Correlation-Aware Next Basket Recommendation Using Graph Attention Networks

Yuanzhe Zhang[1], Ling Luo[2](✉), Jianjia Zhang[3], Qiang Lu[1], Yang Wang[3], and Zhiyong Wang[1]

[1] The University of Sydney, Sydney, NSW 2006, Australia
yzha9691@uni.sydney.edu.au, {steven.lu,zhiyong.wang}@sydney.edu.au
[2] The University of Melbourne, Melbourne, VIC 3010, Australia
ling.luo@unimelb.edu.au
[3] University of Technology Sydney, Sydney, NSW 2007, Australia
{Jianjia.Zhang,Yang.Wang}@uts.edu.au

Abstract. With the increasing number of commodities in our daily life, the recommender system plays a more and more important role in selecting items of users' interests. For the next basket recommendation task, in this work, we propose the first end-to-end correlation-aware model to predict the next basket considering intra-basket correlations using graph attention networks. Specifically, items and correlations between items are viewed as nodes and edges in a graph, respectively. By estimating and aggregating the intra-basket correlations using the attention layer of the self-attention model, the recommendation can be conducted at the basket level, instead of at the item level. We conduct comprehensive experiments on a real-world retailing dataset to show the improvement from state-of-the-art baselines using our proposed method.

Keywords: Recommender systems · Next basket recommendation · Graph convolutional neural networks

1 Introduction

People have been facing more and more options in their daily life, such as reading news, purchasing books, watching movies and listening to music. A recommender system is a powerful tool to feed people the items of their interests. However, few works have considered that, in many scenarios, multiple items are purchased together in one transaction. For example, when users go to a supermarket, they usually purchase a group of items instead of a single one. Such a group of items bought at the same time by the same user are referred to as a basket. The next basket recommendation is a task of recommending a basket of items when the historically purchased baskets are provided as input, which will be formally defined later.

H. Yang et al. (Eds.): ICONIP 2020, CCIS 1332, pp. 746–753, 2020.
https://doi.org/10.1007/978-3-030-63820-7_85

Figure 1 is an example of the next basket recommendation in the supermarket scenario. For each user, his transaction history consists of a sequence of baskets. The target is to predict what items belong to the next basket to be purchased in the future. In the transaction data, the size of baskets and the length of basket sequences are usually variable. As shown in Fig. 1, items in a particular basket are usually correlated and show a certain purpose. The first basket of *Cereal, Milk, Bread* and the last basket of *Tooth Paste, Tooth Brush* may imply a latent intention of preparing for breakfast and brushing teeth, respectively. Inspired by this observation, the next basket recommendation should involve related items, rather than independent items. In this way, modeling the intra-basket correlations between items are essential for the next basket recommendation.

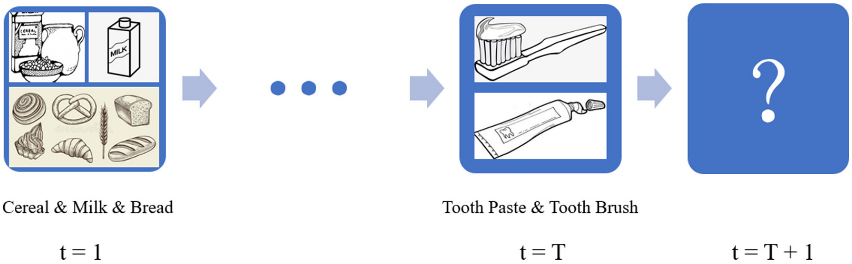

Fig. 1. Illustration of the next basket recommendation task

To explore the correlation among items, a deep-learning-based model called Beacon [5] was recently proposed. In that model, the correlation among items is considered when encoding each basket. The correlation is firstly computed by counting the frequency of co-occurrences of item pairs and then fed into the basket encoder to create the correlation sensitive basket representation. However, in this work, the correlation estimation model and the main deep recommendation model are independent, i.e. the correlation among items needs to be computed externally. In order to better integrate the entire model, we propose an end-to-end graph deep learning model, in which nodes and edges represent items and correlations, respectively. Since there should exist an edge (correlation) between relevant items, in our model, the edge is estimated by self-attention from relevant item embeddings. The embeddings of relevant items are then aggregated with weights of correlations to form the correlation-aware basket representation.

