% items appearing in all the baskets. For Dunnhumby, we use the version with 50k users. For Instacart, we remove the least frequent items. We treat all the items bought in the same order as a basket. The overall time span of Dunnhumby and Tafeng is greater than 4 months, and we divide the time span by weeks; Instacart divides it by days because its overall time span is less than 1 month. The statistics of our datasets after pre-processing are shown in Tab. 2. *#users*, *#items*, and *#baskets* represent the total number of users, items, and baskets respectively. *avg. basket size* represents the average number of items in each basket. *avg. basket/user* indicates how many baskets each user owns on average. *repeated ratio (all)* is the ratio of repeated items to all purchased items. *repeated ratio (target)* is the ratio of repeated purchase items in target baskets (test set).

Comparison Methods. We compare BRL with eight baselines, including two early baselines (POP[6], FPMC [Rendle *et al.*, 2010]), five RNN-based models (DREAM [Yu *et al.*, 2016], Beacon [Le *et al.*, 2019], Sets2Sets [Hu and He, 2019], CLEA [Qin *et al.*, 2021], MBN [Shen *et al.*, 2022]), and a hypergraph based model (NBRR) [Katz *et al.*, 2022].

Evaluation Metrics. Following [Shao *et al.*, 2022; Katz *et al.*, 2022], we adopt three commonly used metrics for evaluation, namely Recall@K, NDCG@K and MAP@K, where K is set to 10 and 20 in our experiments. Generally, *greater* values of evaluation metrics indicate *better* accuracy.

[3]https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset

[4]https://www.dunnhumby.com/source-files/

[5]https://www.kaggle.com/c/instacart-market-basket-analysis

[6]It always suggests users with the most popular item.

Dataset	Metrics	POP	FPMC	DREAM	Beacon	MBN	NBRR	Sets2Sets	CLEA	BRL	Improve
TaFeng	Recall@10	2.59	3.12	7.83	8.21	9.83	10.23	10.94	11.81	**11.98**	1.44%
	NDCG@10	2.00	2.96	7.17	8.51	8.09	7.72	8.47	9.57	**9.75**	1.88%
	MAP@10	3.97	6.12	11.53	11.62	15.52	16.02	17.58	20.12	**21.29**	5.81%
	Recall@20	3.87	4.53	11.58	11.76	13.04	14.67	15.51	15.58	**16.04**	2.95%
	NDCG@20	2.51	3.57	8.02	8.51	8.87	8.85	9.57	10.36	**10.79**	4.15%
	MAP@20	2.41	4.83	6.72	7.87	9.91	8.54	8.71	10.73	**11.92**	11.09%
Dunnhumby	Recall@10	7.78	8.64	9.72	9.13	13.85	15.62	16.97	16.93	**17.48**	3.01%
	NDCG@10	8.99	9.53	11.39	10.86	14.28	14.83	15.06	15.83	**17.54**	10.80%
	MAP@10	2.73	2.95	3.54	3.42	4.64	4.52	4.62	5.14	**5.85**	13.81%
	Recall@20	10.19	11.21	13.26	12.35	19.66	21.78	22.56	22.13	**23.19**	2.79%
	NDCG@20	9.23	9.92	12.17	11.42	14.91	15.50	15.54	16.11	**18.73**	16.26%
	MAP@20	1.50	1.68	1.84	1.73	2.85	2.77	2.80	3.19	**3.61**	13.17%
Instacart	Recall@10	7.49	7.28	11.86	11.21	27.36	26.53	26.13	28.63	**29.59**	3.35%
	NDCG@10	9.36	10.07	13.13	13.39	26.45	27.32	25.65	27.62	**29.70**	7.53%
	MAP@10	3.10	3.87	5.91	6.64	9.11	9.11	8.89	10.56	**11.25**	6.53%
	Recall@20	10.23	10.05	13.73	14.04	36.61	36.72	35.58	38.37	**40.68**	6.02%
	NDCG@20	9.81	10.58	13.76	13.81	28.96	28.87	28.64	29.27	**33.12**	13.15%
	MAP@20	1.78	2.03	3.81	3.93	5.17	6.22	5.60	6.23	**7.15**	14.77%

Table 1: Performance comparison on three datasets in terms of Recall, NDCG and MAP. The best performance is bolded while the second is underlined. Column 'Improve' shows the percentage of improvements obtained by the best performance relative to the second.

