Basket Booster for Prototype-based Contrastive Learning in Next Basket Recommendation

Ting-Ting Su, Zhen-Yu He, Man-Sheng Chen, and Chang-Dong Wang(✉)

School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
Guangdong Province Key Laboratory of Computational Science, Guangzhou, China
Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
sutt8@mail2.sysu.edu.cn, hezhy65@mail2.sysu.edu.cn,
chenmsh27@mail2.sysu.edu.cn, changdongwang@hotmail.com

Abstract. Next basket recommendation seeks to model the correlation of items and mine users' interests hidden in basket sequences, and tries to infer a set of items that tend to be adopted in the next session with the mined information. However, the feedback provided by users often involves only a small fraction of millions to billions of items. Sparse data makes it hard for model to infer high-quality representations for basket sequences, which further leads to poor recommendation. Inspired by the recent success of representation learning in some fields, e.g., computer vision and clustering, we propose a basket booster for prototype-based contrastive learning (**BPCL**) in next basket recommendation. A correlative basket booster is designed to mine self-supervised signals just from raw data and make augmentation for baskets. To our best knowledge, this is the first work to promote learning of prototype representation through basket augmentation, which helps overcome the difficulties caused by data sparsity and leads to a better next basket recommendation performance. Extensive experiments on three public real-world datasets demonstrate that the proposed BPCL method achieves better performance than the existing state-of-the-art methods.

Keywords: Contrastive Learning · Next Basket Recommendation · Data Augmentation.

1 Introduction

With the rapid growth of the number of entities involved in online platforms, it is difficult for users to find the items that meet their demands. Therefore, recommendation systems are widely used to provide users with more proper items by mining information contained in historical data, e.g., user preferences. There are many previous works trying to model user and item portraits by making use of historical interactions as a set [3, 5, 23, 29], regardless of chronological order. However, chronological order often contributes significantly to recommendation, as the interests of users change over time [10]. To capture evolving interests of users, some attempts have been made in sequential recommendation [1, 4, 10, 12] that tries to learn representation of strictly sequential sequence. Note that user interactions are not strictly sequential over a short period in many cases, e.g., multiple items may be purchased with different intentions in the same shopping session. Next basket recommendation [8, 16, 24, 28] breaks through the bottleneck of sequential recommendation by recommending a set of items simultaneously. Fig. 1 shows an example for next basket recommendation, in which the last basket is expected to be inferred with the preceding basket sequence with length of 2. Note that both of basket and basket sequence have no fixed size, so the size of the predicted next basket is predefined in next basket recommendation.

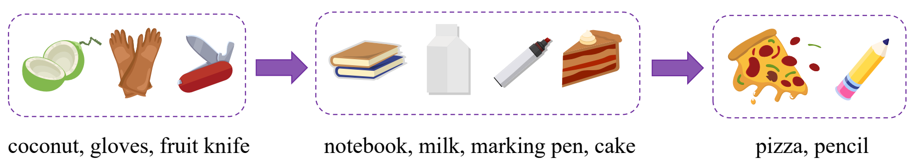

Fig. 1. An example for next basket recommendation.

Data sparsity is a severe challenge for recommendation tasks. The feedback provided by users often involves only a small fraction of millions to billions of items, which makes it hard for model to infer high-quality representations for basket sequences. In the recent years, contrastive learning has been shown to perform well in representation learning in the data-sparse tasks, such as computer vision [14, 17, 18] and clustering [13, 20]. The main idea lies in capturing self-supervised signals from raw data with various data augmentations. In computer vision, it is easy to augment data by rotation, changing color or adding noise, but these methods are unsuitable for recommendation tasks because of the different data types. In order to integrate the advantages of contrastive learning into recommendation tasks, some attempts have been made in developing augmentation methods for item sequence in sequential recommendation [22, 25]. These methods mainly rely on executing random reordering, clipping or insertion of items, which are not applicable for basket sequence since it would destroy the correlation of items within a basket. The augmentation for basket sequences in next basket recommendation remains an unaddressed issue.

To address the issues mentioned above, we introduce contrastive learning to next basket recommendation by proposing a basket booster for prototype-based contrastive learning (**BPCL**). Basket sequences are modeled via an item correlation graph for constructing correlation matrix. With the correlation matrix, we develop a correlation basket booster for basket augmentation, which maintains the correlation between intra-basket items while introducing randomness. The augmented basket sequences as well as the corresponding prototype basket sequences will be encoded to basket sequence representations for prototype-based contrastive learning to improve the performance of next basket recommendation.

To summarize, our contributions are listed as follows:

- We propose a **BPCL** method that introduces contrastive learning to next basket recommendation to mine self-supervised signals from primitive basket sequences, i.e., prototypes. To our best knowledge, this is the first work to promote learning of prototype representation through basket augmentation.
- We propose a basket booster for prototype-based contrastive learning to maintain the correlation between intra-basket items while introducing randomness. The effectiveness of the booster is demonstrated by ablation study.
- Comprehensive experiments on three public real-world datasets demonstrate that the proposed BPCL method achieves better basket recommendation performance than the state-of-the-art methods in terms of the four metrics.

2 Related Work

In this section, we will briefly review previous works related to our work, namely next basket recommendation and contrastive learning.