Our main **contributions** are:

1. We explore the intra-basket correlation which is essential for basket analysis.
2. We propose an end-to-end model to estimate and aggregate correlations automatically during the training procedure.
3. We achieve better performance when comparing with state-of-the-art baselines for next basket recommendation.

2 Related Works

At an early age, the Markov chain was firstly proposed for this task [6]. In this work, matrix factorization was used to learn the general taste of a user, and Markov Chain modeled sequential behavior by learning a transition graph over items. [4] proposed to recommend the next basket based on K-nearest neighbors with the Wasserstein distance, but this model was difficult to apply on large scale datasets due to the huge computational cost. Recently, with emerging deep learning-based methods, a pioneering work [10] firstly used neural networks with fully-connected layers. The basket was represented as the average of item embeddings and they concatenated basket representations for basket sequence representations. After that, Recurrent Neural Networks [11] was proposed for better exploiting the sequential information of baskets. [1] considered the item attributes such as product category based on RNN's architecture. Although the aforementioned models achieved great basket recommendation performance, they ignored to model intra-basket correlation, which is quite essential in basket analysis. [9] considered the compatibility in baskets but was short for capturing long-term dependency. [2] developed a deep learning-based model to take the correlation into account but they ignored the plentiful sequential information. [5] considered the correlation among items for basket representation but they estimated the correlation by counting co-occurrence of items while predicting the next basket by RNNs. These two parts are independent and parallel which hurts the integration of the whole model. To bridge the research gaps, we propose to model the next basket recommendation task by graph attention networks so that the correlation among items could be estimated directly using the end-to-end deep learning model.

3 Methodology

3.1 Problem Formulations and Notations

The next basket recommendation problem is formally formulated as following. Assume there are m users $U = \{u_1, u_2, ..., u_m\}$ and n items $V = \{v_1, v_2, ..., v_n\}$. For each user, a basket B_t is a subset of V ($B_t \subseteq V$) consisting of several items purchased in his t^{th} transaction. The intra-basket correlation among items is estimated and stored in a matrix $C \in R^{n \times n}$, in which the entry c_{ij} is the correlation between v_i and v_j. The transaction history of a user is viewed as a sequence of baskets $< B_1, B_2, ...B_T >$. Note that the length of basket sequence T is different for different users. We aim to predict the next basket of each user B_{T+1} considering the intra-basket correlation among items.

3.2 Overview of the Model

Generally, our model use Graph Attention Network (GAT)[8] to model this problem. The items V are viewed as nodes in the graph and the correlations are items are viewed as edges in the graph which are estimated by the attention layer with scaled dot-product attention [7]. Figure 2 illustrates the overall structure of our model. Firstly, items are embedded into d-dimensional space as vector representations. After that, the item embeddings are fed to the attention layer to compute the correlation. Then we encode each basket for correlation-aware representation according to the embeddings of items inside the basket. For each user, there exists a sequence of baskets and we feed baskets into LSTM networks in terms of their chronological order. The output of LSTM is then used to compute the final probability of each item belonging to this basket.

