Next Basket Recommendation Model Based on Attribute-Aware Multi-Level Attention

TONG LIU, XIANRUI YIN, AND WEIJIAN NI
College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China
Corresponding author: Weijian Ni (niweijian@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71704096 and Grant 61602278, and in part by the Qingdao Philosophy and Social Sciences Planning Project under Grant QDSKL1801122.

ABSTRACT Next basket recommendation is a challenging problem, mainly due to the relationships among the items in a basket almost not being considered in current research. In this article, we address next basket recommendation with a novel deep learning architecture. In particular, we consider both the short-term user interests and the long-term user preferences, and we design a new attention that considers the relationships among the items in a basket. We extensively evaluated the proposed model on two benchmark data sets, the Ta-Feng and JingDong datasets. The experimental results show that the proposed model outperforms several state-of-the-art next basket recommendation models. In the experiments of the modules’ effects, we also verify the effectiveness of each module. The model significantly improves the NDCG by 25.5 percentage points and 5 percentage points when compared with the uncompleted network model on the JingDong and Ta-Feng datasets, respectively; while in terms of the F1, the performance is improved by 14.6 and 23.4 percentage points, respectively.

INDEX TERMS Next basket recommendation, deep learning, self-attention, item attributes.

I. INTRODUCTION
In recent years, increasingly e-commerce platforms have emerged, such as Taobao, JingDong and Amazon, and recommendation systems play important roles in them. Recommendation systems can identify users’ needs and increase sales in these platforms. Although recommendation systems have been proven useful [1], [2], recommending items that are highly related to users is a challenging problem [3]. In addition, recent studies have focused on recommending lists of items to users [4], [5], where each item is independent. However, in the real scenario, the items in a user’s basket are often related. The task of recommending a basket of items to users is called “next basket recommendation” [6]. It uses a user’s historical purchase to predict what the user will buy next. Next basket recommendation helps e-commerce websites to present the right items to users, thus promoting the growth of their trading volumes and profits, and it can help users quickly select the right items from massive numbers of items.

The difference between next basket recommendation and general recommendation is that it should recommend a basket of items to users each time; therefore, the relationships between the items inside the basket should be considered. Therefore, when analyzing the historical baskets of users, the implicit information between the items and users should be more fully mined. First, users’ short-term purchase behavior and their long-term purchase behavior should be considered, both of which reflect users’ preferences for items. Second, the preferences of users change over time, and so we need to mine the implicit information using the purchase history sequence of users. In addition, it is also important to find the dependencies between the baskets and between the items in the basket. In combination with the basket recommendation application scenario, we summarized two points to consider in modeling:

- Mining users’ long-term and short-term preferences from their historical purchase sequence, and
- Considering the relationships between the items in the basket and the relationships between baskets.

Using machine learning in recommendation systems has been a hot research area. The model based on collaborative filtering can find the overall interest of the user, but it is difficult to consider the sequential characteristics of historical transactions. The order feature can be considered in
the Markov model, but the high order Markov model is too complex. In recent years, deep learning has also been widely used in the recommendation field. A session-based recommendation system used an RNN to model users’ behavioral sequences [8]. This work regards the behavior of clicking on a commodity in a session as a sequence and applies the GRU to mine the sequence information. A hierarchical attention network to solve the sequential recommendation problem was proposed [9]. Through the hierarchy layer, this work combined the long-term and short-term preferences of users to generate an advanced hybrid representation of users. A hybrid model was proposed that incorporated the gating mechanism of learning, which can combine different models given different context information [10]. Regarding next basket recommendation, many models have also been proposed: the HRM aggregates the representation of items and the representation of users through nonlinear maximum pooling [11], the DREAM uses an RNN to model the next basket sequence for the first time [12], and the ANAM integrates the attribute information of the item into the item representation [13]. However, what the above works miss is that they do not explicitly focus on the relationships between the items inside the basket. To solve the problem, this article proposes a novel Next basket Recommendation Model Based on Attribute-aware Multi-level Attention (NbR-AM). Firstly, we pay attention to the item information inside the basket, construct the basket hybrid representation with the self-attention mechanism at the basket level, then apply the attention mechanism to a user’s basket sequence level to pay attention to different baskets, and finally recommends the next basket.