Feature	TaFeng	Dunnhumby	Instacart
#users	7,384	36,240	19,825
#items	4,999	4,995	7,999
#baskets	40,393	181,199	178,643
avg. basket size	5.38	7.46	8.88
avg. basket/user	5.43	5.00	9.01
repeated ratio (all)	11.4%	20.9%	52.0%
repeated ratio (target)	14.9%	24.9%	56.6%

Table 2: The characteristics of experimental datasets

Parameter Settings. To make a fair comparison, we adopt the following settings for all methods: the embedding size is set to 128, item embedding parameters are randomly initialized in the range of (0, 1), 10% data of users are reserved as the testset and the left data is applied to train recommenders. For Sets2Sets, we set the hidden size of the encoder-decoder to 128. For CLEA, the initial temperature in Gumbel Softmax is set to 10. For MBN, we set the learning rate to 0.001 with step decay by 0.99 and the batch size to 32. For NBRR, the number of nearest neighbors is set to 100. For our model, the initial learning on the three datasets (TaFeng, Dunnhumby, Instacart) is separately set to $1e{-}4, 1e{-}5, 1e{-}5$, respectively. We empirically set hyperparameters $\lambda = 1/16$.

4.2 Performance Comparison

We can draw the following conclusions according to the results in Tab. 1. First, traditional recommendation algorithms perform poorly. We observe that the simple POP method is the worst performer on almost all metrics in all datasets, and only achieves a performance close to FPMC on Recall@10 and Recall@20 in Instacart dataset (i.e., 7.49 v.s. 7.28, 10.23 v.s. 10.05). This is because POP makes recommendations based on the popularity of items without considering the user's personalized preferences. FPMC is a classic first-order Markov-based personalized recommendation algorithm. We observed that FPMC is significantly better than POP, but there is still a clear gap with RNN-based methods. Because the Markov assumption restricts FPMC only to consider historical basket information in the recent time window, but cannot consider very long basket information.

Second, the method based on RNN and hypergraph representation significantly improves the quality of next-basket recommendations. More specifically, 1) Because RNN can capture basket information in a longer history, the early RNN-based recommendation methods DREAM and Beacon have achieved significant performance improvement over Markov-based FPMC. 2) Furthermore, MBR and NBRR model the multi-behavioral information in the next-basket recommendation and the repeated purchase cycle of items, and capture user preferences more comprehensively. 3) The Sets2Sets method considering repeated purchases of items can also effectively improve the recommendation performance, which means that repeated purchases usually reflect users' long-term preferences more. CLEA uses the contrastive learning technique to strengthen the consistency of the basket representation, and achieves the best results compared to other baselines, which also verifies that contrastive learning is effective for improving the representation of the basket.

Third, our BRL model achieves state-of-the-art performance thanks to better basket representation learning and fine-grained emphasis on the importance of repeated items. Specifically, on the TaFeng, Dunnhumby and Instacart datasets, the proposed BRL achieved an average improvement of 4.55%, 9.97% and 8.56%, respectively, and a maximum improvement of 11.09%, 16.26%, and 14.77%. The improvement in Dunnhumby and Instacart datasets is significantly greater than that of TaFeng dataset, which may be because the amount of data in the front is significantly larger than that in the back, and the average length of each basket is also longer (See Tab. 2 for details). Compared with the sub-optimal CLEA method, our method can excavate the multiple correlations between items, contributing to better basket representation. Although Sets2Sets also noticed repeated items, it is limited to singly predicting purchased items on a coarse-grained approach. Our method can not only analyze interactions among intra-basket items, but also attach importance to the repurchase behavior at a finer-grained level.

4.3 Ablation Study

We study the effectiveness of key components of the proposed model by generating several variants. Tab. 3 presents the results on two different datasets.

-HGC. This variant does not include hyperedge and hypergraph convolutional networks, and is used to study the effect of hypergraph representations and hypergraph convolutional networks on the quality of basket representations. According to Tab. 3, we observe a clear performance drop compared to BRL, illustrating the importance of considering correlations among items within the basket.

-Sub. This variant simplifies the hypergraph by removing the hyperedges generated by repurchasing sub-basket, that is, we do not consider the closer correlations among repurchased items. When the hyperedges from repurchasing sub-basket are removed, our model will achieve worse results on two datasets simultaneously. This is because more repurchased items can provide clearer clues about user interests.