2.1 Next Basket Recommendation

Temporal recommendation focusing on modeling interactions with timeline has shown competitive performance in many time-sensitive scenarios [2]. According to the target of recommendation, temporal recommendation can be divided to sequential recommendation and basket recommendation. Specially, next basket recommendation is a type of basket recommendation that predicts next basket without any information from next basket. Sequential recommendation [7, 27, 30, 32] seeks to predict next item with the representation of item sequence, while next basket recommendation [16, 21, 24, 26] tries to predict next basket with the representation of basket sequence. In many real-world scenarios, interactions do not in strict chronological order over a short period, e.g., a shopping session. Hence, next basket recommendation that fits the situation has gained increasing attention in recent years.

There are some previous works [8, 16, 31] focusing on modeling qualified representation of basket sequence to guide the prediction of next basket. Hu et al. [8] design a KNN-based method to model basket sequence representation with the information from similar users' interactions. Le et al. [16] model basket sequence representation by developing a hierarchical network considering both inter-basket association and intra-basket association. Recently, advanced representation learning methods such as graph embedding [10, 19], attention mechanisms [7, 12] are widely applied in sequential recommendation for their outstanding performance in learning representation. However, the performance of these methods depends on data in a high degree, while the feedback provided by users often involves only a small fraction of millions to billions of items leading to data sparsity. It is necessary to develop a method to alleviate the issues caused by sparse data in next basket recommendation.

2.2 Contrastive Learning

Contrastive learning aims to mine useful signals from unlabeled data to alleviate sparse data problems. This method has been widely used in some fields, e.g., computer vision [14, 17, 18] and clustering [13, 20]. For computer vision, contrastive learning can be adopted to promote performance of domain adaptation [11]. For clustering, contrastive learning can help with learning representation of different clusters [20]. When it comes to basket recommendation, only a few attempts has been made in leveraging contrastive learning to alleviate the problem of data sparsity. To our best knowledge, the only attempt for integrating contrastive learning to next basket recommendation is [24]. It tries to split a target basket sequence into two sub-basket sequences according to the correlation between the item in a basket and the item in candidate set, and aims to learn a qualified representation for the filtered pos-basket sequence. However, in next basket recommendation, augmentation method for learning better representation of prototype is less well-studied.

3 The Proposed Method

In this section, we present the proposed BPCL method for next basket recommendation. We start with formulating next basket recommendation problem in Section 3.1. Then, we introduce the proposed BPCL method in detail in Section 3.2 with six parts, i.e., correlative basket booster, basket encoder, dynamic context encoder, basket predictor, prototype-based contrastive learning and multi-task training. The overall architecture of BPCL is illustrated in Fig. 2.

3.1 Problem Statement

Let $\mathcal{I} = \{i_1, \ldots, i_N\}$ denote the set of items and $\mathcal{B}_t = \left\{i_1, \ldots, i_{|\mathcal{B}_t|}\right\}$ denote the basket at time step t, where N is the number of items and $\mathcal{B}_t \subset \mathcal{I}$. Given a basket sequence $S = [\mathcal{B}_1, \ldots, \mathcal{B}_{|S|}]$ consisting of several baskets, where $|S|$ is the length of S, next basket recommendation aims to predict several items to be adopted at time step $(|S| + 1)$ as:

$$\hat{\mathcal{B}}_{|S|+1} = \mathcal{K}_{i\in\mathcal{I}}(\mathrm{P}(i \in \mathcal{B}_{|S|+1}|S)) \tag{1}$$

where $\mathcal{K}$ denotes k items picked with the highest probability. Note that k is a predefined number indicating the size of the next basket, and next basket recommendation predicts all items in the next basket simultaneously.

3.2 BPCL

Correlative Basket Booster Given a basket $\mathcal{B}_t$, it will be converted to a binary vector as $\mathbf{b}_t \in \{0, 1\}^N$ in which the n-th entry is set to 1 if $i_n \in \mathcal{B}_t$. For basket augmentation, one of the most direct ways is random masking, i.e., randomly set some 1 to 0 in the vector. We argue that using random masking alone

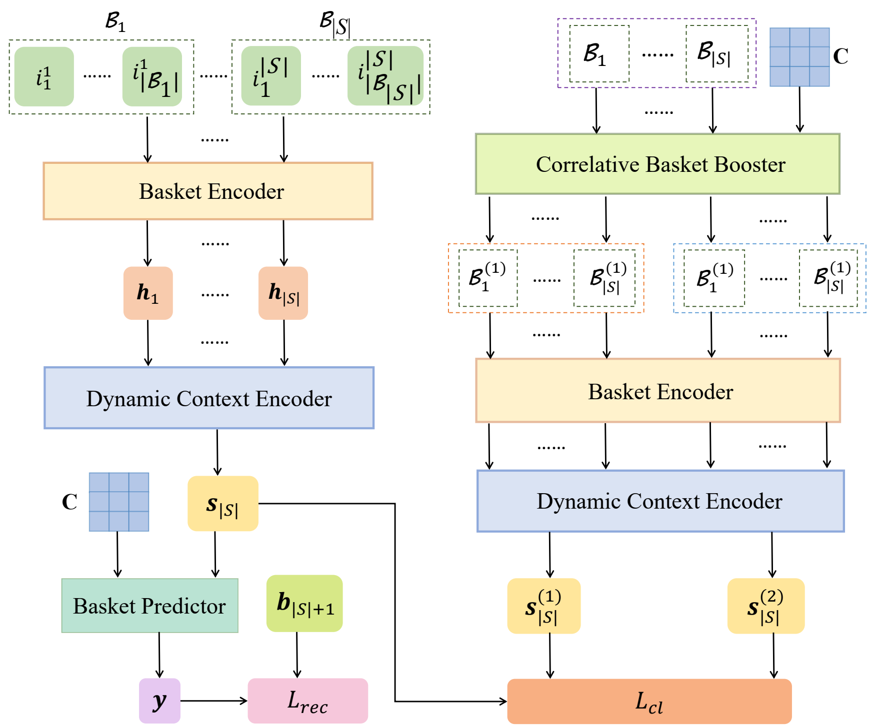

Fig. 2. The architecture of the proposed BPCL method.

can destroy the correlation between items within the same basket. Instead of augmenting basket by random masking, we propose a correlative basket booster to maintain the correlation and introduce randomness for data augmentation.