The main contributions of this article are summarized as follows:

- The basket representation is constructed by applying self-attention to the items in the basket, basket representations are encoded by bi-LSTM as a latent representation.
- The second attention layer focuses on the user’s different baskets to construct the representation of the user’s purchase intention.
- When recommend the item in the next basket in the recommendation layer, the context representation and previous time step recommended item are fully considered.

We performed experiments on two common datasets and compared them with other existing models. The experimental results show that the proposed model improves the effectiveness of basket recommendations.

The remainder of this article is organized as follows. Section II briefly reviews the key research works in the literature related to this area. The Problem Formulation is presented in Section III. In Section IV, a next basket recommendation model based on attribute-aware multi-level attention is proposed and its algorithm are described in detail. Extensive experimental results and discussions is reported in Section V. Finally, the conclusions related to this research study are drawn in Section VI.

II. RELATED WORK

In order to provide personalized recommendation to users, review-based models, social-based models and behavior-based models have been used. The SliRec explores the possibility of fusing representation learning-based CF models and matching function learning-based CF models [21]. In addition, the Sli-Rec has done an excellent job in finding a user’s long-term and short-term preferences in the recommendation system [22]. Various

Deep learning based models are a hot research direction of basket recommendation in recent years. There is also a lot of excellent work combine deep learning and traditional models. The DeepCF explored the possibility of fusing representation learning-based CF models and matching function learning-based CF models [21]. In addition, the Sli-Rec has done an excellent job in finding a user’s long-term and short-term preferences in the recommendation system [22]. Various
deep neural network models have been applied to solve the basket modeling and sequence modeling problems in basket recommendation. The HRM first conducts the maximum pooling operation for the embedding vector of the recently purchased items, and then it constructs the user behavior representation using the maximum pooling operation between the user’s embedding vector and the first layer representation to form the next item recommendation [11]. However, both FPMCs and the HRM are based on Markov chain approaches, which model the sequential behavior of users only using adjacent transactions, which is not sufficient to capture the long-term trend of the basket. To solve this problem, the DREAM uses an RNN to model the global sequential characteristics between baskets and uses the hidden states of the RNN to represent the dynamic interests of users over time [12]. The ANAM uses a hierarchical structure to apply an independent attention mechanism to items and their respective attributes, uses an RNN to model the user’s sequential behavior over time, and passes the user’s preferences for items and their attributes to the next basket by sharing the attention weight between two baskets in different hierarchies [13].

III. PROBLEM FORMULATION
In the next basket recommendation scenario, we have a set of users $U = \{u_1, u_2, \ldots, u_{|U|}\}$ and items $I = \{i_1, i_2, \ldots, i_{|I|}\}$, where $|U|$ and $|I|$ are the number of users and the number of items, respectively. Given a user $u$ and this user’s existing historical basket record $B^u_t = \{B^u_{t1}, B^u_{t2}, \ldots, B^u_{tn}\}$, where $B^u_{ti}$ is the basket of user $u$ at time step $t$ in the basket sequence, $B^u_t \subseteq I$ are sorted chronologically. The historical basket records of all users denoted as $B = \{B^u_i\}_{u=1}^{|U|}$. Each item has some attribute information, such as the category information of the item. The attributes of item $i$ are expressed as $a_i \in A$, where $A$ is the set of attributes. The attribute information of basket $B^u_{ti}$ is expressed as $A^u_{ti} = \{a_i| i \in B^u_{ti}\}$. The goal of next basket recommendation is use a user’s historical transaction records to predict the next basket $B^u_{t+1} = \{i_1, i_2, \ldots, i_m\}$.