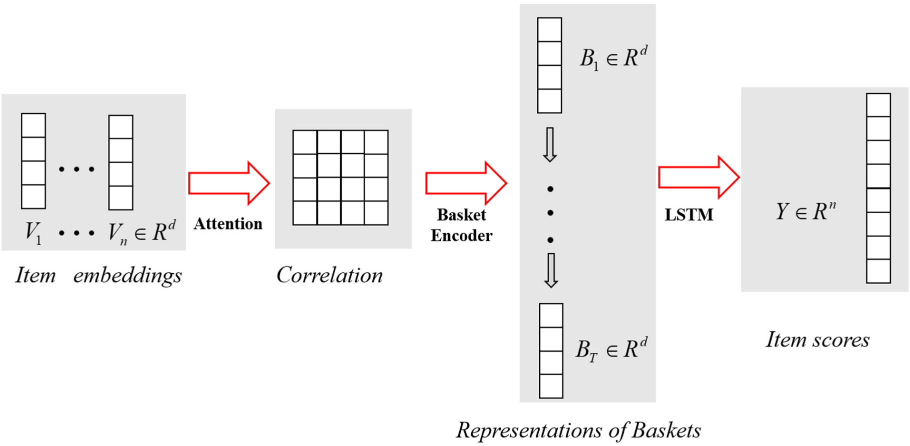

Fig. 2. Overall structure of our model

3.3 Item Embeddings and Correlation Estimation

We use a d-dimensional embedding vector $v_i \in R^d$ to represent each node (item) in the graph. Then the correlation among items can be estimated by the attention layer using the embedding vectors of items. Given the d-dimensional embedding vectors of n items $\{v_i\}_n$, a matrix $I \in R^{n \times d}$ can be generated to store the embeddings of all items. Correlation $\mathbf{C} \in R^{n \times n}$ between items is computed by the scaled dot-product attention [8] with normalization of $\sqrt{d}$ by

$$\mathbf{C} = \text{softmax}\left(\frac{IW_k(IW_q)^\top}{\sqrt{d}}\right), \tag{1}$$

where the $W_k \in R^{d \times d}$ and $W_q \in R^{d \times d}$ are learnable parameters for asymmetric item embeddings. The softmax is row-wise, i.e. conducted on each row of the matrix for normalization. This equation implies that the correlation between items is computed from the similarity between nodes.

3.4 Correlation-Aware Basket Encoder

In this section, we would introduce the basket encoder to get the correlation-aware representation of a basket from correlation C and item embeddings. For each basket B_t, a binary row vector $\mathbf{x}_t \in \{0, 1\}^n$ is generated to indicate whether an item is inside this basket, where the i^{th} value is 1 only when the i^{th} item is inside the basket. The correlation-aware representation $\mathbf{b}_t$ of the basket B_t is a d-dimensional vector computed by:

$$\mathbf{b}_t = \mathbf{x}_t \mathbf{C} I. \tag{2}$$

This equation indicates the basket representation is computed by aggregating the embeddings of the items belonging to this basket and their neighboring nodes with the weight of correlation.

Additionally, in practice, weak correlation is more likely to be noise that may negatively impact the basket representation. Therefore, a trainable scalar parameter $\eta \in \mathbb{R}^+$ is leveraged to filter out weak correlation.

$$\mathbf{b}_t = \mathrm{ReLU}\left(\mathbf{x}_t \mathbf{C} - \eta \mathbf{1}\right) I, \tag{3}$$

where $\mathbf{1} \in \{1\}^n$ is a row vector of ones and ReLU is applied for each element.

3.5 Next-Basket Recommendation

After obtaining representations of all baskets, Long-Short Term Memory networks (LSTM) is used to model the sequence of baskets. Given the representations of a sequence of baskets $< B_1, B_2, ...B_T >$, the recurrent H-dimension hidden output $\mathbf{h}_t \in \mathbb{R}^H$ at step t is computed by

$$\mathbf{h}_t = \tanh\left(\mathbf{b}_t \varPsi + \mathbf{h}_{t-1} \varPsi' + \boldsymbol{\psi}\right), \tag{4}$$

where $\varPsi \in \mathbb{R}^{d \times H}$, $\varPsi' \in \mathbb{R}^{H \times H}$ and $\psi \in \mathbb{R}^H$ are weight and bias parameters to be learned.