When modeling correlation between items by a graph, each item can be regarded as a node in the graph. For item p and q, the weight of their connection edge can be set to the number of times that they are in the same basket. And then we can obtain a weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ according to the graph. Further, the adjacency matrix can be normalized to correlation matrix $\mathbf{C} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$, with each $\mathbf{C}_{pq}$ defined as:

$$\mathbf{C}_{pq} = \begin{cases} 0, & p = q; \\ \frac{\mathbf{A}_{pq}}{\sqrt{\sum_p \mathbf{A}_{pq}}\sqrt{\sum_q \mathbf{A}_{pq}}}, & p \neq q. \end{cases} \tag{2}$$

With the correlation matrix, we can distinguish the items associated with a basket in a higher degree for basket augmentation as:

$$\mathbf{b} = \max\{0, \mathbf{b} \cdot \mathbf{C}\}. \tag{3}$$

However, it is unreasonable for that $\mathbf{b}$ and $\mathbf{C}$ are in different representation spaces. And we should map them to the same latent space before matching. Finally, the formulation of the proposed correlative basket booster is:

$$\mathbf{b}^{aug} = \mathbf{b}^{'} \circ \mathcal{B}(\mathbf{m}) \tag{4}$$

$$\mathbf{b}^{'} = \max(\mathbf{0}, (\mathbf{b}\boldsymbol{\Pi} + \boldsymbol{\delta}) \cdot (\mathbf{C}\boldsymbol{\Upsilon} + \boldsymbol{\xi})^T) \tag{5}$$

where $\circ$ means element-wise product, $\mathbf{m} \in \mathbb{R}^N$ conforms to the uniform distribution, i.e., $m_n \sim U(0,1)$, $\boldsymbol{\Pi} \in \mathbb{R}^{N\times L}$, $\boldsymbol{\delta} \in \mathbb{R}^L$, $\boldsymbol{\Upsilon} \in \mathbb{R}^{N\times L}$ and $\boldsymbol{\xi} \in \mathbb{R}^{N\times L}$ are learnable parameters, L is the correlation latent dimension, and $\mathcal{B}$ is defined as:

$$\mathcal{B}(\mathbf{x}_p) = \begin{cases} 0, & \text{if } \mathbf{x}_p \leq \mu; \\ 1, & \text{if } \mathbf{x}_p > \mu \end{cases} \tag{6}$$

where $\mu \in \mathbb{R}^+$ is a hyper-parameter of mask threshold. The correlation matrix contributes to maintain the correlation between items within the same basket, while masking introduces some randomness.

The correlative basket booster will work for the data augmentation in contrastive learning task.

Basket Encoder Although the vectorized basket representation $\mathbf{b}_t$ of $\mathcal{B}_t$ includes the direct relationships between the basket and all items, the information contained in the high-dimensional vector is far less than the maximum information it can store. In order to reduce data redundancy, we compress it to a low-dimensional space and obtain the hidden representation $\mathbf{h}_t$ of basket $\mathcal{B}_t$ as:

$$\mathbf{h}_t = \mathcal{F}(\mathbf{b}_t \cdot \boldsymbol{\Phi} + \boldsymbol{\theta}) \tag{7}$$

where $\mathcal{F}$ is an activation function, e.g., Relu, $\boldsymbol{\Phi} \in \mathbb{R}^{N\times H}$, $\boldsymbol{\theta} \in \mathbb{R}^H$ are learnable parameters, and H is the basket latent dimension.

Dynamic Context Encoder Since basket sequence consists of a string of chronological sets of interactions, it is convenient to mine dynamic user interest hidden in the sequence to guide next basket recommendation. We utilize a Recurrent Neural Network (RNN) [1, 6, 15] to model basket sequence as:

$$\mathbf{s}_t = \mathcal{G}(\mathbf{h}_t\boldsymbol{\Psi} + \mathbf{s}_{t-1}\boldsymbol{\Omega} + \boldsymbol{\varphi}) \tag{8}$$

where $\mathbf{s}_t$ is the sequence hidden representation at time step t, $\mathcal{G}$ is an activation function, $\boldsymbol{\Psi} \in \mathbb{R}^{H\times K}$, $\boldsymbol{\Omega} \in \mathbb{R}^{K\times K}$ and $\boldsymbol{\varphi} \in \mathbb{R}^K$ are learnable parameters, K is the sequence latent dimension. The sequence hidden state at time step $|S|$, i.e., $\mathbf{s}_{|S|}$, encodes the interaction context of basket sequence S.