IV. THE PROPOSED APPROACH
In this article, the Next basket Recommendation Model Based on Attribute-Aware Multi-Level Attention is proposed. Figure 1 shows the overview of the architecture of the proposed model, which consists of four layers: the basket layer, the sequence layer, the recommendation layer, and the output layer. In the basket layer, the basket embedding $v_{basket}$ is calculated by the self-attention mechanism. The input of the self-attention mechanism is the embedding sequence of items $\{i_1, i_2, \ldots, i_m\}$ and the output of the self-attention is basket embedding $v_{basket}$. In the sequence layer, user’s basket representation sequence $\{v_{basket}^1, v_{basket}^2, \ldots, v_{basket}^m\}$ is input into the bi-LSTM so as to determine the long-term interests and short-term preferences in the user’s sequential behavior. These basket representations are encoded by bi-LSTM as a latent representation $h_t$. In the recommendation layer, the second level of the attention mechanism is used to pay different degrees of attention to these latent representations to model the relationships between baskets so as to obtain the current basket’s context representation $c_t$. The next hidden state $s_t$ of the decoder at time step $t$ is computed by the context representation $c_t$, the previous time step hidden state $s_{t-1}$ of LSTM and the global embedding (GE). In the output layer, the hidden state of LSTM in each time step is used the recommendation layer to calculate the probability that each item may be purchased. In each time step, we select the item with the highest probability as the recommended item for users. To prevent repeated predictions, we use the masked softmax (MS) to output the probability distribution.

A. THE BASKET LAYER
The self-attention mechanism was first proposed by Google for machine translation and it achieved the optimal results at that time [23]. Self-attention uses an attention mechanism to calculate the association between each word and other words. The basic structure is shown in Figure 2.

For self-attention, $Q, K$ and $V$ come from the same input. First, we need to calculate the dot-product between $Q$ and $K$; and then to prevent the result from being large, it is divided by a scale $\sqrt{d_k}$, where $d_k$ is a dimension of a Query and Key vector. The result is then normalized to a probability distribution using the softmax operation, and then multiplied by matrix $V$ to produce a sum of the weights, which can be formalized as follows:

$$\text{self-attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

(1)
Multi-head attention can have different representations $\text{head}_i$ of $Q, K$ and $V$ for the same input, and then combine the self-attention results. The structure of multi-head attention is shown in Figure 3 and the formula is as follows:

$$\text{MultiHead}(Q, K, V) = [\text{head}_1, \ldots, \text{head}_h]W^o,$$  \hspace{1cm} (2)

where $\text{head}_i = \text{self-attention}(QW^Q_i, KW^K_i, VW^V_i)$.

The information of items and their attributes in the basket can reflect the purchase intention or preference of the users from different aspects, which play a complementary role in the information representation of the basket. In Equations (1) and (2), we take the category and id of the items in the basket as the input of the self-attention mechanism to get two kinds of basket representations $v_{\text{basket}}^{\text{category}}$ and $v_{\text{basket}}^{\text{item}}$, respectively, and then concatenate the two vector representations together to get the final attribute-aware basket representation $v_{\text{basket}}$. The implementation of the lookup operation is as follows:

$$v_{\text{basket}}^{\text{category}} = \text{self-attention}(Q^{\text{category}}, K^{\text{category}}, V^{\text{category}}),$$ \hspace{1cm} (3)

$$v_{\text{item}}^{\text{category}} = \text{self-attention}(Q^{\text{item}}, K^{\text{item}}, V^{\text{item}}),$$ \hspace{1cm} (4)

$$v_{\text{basket}} = [v_{\text{category}}, v_{\text{item}}].$$ \hspace{1cm} (5)

B. THE SEQUENCE LAYER

The Sequence to Sequence model is a popular RNN model in recent years, which is widely used in machine translation, automatic question answering systems and other fields, and has achieved good results [24], [25]. As we all know, LSTM is a popular model to deal with sequence labeling because it solves the disappearing disappearance and exploding gradient problems of the traditional RNN in the long sequence training process [26]. Through LSTM’s three gates mechanisms, each hidden state can express the effective information of the current basket and the previous basket. Further, bi-LSTM can perform long-term memory and short-term memory simultaneously from two directions [27], which is conducive to mining more potential preference relationships from the user’s behavior sequence. We use bi-LSTM in basket sequence modeling. The hidden state representation $h_i$ not only contains the basket information from the front of the current basket, but it also contains the basket information from the back of the current basket. This helps the model to learn the purchase preferences of users as a whole. For the basket representation sequence $[v_1^{\text{basket}}, v_2^{\text{basket}}, \ldots, v_m^{\text{basket}}]$, in order to determine the user’s long-term and short-term purchasing preferences, we used LSTM from two directions to get the hidden states $\overrightarrow{h}_i$ and $\overleftarrow{h}_i$ of each basket and concatenate these two vectors to get the current basket representation $h_i$, which reflects the sequence information centered on the current basket.