The item probability is then computed by the output of the final layer of LSTM $\mathbf{h}_T \in R^H$:

$$\mathbf{S} = \sigma(\mathbf{h}_T W), \tag{5}$$

where $W \in \mathbb{R}^{H \times n}$ and the σ is the sigmoid function to ensure the probability is in the range from 0 to 1. The output $\mathbf{S} \in R^n$ implies the probability of each item belonging to the predicted next basket.

Additionally, to get the correlation-aware output, we combine the output probability S with the correlation C with a hyperparameter α controlling the trade-off.

$$\mathbf{y} = \alpha(\mathbf{S}) + (1 - \alpha)(\mathbf{SC}), \tag{6}$$

The final recommendation is based on the predicted item scores $\mathbf{y}$. Since the size of the next basket is various and is often noncritical in the setting of previous works. In our work, we follow the traditional way of setting the basket size as a constant small number k. The final predicted next basket takes k items with the highest probability in $\mathbf{y}$.

3.6 Loss Function

Our goal is to make the predicted basket similar to the ground truth. The last basket B_{T+1} in each sequence is removed as ground truth and the remaining sequence $< B_1, B_2, ...B_T >$ is for training.

Generally, to make the predicted scores $\mathbf{y}$ closer to the ground truth B_{T+1}, we adopt weighted cross-entropy as the loss function.

$$\mathcal{L} = -\frac{1}{|B_{T+1}|} \sum_{i \in B_{T+1}} \log(\mathbf{y}_i) - \frac{1}{|V \backslash B_{T+1}|} \sum_{j \in V \backslash B_{T+1}} \log(1 - \mathbf{y}_j) \tag{7}$$

During training, the probability $\mathbf{y}$ of adopted items in the ground truth basket B_{T+1} is encouraged to increase (the first term) while the probability of other negative items in $V \backslash B_{T+1}$ is decreasing (the second term). The weights $|B_{T+1}|$ and $|V \backslash B_{T+1}|$ are used to fix the biases in the unbalanced data, i.e. the purchased items are much less than unpurchased ones.

4 Experiments

4.1 Dataset

The dataset used in our experiments is the TaFeng dataset[1], which is the most popular in the basket analysis research domain. TaFeng is a transaction dataset of a grocery store from November 2000 to February 2001 and each transaction can be viewed as a basket in our experiments. There are 32,266 users and 23,812 items with 817,741 transactions in total.

For preprocessing, as [5], users who bought at least 10 items and items which were bought by at least 10 users are selected for experiments. Additionally, each user must purchase at least 3 baskets to ensure a sequence of at least 2 baskets for training. In the preprocessed dataset, the average basket size is 5.9 and the average length of sequences of baskets is 7.0. To get the training, validation and test sets, we chronologically split the whole dataset. Specifically, since the TaFeng dataset spanned 4 months, the basket sequences in the first 3 months are considered for training. The baskets purchased from the 3 to 3.5 months and from the 3.5 to 4 months are for validation set and test set, respectively.

4.2 Evaluation Metrics

The evaluation metrics we use is the F1 score, which measures the similarity between the ground truth and the predicted basket. The F1 score combines the recall rate and precision rate. A large predicted basket hurts the precision rate because of including many irrelevant items outside the true basket while a small one is hard to include all relevant items, which hurts the recall rate. Since the average basket size is 5.9 in the TaFeng dataset, the F1 score at 5 is used in the experiments, which means the size of the predicted next basket is set as 5.

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

4.3 Baselines

We compare our model with a series of state-of-the-art baselines.

- POP takes the most popular items as the predicted basket.
- triple2vec[2] [9] considers the correlation among items in baskets without encoding the sequential information of baskets.
- DREAM[3] [11] is a basket recommendation model using recurrent neural networks without considering correlation.
- Beacon[4] [5] is a recent deep learning-based model. The basket representation employs correlation among items which is computed by counting the frequency of being purchased together.