Basket Predictor In intuition, we can get the probability that each item belongs to the next basket by linearly mapping the sequence representation $\mathbf{s}_{|S|}$ to probability space. However, the method ignores the importance of correlation between items within the same basket for next basket recommendation. So we utilize the correlation matrix $\mathbf{C}$ defined in Eq. (2) to provide information about the correlation between items:

$$\mathbf{y} = \mathbf{s}_{|S|} \cdot \frac{1}{1 + e^{-\mathbf{\Lambda C}}} \tag{9}$$

where $\mathbf{\Lambda} \in \mathbb{R}^{K \times N}$ is a learnable weight matrix.

We choose the recommendation loss $\mathcal{L}_{rec}$ with the expectation that items belonging to the next basket are assigned with a higher probability, and encourage them to maintain the greatest possible advantage over items not in the next basket. Then the objective function is formulated as:

$$\mathcal{L}_{rec} = -\frac{N - |\mathcal{B}_{|S|+1}|}{|\mathcal{B}_{|S|+1}|} \sum_{p \in \mathcal{B}_{|S|+1}} \log(\frac{1}{1 + e^{-\mathbf{y}_p}}) - \sum_{q \notin \mathcal{B}_{|S|+1}} \log(\frac{e^{r - \mathbf{y}_q}}{1 + e^{r - \mathbf{y}_q}}) \tag{10}$$

where r is the minimum probability of items belonging to the next basket.

To help the recommendation task encode basket sequence, and learn a better representation for final prediction, we try to integrate contrastive learning into recommendation task in the next section.

Prototype-based Contrastive Learning With the correlative basket booster defined by Eq. (4), two augmentations can be obtained for a basket sequence $S = [\mathcal{B}_1, \dots, \mathcal{B}_{|S|}]$ by executing augmentation twice, namely $S^{(1)} = [\mathcal{B}_1^{(1)}, \dots, \mathcal{B}_{|S|}^{(1)}]$ and $S^{(2)} = [\mathcal{B}_1^{(2)}, \dots, \mathcal{B}_{|S|}^{(2)}]$. There are $(2 \times M)$ augmentations corresponding to a batch of basket sequences with batch size of M. To construct contrastive signals, the two augmentations from the same basket sequence will be treated as a positive pair, and the remaining $2(M-1)$ augmentations from different basket sequences will be treated as negative pairs.

All the augmentations in the batch can be encoded by the context encoder defined in Eq. (8), denoted as $\mathcal{S}_{aug} = \left\{\mathbf{s}_1^{(1)}, \dots, \mathbf{s}_M^{(1)}, \mathbf{s}_1^{(2)}, \dots, \mathbf{s}_M^{(2)}\right\}$. The contrastive loss $\mathcal{L}_{cl}$ is designed to maximize the similarity of representations from the same basket, and minimize the similarity of representations from different baskets. To achieve the goal, we can define contrastive loss for the augmented basket sequence $S_p^{(1)}$ in the batch as:

$$\mathcal{L}_{cl}(\mathbf{s}_p^{(1)}, \mathbf{s}_p^{(2)}) = -\log \frac{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p^{(2)}}^T)}{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p^{(2)}}^T) + \sum_{\mathbf{s}^- \in \mathcal{S}_{aug}^-} \exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}^-}^T)} \tag{11}$$

where $\mathcal{S}_{aug}^-$ is the set of sequence representations of augmentations from different basket sequences with $S_p^{(1)}$ in the batch.

Note that introducing contrastive learning into the recommendation task would help the recommendation model learn the representation of basket sequence that is more conducive to predict recommendation probability. However, augmented basket sequences inevitably lose some information of their primitive basket sequences. In order to preserve as much information as possible of prototype, we further define a prototype-based contrastive learning loss as:

$$\begin{aligned}\mathcal{L}_{cl}(\mathbf{s}_p, \mathbf{s}_p^{(1)}, \mathbf{s}_p^{(2)}) = &-\log \frac{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p}^T)}{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p}^T)) + \sum_{\mathbf{s}^- \in \mathcal{S}^-} \exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}^-}^T)} \\ &-\log \frac{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p^{(2)}}^T)}{\exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}_p^{(2)}}^T) + \sum_{\mathbf{s}^- \in \mathcal{S}_{aug}^-} \exp(\mathbf{s}_p^{(1)} \cdot {\mathbf{s}^-}^T)}\end{aligned} \tag{12}$$

where $\mathcal{S}^-$ is the set of sequence representations of prototypes and $\mathcal{S}_{aug}^{(1)} = \left\{\mathbf{s}_1^{(1)}, \ldots, \mathbf{s}_M^{(1)}\right\}$ in the batch except $\mathbf{s}_p$ and $\mathbf{s}_p^{(1)}$.

Multi-Task Training In the previous sections, we introduced the correlative basket booster for augmentation, the part of model for recommendation task, and the part of model for contrastive learning task, respectively. In this section, we adopt a multi-task strategy to combine them by optimizing them jointly as:

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

where $\lambda \in \mathbb{R}$ is a hyper-parameter to adjust the intensity of contrastive learning.

4 Experiments

In this section, we design extensive experiments to evaluate the performance of the proposed BPCL method against six baseline methods on three real-world datasets. In particular, we aim to answer the following three research questions:

- **RQ1**: How does the proposed BPCL method perform on next basket recommendation compared with existing methods?
- **RQ2**: Whether the prototype-based contrastive learning with correlative basket booster promotes the model to recommend the next basket?
- **RQ3**: How do the key hyper-parameters, i.e., the correlation latent dimension L, the weight of contrastive learning λ and the mask threshold μ, affect model quality?