$$\overrightarrow{h}_i = \text{LSTM}(\overrightarrow{h}_{i-1}, v_i^{\text{basket}}),$$ \hspace{1cm} (6)

$$\overleftarrow{h}_i = \text{LSTM}(\overleftarrow{h}_{i-1}, v_i^{\text{basket}}),$$ \hspace{1cm} (7)

$$h_i = [\overrightarrow{h}_i, \overleftarrow{h}_i].$$ \hspace{1cm} (8)

C. THE RECOMMENDATION LAYER

In the traditional RNN structure, when the input sequence is very long, the context vector is not enough to represent the whole input sequence, which makes it difficult to deal with long sequences. To solve this problem, the attention mechanism is introduced. The attention mechanism has been widely used in the fields of computer vision and natural language processing [28]–[31]. In the next basket recommendation problem, many users have very long histories of baskets of items, with dozens of baskets. Due to the existing basket sequence having different contributions to the next basket prediction, we use the attention mechanism to pay attention to different parts of the basket representation sequence and aggregate $m$ basket representations to generate the context representation $c_t$, which can more effectively express the purchase intentions of users at a certain time, and thus recommend items to the users more accurately. In particular, the attention mechanism at time step $t$ assigns weights $\alpha_t$ to the $i$-th basket representation.

$$e_t = U_a^T \tanh(W_a \cdot s_{t-1} + H_a \cdot h_i),$$ \hspace{1cm} (9)

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^{m} \exp(e_j)},$$ \hspace{1cm} (10)
where \( W_a, H_a, \) and \( U_a \) are the weight parameters, and \( s_{t-1} \) is the hidden state of the LSTM at step \( t-1 \). \( e_t \) is the weight influence factor. We believe that the weight of the current basket representing \( h_t \) is not only related to \( h_t \) itself, but also to the state of other time steps. The final context vector \( c_t \), associated with the recommendation at time \( t \) is calculated as follows:

\[
c_t = \sum_{i=1}^{m} \alpha_i h_i, \tag{11}
\]

In our model, we use the users’ history basket in time steps \( \{1, 2, \ldots, t\} \) to recommend the next basket in time step \( t+1 \). The process of recommending a basket of several items is loops, as shown in Algorithm 1. In the recommendation layer, each step in the loops consider the recommended item of previous steps to recommend the item by the global embedding. In this layer, unidirectional LSTM is used. The context representation \( c_t \), the last step hidden state \( s_{t-1} \) of LSTM and the global embedding \( g_{t-1} \) of the output layer are used as the inputs to calculate the next hidden state representation \( s_t \) of LSTM.

\[
s_t = \text{LSTM}(s_{t-1}, [g_{t-1}; c_t]), \tag{12}
\]

where \([g_{t-1}; c_t]\) means the concatenation of the vectors \( g_{t-1} \) and \( c_t \). The global embedding \( g_{t-1} \) is the average of the embeddings of the predicted items at previous \( t-1 \) steps in the loops. Each step in the loops, we select the item which have the highest probability as the recommend item and use the items’ embedding to calculate the global embedding \( g_{t-1} \). When \( t = 1 \), \( g_{t-1} \) is a random initialized vector. When \( t \geq 2 \), \( g_{t-1} \) is computed as follows:

\[
g_{t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} \text{emb} \left( \text{argmax}_{j=1, \ldots, |l|} (y_{i,j,1}, \ldots, y_{i,j,i}, \ldots, y_{i,j,|l|}) \right), \tag{13}
\]

where \( y_{i,j} \in [0, 1] \) s.t. \( \sum_{j=1}^{|l|} y_{i,j} = 1 \). The \( y_{i,j} \) is the probability that item \( j \) being purchased by users in the \( i \)-th loop step.