4.4 Experimental Results

The experimental results are shown in Table 1. The final results on the test set are from the model with the best performance on the validation set after convergence. The optimizer we use is Adam [3]. Our model is trained for 50 epochs in total with batch size of 32. The hyperparameters are tuned by grid search, and the values used in our experiments are $d = 10$, $\alpha = 0.5$, $H = 16$ and learning rate $= 0.001$.

Table 1. Comparison of F1 score at 5 with baselines

Models	POP	triple2vec	DREAM	Beacon	**Our Model**
F1@5(%)	4.66	4.66	5.85	6.27	**6.44**

It is observed that our correlation-aware graph deep learning model performs better than other baselines[5]. The POP method naively takes the most popular items without modeling correlation and sequential information so that it achieves the lowest performance. The triple2vec takes the correlation into account but performs also poorly due to the lack of sequential information. DREAM leverages RNNs to encode the basket sequence information but ignores the correlation. Beacon considers both sequential information and correlation among items so that it gets improvement from DREAM's basic RNN structure. Note that the correlation in Beacon is estimated by counting the co-occurrence of item pairs. Benefit from the great expression ability of the attention model, our model can capture the correlations well with abundant transaction data. Even if our model does not leverage external correlation information, the prediction can still be more accurate. This implies that our end-to-end model can estimate and aggregate more helpful correlations automatically for the recommendation than the external items co-occurrence information.

[2] https://github.com/MengtingWan/grocery.
[3] https://github.com/LaceyChen17/DREAM.
[4] https://github.com/PreferredAI/beacon.
[5] Results of POP, triple2vec and DREAM are from [5].

5 Conclusion

In this paper, we present an end-to-end correlation-aware model for the next basket recommendation using graph attention networks. In our model, the correlation is estimated by the attention model directly during learning. The correlation-aware basket representation is then computed by aggregating embeddings of the items inside and their neighboring items with the weight of correlation. After that, the basket sequence is fed into LSTM for sequential information encoding. We compare our model with state-of-the-art baselines on a popular real-world basket grocery dataset and achieves the best performance, which demonstrates the benefits of taking the correlation between items into account for the next basket recommendation. In future work, we will make the size of the predicted basket adaptive for each user instead of a constant number.

References

1. Bai, T., Nie, J.Y., Zhao, W.X., Zhu, Y., Du, P., Wen, J.R.: An attribute-aware neural attentive model for next basket recommendation. In: ACM SIGIR (2018)
2. Gong, Y., et al.: Exact-k recommendation via maximal clique optimization. In: ACM Knowledge Discovery and Data Mining (KDD) (2019)
3. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization (2015)
4. Kraus, M., Feuerriegel, S.: Personalized purchase prediction of market baskets with Wasserstein-based sequence matching. In: ACM Knowledge Discovery and Data Mining (KDD) (2019)
5. Le, D.T., Lauw, H.W., Fang, Y.: Correlation-sensitive next-basket recommendation. In: International Joint Conference on Artificial Intelligence (IJCAI) (2019)
6. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized Markov chains for next-basket recommendation. In: International World Wide Web Conference (WWW) (2010)
7. Vaswani, A., et al.: Attention is all you need. In: Conference on Neural Information Processing Systems (NIPS) (2017)
8. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. The International Conference on Learning Representations (ICLR) (2018)
9. Wan, M., Wang, D., Liu, J., Bennett, P., McAuley, J.J.: Representing and recommending shopping baskets with complementarity, compatibility and loyalty. In: International Conference on Information and Knowledge Management (CIKM)(2018)
10. Wan, S., Lan, Y., Wang, P., Guo, J., Xu, J., Cheng, X.: Next basket recommendation with neural networks. In: ACM Conference on Recommender Systems (RecSys) (2015)
11. Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T.: A dynamic recurrent model for next basket recommendation. International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2016)