4.1 Experiments Settings

Datasets We conduct experiments on three public real-world datasets: Delicious[1], Beauty[2] and TaFeng[3]. Delicious contains tagging information from Nov

[1] http://www.delicious.com
[2] https://www.kaggle.com/skillsmuggler/amazon-ratings
[3] https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset

Table 1. Statistics for the Delicious, Beauty and TaFeng datasets

Dataset	Num of items	Average basket size	Average basket sequence length
Delicious	8920	3.78	31.66
Beauty	19340	1.50	6.33
TaFeng	14313	5.76	4.90

2003 to Nov 2010 of a social bookmarking system, in which the tags assigned to the same bookmark is regarded as a basket. Beauty consists of interactions from May 1996 to Jul 2014 of subcategory "Beauty" on Amazon, and TaFeng contains the transaction data of a Chinese grocery store from Nov 2000 to Feb 2001. We define the set of items that are interacted with the same user within the same day as a basket for the two datasets. Each basket sequence consists of the baskets from the same user in chronological order. The statistics of the three datasets are described in Table 1.

Following [16], we divide the basket sequences into three non-overlapping time periods as training set, validation set and testing set. The items and users with less than 5 interactions, as well as the basket sequences of less than 3 in length are filtered out. For Delicious, the part of interactions before Sep 2010 is treated as training set, the part from Sep to Oct of 2010 is validation set and the part after Oct 2010 is testing set. For Beauty, the part of interactions before Jun 2013 is treated as training set, the part after Jul of 2013 is validation set and the part of 2014 is testing set. For TaFeng, the part of interactions of 2000 is treated as training set, the part before Jan of 2001 is validation set and the part after Feb 2001 is testing set.

Evaluation Given a basket sequence $S = [\mathcal{B}_1, \ldots, \mathcal{B}_{|S|}]$, the preceding $(|S| - 1)$ baskets are used to predict the last basket, and any sequence with more than 30 baskets will be truncated with the prefix cut off. To evaluate the performance of the proposed method and baselines, we adopt the widely used Hit Ratio ($HR@K$) [4, 9, 22] and F-measure ($F1@K$) [15, 24] as evaluation metrics with $K = 5$ and $K = 20$, which can be formulated as:

$$HR@K = \frac{\sum Hit(\mathcal{B}_{pred}^{K})}{\sum_{\mathcal{B}_{target} \in \mathcal{T}} |\mathcal{B}_{target}|} \tag{14}$$

$$F1@K = \frac{2 \times Recall@K \times Precision@K}{Recall@K + Precision@K} \tag{15}$$

where $\mathcal{B}_{target}$ and $\mathcal{B}_{pred}^{K}$ denote a target basket and the corresponding predicted basket with the size of K respectively, Hit counts the number of items that appear in both $\mathcal{B}_{pred}^{K}$ and $\mathcal{B}_{target}$, and $\mathcal{T}$ is the set of the last basket in the testing set. The $Recall@K$ and $Precision@K$ are defined as:

$$Recall@K = \frac{1}{|\mathcal{T}|} \sum_{\mathcal{B}_{target} \in \mathcal{T}} \frac{Hit(\mathcal{B}^K_{pred})}{|\mathcal{B}_{target}|} \tag{16}$$

$$Precision@K = \frac{1}{|\mathcal{T}|} \sum_{\mathcal{B}_{target} \in \mathcal{T}} \frac{Hit(\mathcal{B}^K_{pred})}{|\mathcal{B}^K_{pred}|}. \tag{17}$$

Baselines We adopt the following six recommendation models for comparison.

- **POP-K** It is a non-personalized recommendation model that recommends K items with the highest popularity in terms of basket for users.
- **MCNet** It learns transition probability between the latest basket and candidate items based on the Markov-chain implemented by a neural network.
- **BSEQ**[15] It recommends next basket based on the corresponding basket sequence representation learned by making use of recurrent neural network.
- **SASRec**[12] It makes an adaption for transformer layer to learn the correlation of items in sequences, and recommends items based on the correlation.
- **CoSeRec**[22] It introduces two informative augmentations for item sequence to construct self-supervised signals, and applies transformer encoder to promote recommendation performance.
- **Beacon**[16] It is a state-of-the-art model for next basket recommendation that utilizes correlation between items to encode basket sequences and conduct next basket recommendation.

Implementation Details We choose LSTM with 0.3 dropout probability as the type of recurrent layer units, and set the latent dimension L, H and K as 32. The RMSProp optimizer with a learning rate of 0.01 is adopted for optimizing the model. For hyper-parameters, we tune mask threshold μ within the range of $\{0.5, 0.6, 0.7, 0.8, 0.9\}$, and contrastive learning weight λ within the range of $\{0.05, 0.1, 0.2, 0.3, 0.4\}$. To be fair, the hyper-parameters of baseline methods share the same experimental settings as mentioned above. As for the others, we tune them on the validation set applying early stopping with patience of 5 epochs, and report results on the testing set.

4.2 Performance Comparison

Table 2 presents the experimental results of the all methods on the three real-world datasets. We can observe that the proposed BPCL method obtains the best performance on the three datasets in terms of the four metrics. For all datasets, our model achieves at least 7% improvement in every metric, which helps answer **RQ1**. In particular, it achieves at least 44% improvement on Beauty. We can find that Beauty contains the largest number of items and the least average interactions among the three datasets from Table 1, so the experimental results suggest

Table 2. Performance comparison of different methods on next basket recommendation. The best scores on all datasets are highlighted in bold and the second best is labeled with *. The last column is the improvement compared with the best baseline results.