### D. THE OUTPUT LAYER AND MODEL LEARNING

In this layer, we recommend a basket with five items to user. In the output layer, \( y_i \) is the probability distribution that each item is likely to be bought at time step \( t+1 \) and is calculated as follows:

\[
o_t = W_{of}(W_d \cdot s_t + V_d \cdot c_t), \tag{14}
\]

\[
y_t = \text{soft max}(o_t + I_t), \tag{15}
\]

where \( W_o, W_d, \) and \( V_d \) are weight parameters, and \( I_t \) is the mask vector that is used to avoid recommend repeated items in the output layer. If the item has been recommended in the previous \( t-1 \) steps, \( (I_t)_i \) set to \(-\infty\). In addition, \( f \) is a nonlinear activation function.

\[
(I_t)_i = \begin{cases} 
-\infty, & \text{if the item } i \text{ has been recommended} \\
0, & \text{otherwise}. 
\end{cases} \tag{16}
\]

In order to effectively learn from the training data, we use the weighted cross-entropy as the loss function, which is defined as follows:

\[
L = \sum_{u \in U} \sum_{B^u_t \in B^u} \sum_{i \in I_t} (-k \cdot y_i \cdot \log p_i - q \cdot (1 - y_i) \cdot \log(1 - p_i)). \tag{17}
\]

where \( k \) and \( q \) are the weights of the positive and negative classes, respectively. \( p_i \) is the probability of item \( i \) being purchased in the next basket. The \( y_i \) indicates whether the user purchased the item \( i \). If the user purchased item \( i, y_i = 1 \); otherwise, \( y_i = 0 \).

We further formalize the above two levels of the proposed approach in Algorithm 1:

### Algorithm 1 Next Basket Recommendation Model Based on Attribute-Aware Multi-Level Attention

**Input:**
- Users’ basket sequence \( \{B^u_1, B^u_2, \ldots, B^u_m\} \)

**Output:**
- Recommend the next basket \( R(u) \) for each user

**Begin**
1. for each \( u \in U \) do
2. construct the basket representations \( v_{\text{basket}} \), \( v_{\text{category}} \) and \( v_{\text{item}} \) via self-attention with items and categories, respectively
3. \( v_{\text{basket}} \leftarrow [v_{\text{basket}}] \)
4. the basket representation sequence \( \{v_1^{\text{basket}}, v_2^{\text{basket}}, \ldots, v_m^{\text{basket}}\} \) is input into the bi-LSTM to get the hidden representation \( \{h_1, h_2, \ldots, h_m\} \)
5. the hidden representation put into the second attention mechanism to get the context vector \( c_t \) via Equation (11)
6. The context vector \( s_t \) is obtained via Equations (12) and (13).
7. get the probability \( y_t \) of item maybe purchased via Equations (14) and (15)
8. select one item with highest probability and put it into \( R(u) \)
9. end for
10. return \( R(u) \)
11. end for
**End**

### V. EXPERIMENTS

#### A. DATASETS

In order to evaluate the performance in next basket recommendation, we use two real-world transaction data sets to conduct the experiments.

- The Ta-Feng dataset is a public dataset released by the RecSys, which contains a Chinese grocery store transaction data from November 2000 to February 2001.
The dataset contains 817,741 transactions, belonging to 32,266 users and 23,812 items.

- The JD dataset provides the behavioral, evaluation and user data of some commodities from February 1, 2016 to April 15, 2016.

We preprocessed the above two datasets to remove those items that were seldom purchased and those users who seldom purchased items. Users who purchased less than 15 items were deleted, and items that were purchased less than 15 times were deleted. Table 1 summarizes the statistics of the two datasets after data preprocessing.

Finally, we divided all the data sets into training sets and test sets. The test sets contain only the last transaction for each user while all other transactions are used the training sets.