-CL. This variant discards the contrastive learning task and relies merely on RNNs to model behavioral basket sequences. It achieves close performance to *(-) Sub* on both datasets in Tab. 3, validating the usefulness of contrastive learning and hyperedges on repeated items.

-DA. This variant by replacing our data augmentation with a straightforward method, i.e., negative samples are constructed by randomly replacing any items. It is significantly worse than our method. This is because random replacement cannot capture the influence of repeatedly purchased items on the basket representation. In contrast, our method effectively captures such an effect by replacing repeated items to construct negative samples.

BRL(λ). This variant analyzes the impact of different contrastive learning loss strength coefficients λ in Eq. 13 on the overall recommendation performance. We can find that the optimal choices have different distributions on the different metrics for these two datasets: $\lambda = 1/4$ is a bit better on Recall while $\lambda = 1/16$ works better on NDCG. We speculate that $\lambda = 1/4$ emphasizes the value of user repurchase behaviors and thus promotes the ranking of repurchased items. Overestimating the ranking of repurchased items across the entire collection reduces user interest exploration and ultimately leads to a decline in ranking ability.

Variants		Instacart		Dunnhumby	
		Recall	NDCG	Recall	NDCG
-HGC		27.38	26.96	17.14	16.20
-Sub		28.85	29.01	17.23	16.88
-CL		28.73	28.72	17.32	16.72
-DA		28.86	28.84	17.36	16.89
BRL(λ)	λ =1/4	**29.62**	29.29	**17.50**	17.41
	λ =1/16	29.59	29.70	17.48	17.54
	λ =1/64	29.26	29.38	17.37	17.03
BRL		<u>29.59</u>	**29.70**	<u>17.48</u>	**17.54**

Table 3: Ablation study on key components.

5 Related Work

Next basket recommendation has been widely studied in recent years. One approach [Yu *et al.*, 2016] is to use an RNN to obtain basket representations by pooling items, where all the items are treated independently. As a result, the basket features may be suboptimal. Le *et al.* [2019] develop a hierarchical network architecture to capture high-order dependencies across baskets by incorporating pair-wised information across baskets. Schulz *et al.* [2020] construct a model to learn heterogeneous user behaviors, which capture the user's multiple intentions from a single basket. However, these works pay little attention to the item correlations in a basket, and lead to a sub-optimal representation of baskets. Meanwhile, some methods try to exploit repeated items for performance improvements. Hu *et al.* [2020] consider personalized item frequency and neighbor preference to predict the next basket while Hu and He [2019] predict next items by combining an encoder-decoder framework with an attention mechanism based on item frequency. Shen *et al.* [2022] capture meta-knowledge from multi-behavior sequences, from which repeat frequency is considered. Faggioli *et al.* [2020b] utilizes collaborative filtering techniques to model the recency of items and thus to capture users' shopping patterns. The above methods do not well explore the relationship among repeated purchases, while our method considers repeated items in a fine-grained manner. Most recently, although hypergraph convolution has been adopted to enhance traditional recommendation [Yu *et al.*, 2021; Xia *et al.*, 2021], it has not been well studied in NBR. Katz *et al.* [2022], where hypergraph is used to model the items repeatedly purchased periodically. In contrast, we focus on repeated items in a fine-grained manner and provide a more nuanced approach to capturing correlations among items within the same basket.

6 Conclusions and Future Work

This paper advocated the importance of intra-basket item correlations, which were validated by data analysis on a real dataset. Then, a novel model called BRL was realized by constructing hyperedges from all the (repeated) items in a basket. We designed a contrastive learning task to stress more on repeated items. Experimental results demonstrated that our approach outperformed eight baselines. For future work, we intend to consider time intervals between repurchased behaviors to further enhance basket representations.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under Grants No. 62032013 and No. 61972078, and the Fundamental Research Funds for the Central Universities under Grants No. N2217004 and No. N2317002.

Contribution Statement

Yalin Yu*, Enneng Yang* conceived, designed and performed the experiments. Guibing Guo† supervised the project. Linying Jiang and Xingwei Wang analyzed the data. All authors contributed to writing the paper.
Note: Yalin Yu*, Enneng Yang* marked with an asterisk (*) are co-first authors and contributed equally to this work. Guibing Guo† is the corresponding author.