Model	Metrics		POP	BSEQ	MCNet	SASRec	CoSeRec	Beacon	BPCL	**Improve**
Delicious	HR@K	5	1.84	3.28	3.66*	2.89	0.16	3.09	**4.00**	9.29%
	(%)	20	7.73	10.5*	10.4	8.86	0.59	9.25	**11.4**	8.57%
	F1@K	5	1.83	2.24	2.57*	2.19	0.11	2.22	**2.75**	7.00%
	($\times 10^2$)	20	2.24	2.93*	2.91	2.50	0.17	2.61	**3.16**	7.85%
Beauty	HR@K	5	0.04	0.00	0.23	0.11	0.15	0.41*	**0.68**	65.9%
	(%)	20	0.83	0.19	0.26	0.34	0.38	1.02*	**1.47**	44.1%
	F1@K	5	0.02	0.00	0.09	0.04	0.06	0.16*	**0.28**	75.0%
	($\times 10^2$)	20	0.12	0.03	0.03	0.05	0.05	0.13*	**0.19**	46.2%
Tafeng	HR@K	5	4.65	4.66	4.73	4.32	3.67	5.11*	**5.52**	8.02%
	(%)	20	6.03	6.41	6.65	6.24	5.98	7.29*	**7.88**	8.09%
	F1@K	5	4.79	4.80	4.84	4.40	3.72	5.20*	**5.63**	8.27%
	($\times 10^2$)	20	2.75*	2.45	2.52	2.37	2.27	2.75*	**2.98**	8.36%

that BPCL can effectively overcome the difficulties caused by data sparsity in next basket recommendation.

Note that CoSeRec is a sequential recommendation model that introduces informative augmentation for contrastive learning, but achieves poor performance comparing to the others on Delicious which contains the most average interactions among the three datasets. This is perhaps due to its intent to predict the next item by modeling item sequence with the assumption that interactions are in strict chronological order which is at odds with the reality, and improper augmentation for item sequences destruct the intra-basket correlation. BSEQ performs well on Delicious and Tafeng, but fails on Beauty, a dataset with large data sparsity. This indicates that although next basket recommendation can improve the recommendation effectiveness by fitting actual situation, it is prone to be affected by data density and is in great instability. All of the above suggest that the proposed correlative basket booster could promote the model to recommend the next basket. In order to further verify this conjecture, ablation study is conducted.

4.3 Ablation Study

The performance of BPCL and its variants on all three datasets are shown in Table 3. BPCL-CL indicates BPCL without contrastive learning, and BPCL-prototype indicates BPCL adopting contrastive loss without the introduction of prototype, i.e., Eq. (11). It is clear that the performances of BPCL-CL and BPCL-prototype are no better than that of BPCL in terms of the four metrics on all datasets. More specially, sparse data magnifies the advantage, i.e., BPCL achieves the greatest improvement of performance on Beauty. Although BPCL-prototype attains better performance than BPCL-CL because of the integration of contrastive signal, it is still hard to surpass BPCL due to the inevitable loss

Table 3. Performance of the proposed BPCL and its variants for ablation studies.

Dataset	**Model**	HR@K (%)		F1@K ($\times 10^2$)	
		5	20	5	20
Delicious	BPCL-CL	3.732	10.02	2.608	2.810
	BPCL-prototype	3.237	10.93	2.202	3.036
	BPCLR	**3.998**	**11.35**	**2.751**	**3.159**
Beauty	BPCL-CL	0.000	0.000	0.000	0.000
	BPCL-prototype	0.451	1.278	0.192	0.170
	BPCL	**0.677**	**1.466**	**0.280**	**0.194**
Tafeng	BPCL-CL	2.775	4.323	2.785	1.620
	BPCL-prototype	5.517	7.676	5.626	2.886
	BPCL	**5.517**	**7.879**	**5.631**	**2.976**

of prototype information. We can conclude that the proposed prototype-based contrastive learning with correlative basket booster promotes the model to recommend the next basket (**RQ2**).

4.4 Hyper-Parameter Study

In this section, we analyze three key hyper-parameters of the proposed BPCL method to answer **RQ3**, including the correlation latent dimension L, the weight of contrastive learning λ and the mask threshold μ.

Firstly, we explore how the correlation latent dimension L influences the performance of BPCL and show the results in Fig. 3. It is obvious that BPCL holds steady performance on the three datasets with different L and keeps a slow lift on the whole as the increase of L until it reaches 32. The similar performance of BPCL with $L = 32$ and BPCL with $L = 8$ implies that a small correlation latent dimension is sufficient for embedding the correlation information, which is consistent with the fact of data sparsity. The performance tends to decrease when L is larger than 32, which indicates that too large latent dimension makes it difficult for the model to capture the most valuable information for augmentation.

Next, we investigate how the different combinations of the weight of contrastive learning λ and the mask threshold μ in correlative basket booster affect recommendation performance. The λ is tuned among $\{0.05, 0.1, 0.2, 0.3, 0.4\}$, while the μ is tuned among $\{0.5, 0.6, 0.7, 0.8, 0.9\}$. We adopt heatmaps to show the results of the proposed BPCL method with different combinations on Beauty visually, and the results in terms of HR are shown in upper Fig. 4 and the results in terms of F1 are shown in lower Fig. 4. In general, better performance is achieved by λ larger than 0.2 with $\mu = 0.5$, suggesting that proper combination of the weight of contrastive learning and randomness helps the model capture useful self-supervised signal from prototype to learn better basket sequence representation indeed, thus can contribute to overcome the difficulties caused by sparsity data and get a better next basket recommendation performance.