### B. BASELINES

We use the following representative start-of-the-art models as the baselines for the experiments:

- **NMF**: The recommended model based on non-negative matrix factorization is proved to be the best model among the traditional models [32].
- **HRM**: It is a hierarchical presentation model that uses a shallow neural network and nonlinear operations to aggregate the embedding representation of the user and the embedding representation of the last purchased items [11].
- **DREAM**: The model uses an RNN to model the global sequential characteristics between baskets and uses the hidden states of the RNN to represent the dynamic interests of users over time [12].
- **ANAM**: The model uses a hierarchical structure to apply an independent attention mechanism to items and their attributes respectively; uses an RNN to model the user’s sequential behavior over time; and passes the user’s preference for items and their attributes to the next basket by sharing the attention weights between two baskets in different hierarchies [13].
- **IIAAN**: The model applies an attention mechanism on the basket sequence to act on all historical user baskets to model the user’s long-term preferences, and within the basket to act on the commodity level in the most recent basket to model the user’s dynamic interests and short-term preferences [33].

The traditional NMF model is based on the scikit-learn 0.20.3 machine learning library. The HRM, DREAM, ANAM and IIAAN models are based on deep learning and implemented with PyTorch 1.0.

The model proposed in this article are implemented using PyTorch 1.0 [34]. In our experiments, we recommend 5 items to each user. The embedding dimension is set to 100, and the number of multi-head attention mechanisms is set to 2. The loss function is optimized by the Adam algorithm. Due to the difference in the numbers of positive and negative instances in the loss function, we set \( k \) to 800 times as big as \( q \) in the experiment for the Ta-Feng dataset, and set \( k \) to 1000 times as big as \( q \) for JingDong dataset in Equation (17).

### TABLE 1. Statistics of datasets.

| Datasets | Users | Items | Transactions | Categories |
|----------|-------|-------|--------------|------------|
| Ta-Feng  | 8143  | 7323  | 575690       | 1007       |
| JD       | 16891 | 17718 | 250996       | 10         |

The corresponding learning rates for the two datasets are set as 0.002 and 0.004. Dropout was not used in the experiment. The batch size is 64 in the training stage. The number of epochs is 100.

### C. EVALUATION METRICS

For each model, we recommend a list of \( N \) items \((N = 5)\) for user \( u \) denoted as \( R(u) \), where \( R_i(u) \) represents the \( i \)-th recommended items. The baskets actually purchased by users are recorded as \( T(u) \). The performance is evaluated on the test set using the following metrics:

\[
\begin{align*}
\text{Precision}(T(u), R(u)) &= \frac{|T(u) \cap R(u)|}{|R(u)|}, \\
\text{Recall}(T(u), R(u)) &= \frac{|T(u) \cap R(u)|}{|T(u)|}, \\
\text{F1} - \text{score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \\
\text{Hit} - \text{Ratio} &= \frac{\sum_{u \in U} I(T(u) \cap R(u) \neq \emptyset)}{|U|}, \\
\text{NDCG@K} &= \frac{1}{N_k} \sum_{j=1}^{k} \frac{2^{I(R_j(u) \in T(u))} - 1}{\log_2(j + 1)}.
\end{align*}
\]

In the above evaluation metrics, \( I(\cdot) \) is an indicator function where when the condition is true, the value is 1; otherwise, it is 0. \( N_k \) is a constant that denotes the maximum value of NDCG@K.

### D. THE EXPERIMENTAL RESULTS COMPARED WITH THE BASELINE

Table 2 and table 3 illustrate the experimental results of all compared models on the two datasets. All models were run four times and we use the average values as the results. The following can be concluded from the experimental results:

Firstly, it can be seen that the neural network obviously improves the effect, and the models based on neural networks are better than NMF because NMF does not model the basket and is far inferior to neural networks in complex learning.

Secondly, among the models based on neural networks, several models are better than the HRM because the HRM network is too simple and only uses the sequence information of two adjacent baskets, which is weaker than the other deep learning models in feature representation. In addition, here it is indicate that a reasonable network design can indeed help the model learn more information to improve the effect of the model.

Thirdly, our model has achieved the best results on the two data sets. This shows that designing the properties of the perception of the multilayer attention model can better focus...
on the inside of the basket to recommend item information to build the next basket and the order of the items in the different baskets, can help to recommend the next basket.