References

Mozhdeh Ariannezhad, Sami Jullien, Ming Li, Min Fang, Sebastian Schelter, and Maarten de Rijke. ReCANet: A Repeat Consumption-Aware Neural Network for Next Basket Recommendation in Grocery Shopping. In *SIGIR '22*, pages 1240–1250, Madrid Spain, July 2022. ACM.

Ting Bai, Jian-Yun Nie, Wayne Xin Zhao, Yutao Zhu, Pan Du, and Ji-Rong Wen. An Attribute-aware Neural Attentive Model for Next Basket Recommendation. In *SIGIR '18*, pages 1201–1204, Ann Arbor MI USA, June 2018. ACM.

Song Bai, Feihu Zhang, and Philip H.S. Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, February 2021.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, December 2014. arXiv:1412.3555 [cs].

Guglielmo Faggioli, Mirko Polato, and Fabio Aiolli. Recency Aware Collaborative Filtering for Next Basket Recommendation. In *UMAP '20*, pages 80–87, Genoa Italy, July 2020. ACM.

Guglielmo Faggioli, Mirko Polato, and Fabio Aiolli. Recency Aware Collaborative Filtering for Next Basket Recommendation. In *UMAP '20*, pages 80–87, Genoa Italy, July 2020. ACM.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph Neural Networks, February 2019. arXiv:1809.09401 [cs, stat].

Lukáš Havrlant and Vladik Kreinovich. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, 46(1):27–36, January 2017.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997.

Haoji Hu and Xiangnan He. Sets2Sets: Learning from Sequential Sets with Neural Networks. In *KDD '19*, pages 1491–1499, Anchorage AK USA, July 2019. ACM.

Haoji Hu, Xiangnan He, Jinyang Gao, and Zhi-Li Zhang. Modeling Personalized Item Frequency Information for Next-basket Recommendation. In *SIGIR '20*, pages 1071–1080, Virtual Event China, July 2020. ACM.

Ori Katz, Oren Barkan, Noam Koenigstein, and Nir Zabari. Learning to Ride a Buy-Cycle: A Hyper-Convolutional Model for Next Basket Repurchase Recommendation. In *RecSys '22*, pages 316–326, Seattle WA USA, September 2022. ACM.

Duc-Trong Le, Hady W. Lauw, and Yuan Fang. Correlation-Sensitive Next-Basket Recommendation. In *IJCAI-19*, pages 2808–2814, Macao, China, August 2019.

Yuqi Qin, Pengfei Wang, and Chenliang Li. The World is Binary: Contrastive Learning for Denoising Next Basket Recommendation. In *SIGIR '21*, pages 859–868, Virtual Event Canada, July 2021. ACM.

Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized Markov chains for next-basket recommendation. In *WWW '10*, page 811, Raleigh, North Carolina, USA, 2010. ACM Press.

Alexander Schulz, Fabian Hinder, and Barbara Hammer. DeepView: Visualizing Classification Boundaries of Deep Neural Networks as Scatter Plots Using Discriminative Dimensionality Reduction. In *IJCAI-PRICAI-20*, pages 2305–2311, Yokohama, Japan, July 2020.

Zhufeng Shao, Shoujin Wang, Qian Zhang, Wenpeng Lu, Zhao Li, and Xueping Peng. A Systematical Evaluation for Next-Basket Recommendation Algorithms, September 2022. arXiv:2209.02892 [cs].

Yanyan Shen, Baoyuan Ou, and Ranzhen Li. MBN: Towards Multi-Behavior Sequence Modeling for Next Basket Recommendation. *TKDD'22*, 16(5):1–23, October 2022.

Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation. *AAAI*, 35(5):4503–4511, May 2021.

Jaehyuk Yi and Jinkyoo Park. Hypergraph Convolutional Recurrent Neural Network. In *KDD '20*, pages 3366–3376, Virtual Event CA USA, August 2020. ACM.

Feng Yu, Qiang Liu, and Shu Wu. A Dynamic Recurrent Model for Next Basket Recommendation. In *SIGIR '16*, pages 729–732, Pisa Italy, July 2016. ACM.

Junliang Yu, Hongzhi Yin, and Jundong Li. Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation. In *WWW '21*, pages 413–424, Ljubljana Slovenia, April 2021. ACM.

Shengyu Zhang, Dong Yao, Zhou Zhao, Tat-Seng Chua, and Fei Wu. CauseRec: Counterfactual User Sequence Synthesis for Sequential Recommendation. In *SIGIR '21*, pages 367–377, Virtual Event Canada, July 2021. ACM.