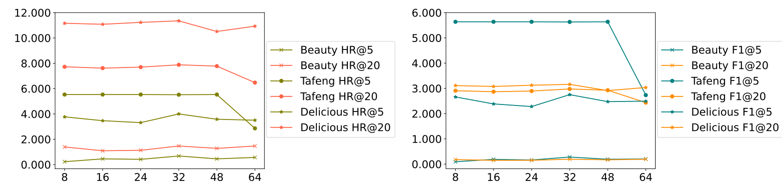

Fig. 3. The performance of BPCL with varying correlation latent dimension L. The left corresponds to $HR@K$ (%) and the right corresponds to $F1@K$ ($\times 10^2$). The x-axis denotes L varies within the range of $\{8, 16, 24, 32, 48, 64\}$, and the y-axis denotes $HR@K$ (%) and $F1@K$ ($\times 10^2$) respectively.

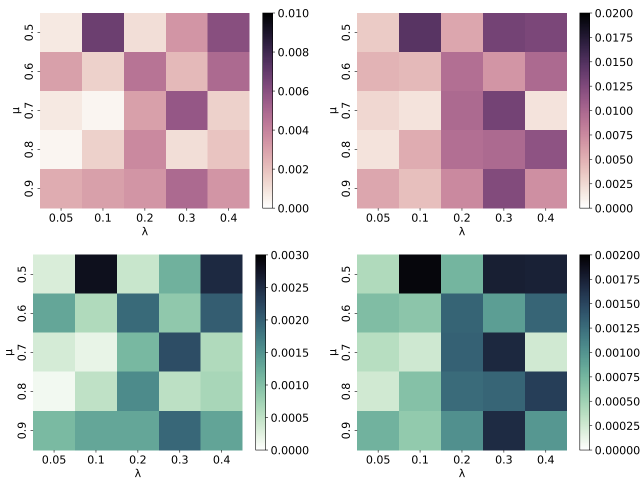

Fig. 4. Heatmap of Hit Ratio (HR) and F1 on Beauty. The upper left corresponds to $HR@5$ (%), the upper right corresponds to $HR@20$ (%), the lower left corresponds to $F1@5$ ($\times 10^2$), the lower right corresponds to $F1@20$ ($\times 10^2$). The x-axis denotes λ varies within the range of $\{0.05, 0.1, 0.2, 0.3, 0.4\}$, and the y-axis denotes μ varies within the range of $\{0.5, 0.6, 0.7, 0.8, 0.9\}$. The darker color represents the better performance.

5 Conclusion

In this paper, we propose a new BPCL method that introduces contrastive learning to next basket recommendation. A correlative basket booster is designed to make augmentation for baskets, which can mine self-supervised signals from primitive basket sequences. The augmentations are utilized by prototype-based contrastive learning for promoting next basket recommendation task. To our best knowledge, this is the first work to promote learning of prototype representation through basket augmentation, which helps overcome the difficulties caused by data sparsity and leads to a better next basket recommendation performance. The proposed method is verified on three public real-world datasets, and show the best performance compared with the baseline methods.

Acknowledgments

This work was supported by NSFC (61876193) and Guangdong Province Key Laboratory of Computational Science at Sun Yat-sen University (2020B1212060032).

References

1. Cui, Q., Wu, S., Liu, Q., Zhong, W., Wang, L.: MV-RNN: A multi-view recurrent neural network for sequential recommendation. IEEE Trans. Knowl. Data Eng. **32**(2), 317–331 (2020)
2. Dai, S., Yu, Y., Fan, H., Dong, J.: Spatio-temporal representation learning with social tie for personalized POI recommendation. Data Sci. Eng. **7**(1), 44–56 (2022)
3. Deng, Z., Huang, L., Wang, C., Lai, J., Yu, P.S.: Deepcf: A unified framework of representation learning and matching function learning in recommender system. In: The Thirty-Third AAAI Conference on Artificial Intelligence, Honolulu. pp. 61–68. AAAI Press (2019)
4. Du, Y., Liu, H., Wu, Z.: Modeling multi-factor and multi-faceted preferences over sequential networks for next item recommendation. In: Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference, Bilbao. vol. 12976, pp. 516–531. Springer (2021)
5. Flanagan, A., Oyomno, W., Grigorievskiy, A., Tan, K.E., Khan, S.A., Ammad-ud-din, M.: Federated multi-view matrix factorization for personalized recommendations. In: Machine Learning and Knowledge Discovery in Databases - European Conference, Ghent. vol. 12458, pp. 324–347. Springer (2020)
6. Gama, R., Fernandes, H.L.: An attentive RNN model for session-based and context-aware recommendations: a solution to the recsys challenge 2019. In: Proceedings of the Workshop on ACM Recommender Systems Challenge, Copenhagen. pp. 6:1–6:5. ACM (2019)
7. He, Z., Zhao, H., Lin, Z., Wang, Z., Kale, A., McAuley, J.J.: Locker: Locally constrained self-attentive sequential recommendation. In: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event. pp. 3088–3092. ACM (2021)
8. Hu, H., He, X., Gao, J., Zhang, Z.: Modeling personalized item frequency information for next-basket recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, Virtual Event. pp. 1071–1080. ACM (2020)
9. Islek, I., Ögüdücü, S.G.: A hybrid recommendation system based on bidirectional encoder representations. In: ECML PKDD 2020 Workshops - Workshops of the European Conference on Machine Learning and Knowledge Discovery in Databases, Ghent. vol. 1323, pp. 225–236. Springer (2020)
10. Ji, Y., Yin, M., Fang, Y., Yang, H., Wang, X., Jia, T., Shi, C.: Temporal heterogeneous interaction graph embedding for next-item recommendation. In: Machine Learning and Knowledge Discovery in Databases - European Conference, Ghent. vol. 12459, pp. 314–329. Springer (2020)
11. Kang, G., Jiang, L., Wei, Y., Yang, Y., Hauptmann, A.: Contrastive adaptation network for single- and multi-source domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell. **44**(4), 1793–1804 (2022)
12. Kang, W., McAuley, J.J.: Self-attentive sequential recommendation. In: IEEE International Conference on Data Mining, Singapore. pp. 197–206. IEEE Computer Society (2018)