E. THE INFLUENCE OF COMPONENTS

To evaluate the contribution of each component module to the network, we remove a module from the proposed network to evaluate the recommendation effect of the remaining network. This study conducted a series of contrast experiments utilizing several neural network models:

- attention: The second attention module in sequence modeling is removed.
- selfAttention: Average pooling replace the self-attention module in basket modeling.
- globalEmbed: Remove the effect of the current prediction on the subsequent prediction in the output layer.
- singleHead: The basket is modeled using the single-headed self-attention module, that is the number of multi-head is set to 1.
- normalAttention: Use the normal attention to model the basket embedding vector. The equations are as follows:

\[ \alpha_i = \frac{\exp(\omega e_i)}{\sum_{j=1}^{L} \exp(\omega e_j)}, \]  
\[ v_{\text{basket}} = \sum_{i=1}^{L} \alpha_i e_i. \]  

where \( \omega \) is the weight parameter, and \( e_i \) is the item embedding in the basket, \( L \) is the number of items in the basket.

Table 4 and table 5 respectively show the effects of each component model on the two datasets, and the following conclusions can be drawn:

Firstly, the complete network obtains the best performance, which indicates that each module of the network we designed has played a role in the network to a certain extent. In the self-attention ablation study, we compared self-attention with normal attention. In terms of effect, the latter is slightly better than the former.

Secondly, in the JingDong dataset, the complete network is about 11.4% more efficient than the -globalEmbed network regarding their F1, it improves the HR by 0.8 percent and it increased the NDCG by 8.4 percent. This shows that it is effective to consider the effect of the recommended item on the subsequent recommended items. Regarding the F1 and NDCG metrics, self-attention showed the most obvious improvement effect, increasing these metrics by 14.6 percent and 25.5 percent, respectively. This shows that self-attention is effective in basket modeling. In the Ta-Feng dataset, the complete network improved the NDCG by up to 5 percent compared to the attention network and at least 0.8 percent compared to the singleHead network. Meanwhile, in terms of the Hit-rate and NDCG, the performance is improved by up to 21.9 and 5 percent points compared to -attention network respectively. Therefore, applying the second attention layer to different baskets in the basket sequence also achieved good results.

F. THE INFLUENCE OF HYPER-PARAMETERS

In the loss function, we use k and q to balance the effects of positive and negative samples in Equation (17), respectively. Figure 4 shows the influence of different k to q ratios on the
F1. We can see that the performance of the model tends to be stable when the ratio of k to q set to 500. This shows that the ratio of k to q helps to balance the sample imbalance.

G. THE STATISTICAL TESTING

In order to make a comprehensive comparison of the performances of different learning algorithms, it is not enough to rely on applying the models to measure the performance on data sets. We need to use hypothesis tests, which provides an important reference for our comparison of the learning algorithms. We choose the Friedman test, which is a non-parametric test. We draw the Friedman test graphs for the four metrics in Figure 5. In the graph, if there is no overlapping area between the algorithms, it is proved that algorithms have obvious differences. It can be seen that the results of the hypothesis tests are basically consistent with our experimental results.

VI. CONCLUSION

In this article, a novel Next basket Recommendation Model Based on Attribute-Aware Multi-level Attention was proposed. This model uses a self-attention mechanism to encoding the basket representation, and incorporates the attribute information of items. In the aspect of modeling the long-term and short-term preferences of users, the model is use of bi-LSTM. Then, the hidden state representation of the encoder is decoded by attention and LSTM to make recommendations. Experiments on two public data sets show that the proposed model are superior to traditional and existing deep learning-based recommendation models.

In the next basket recommendation area, there is still a lot of room for improvement. Although existing deep learning-based recommendation models have achieved better results than traditional machine learning models, there is still a lot of research to be done to improve the recommendation effect and promote the commercial application of the research results. At present, there are still some problems in the field of personalized recommendation, such as the lack of diversification. In addition, in order to be realized in commercial applications, more item attribute information and user attribute information must be integrated; thus, personalized recommendation based on knowledge graphs is an important research direction.