13. Ke, G., Hong, Z., Zeng, Z., Liu, Z., Sun, Y., Xie, Y.: CONAN: contrastive fusion networks for multi-view clustering. In: 2021 IEEE International Conference on Big Data (Big Data), Orlando. pp. 653–660. IEEE (2021)
14. Kim, S., Jeong, S., Kim, E., Kang, I., Kwak, N.: Self-supervised pre-training and contrastive representation learning for multiple-choice video QA. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, 2021. pp. 13171–13179. AAAI Press (2021)
15. Le, D., Lauw, H.W., Fang, Y.: Modeling contemporaneous basket sequences with twin networks for next-item recommendation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholmn. pp. 3414–3420. ijcai.org (2018)
16. Le, D., Lauw, H.W., Fang, Y.: Correlation-sensitive next-basket recommendation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao. pp. 2808–2814. ijcai.org (2019)
17. Lee, S., Lee, Y., Lee, G., Hwang, S.: Supervised contrastive embedding for medical image segmentation. IEEE Access **9**, 138403–138414 (2021)
18. Lee, T., Yoo, S.: Augmenting few-shot learning with supervised contrastive learning. IEEE Access **9**, 61466–61474 (2021)
19. Li, C., Hsu, C., Zhang, Y.: Fairsr: Fairness-aware sequential recommendation through multi-task learning with preference graph embeddings. ACM Trans. Intell. Syst. Technol. **13**(1), 16:1–16:21 (2022)
20. Li, Y., Hu, P., Liu, J.Z., Peng, D., Zhou, J.T., Peng, X.: Contrastive clustering. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, Virtual Event. pp. 8547–8555. AAAI Press (2021)
21. Liu, T., Yin, X., Ni, W.: Next basket recommendation model based on attribute-aware multi-level attention. IEEE Access **8**, 153872–153880 (2020)
22. Liu, Z., Chen, Y., Li, J., Yu, P.S., McAuley, J.J., Xiong, C.: Contrastive self-supervised sequential recommendation with robust augmentation. CoRR **abs/2108.06479** (2021)
23. Lu, Y., Xie, R., Shi, C., Fang, Y., Wang, W., Zhang, X., Lin, L.: Social influence attentive neural network for friend-enhanced recommendation. In: Machine Learning and Knowledge Discovery in Databases: Applied Data Science Track - European Conference, Ghent. vol. 12460, pp. 3–18. Springer (2020)
24. Qin, Y., Wang, P., Li, C.: The world is binary: Contrastive learning for denoising next basket recommendation. In: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event. pp. 859–868. ACM (2021)
25. Qiu, R., Huang, Z., Yin, H., Wang, Z.: Contrastive learning for representation degeneration problem in sequential recommendation. In: WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, 2022. pp. 813–823. ACM (2022)
26. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized markov chains for next-basket recommendation. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh. pp. 811–820. ACM (2010)
27. Tong, X., Wang, P., Li, C., Xia, L., Niu, S.: Pattern-enhanced contrastive policy learning network for sequential recommendation. In: Zhou, Z. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Virtual Event / Montreal. pp. 1593–1599. ijcai.org (2021)
28. Wan, S., Lan, Y., Wang, P., Guo, J., Xu, J., Cheng, X.: Next basket recommendation with neural networks. In: Poster Proceedings of the 9th ACM Conference on

Recommender Systems, Vienna. CEUR Workshop Proceedings, vol. 1441. CEUR-WS.org (2015)
29. Xi, W., Huang, L., Wang, C., Zheng, Y., Lai, J.: BPAM: recommendation based on BP neural network with attention mechanism. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao. pp. 3905–3911. ijcai.org (2019)
30. Xie, Z., Liu, C., Zhang, Y., Lu, H., Wang, D., Ding, Y.: Adversarial and contrastive variational autoencoder for sequential recommendation. In: The Web Conference 2021, Virtual Event / Ljubljana. pp. 449–459. ACM / IW3C2 (2021)
31. Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T.: A dynamic recurrent model for next basket recommendation. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, Pisa. pp. 729–732. ACM (2016)
32. Yuan, X., Chen, H., Song, Y., Zhao, X., Ding, Z.: Improving sequential recommendation consistency with self-supervised imitation. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Virtual Event / Montreal. pp. 3321–3327. ijcai.org (2021)