REFERENCES

[1] S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system: A survey and new perspectives,” ACM Comput. Surv., vol. 52, no. 1, pp. 1–38, Feb. 2019.
[2] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, “A review on deep learning for recommender systems: Challenges and remedies,” Artif. Intell. Rev., vol. 52, no. 1, pp. 1–37, Jun. 2019.
[3] S. Wang, L. Hu, Y. Wang, L. Cao, Q. Z. Sheng, and M. Orgun, “Sequential recommender systems: Challenges, progress and prospects,” in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 6332–6338.
[4] S. Wang, L. Hu, L. Cao, X. Huang, D. Lian, and W. Liu, “Attention-based transactional context embedding for next-item recommendation,” in Proc. Nat. Conf. Artif. Intell., 2018, pp. 2532–2539.
[5] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, “Session-based recommendation with graph neural networks,” in Proc. Nat. Conf. Artif. Intell., vol. 33, no. 1, pp. 346–353, 2019.
[6] A. Gatzioura and M. Sanchez-Marre, “A case-based recommendation approach for market basket data,” IEEE Intell. Syst., vol. 30, no. 1, pp. 20–27, Jan. 2015.
[7] J. Tang, F. Belletti, S. Jain, M. Chen, A. Beutel, C. Xu, and E. H. Chi, “Towards neural mixture recommender for long range dependent user sequences,” in Proc. World Wide Web Conf. (WWW), 2019, pp. 1782–1793.
[8] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, “Session-based recommendations with recurrent neural networks,” 2015, arXiv:1511.06939. [Online]. Available: http://arxiv.org/abs/1511.06939
[9] H. Ying, F. Zhuang, F. Zhang, Y. Liu, G. Xu, X. Xie, H. Xiong, and J. Wu, “Sequential recommender system based on hierarchical attention network,” in Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI), Jul. 2018, pp. 3926–3932.
[10] B. Liu and I. Lane, “Attention-based recurrent neural network models for joint intent detection and slot filling,” 2016, arXiv:1609.01454. [Online]. Available: http://arxiv.org/abs/1609.01454
[11] P. Wang, J. Guo, Y. Lan, J. Xu, S. W. and X. Cheng, “Learning hierarchical representation model for nextbasket recommendation,” in Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retre., Aug. 2015, pp. 403–412.
[12] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan, “A dynamic recurrent model for next basket recommendation,” in Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retre. (SIGIR), 2016, pp. 729–732.
[13] T. Bai, J.-Y. Nie, W. X. Zhao, Y. Zhu, P. Du, and J.-R. Wen, “An attribute-aware neural attentive model for next basket recommendation,” in Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retre., Jun. 2018, pp. 1201–1204.
[14] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, and M. Wang, “A neural influence diffusion model for social recommendation,” in Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retre., Jul. 2019, pp. 235–244.
[15] H. Liu, Y. Wang, Q. Peng, F. Wu, L. Gan, L. Pan, and P. Jiao, “Hybrid neural recommendation with joint deep representation learning of ratings and reviews,” Neurocomputing, vol. 374, pp. 77–85, Jan. 2020.
[16] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009.
[17] X. He, L. Xiao, Z. Zhang, Y. Nie, X. Hu, and T. S. Chua, “Neural collaborative filtering,” in Proc. 26th Int. World Wide Web (WWW), Apr. 2017, pp. 173–182.
[18] C.-N. Hsu, H.-H. Chang, and H.-S. Huang, “Mining skewed and sparse transaction data for personalized shopping recommendation,” Mach. Learn., vol. 57, nos. 1–2, pp. 35–59, Oct. 2004.
[19] E. Lazcorreta, F. Botella, and A. Fernández-Caballero, “Towards personalized recommendation by two-step modified a priori data mining algorithm,” Expert Syst. Appl., vol. 35, no. 3, pp. 1422–1429, Oct. 2008.
[20] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, “Factorizing personalized Markov chains for next-basket recommendation,” in Proc. 19th Int. Conf. World Wide Web (WWW), Raleigh, NC, USA, 2010, pp. 811–820.
[21] Z. H. Deng, L. Huang, C. D. Wang, J. H. Lai, and S. Y. Philip, “DEEPFC: A unified framework of representation learning and matching function learning in recommender system,” in Proc. Nat. Conf. Artif. Intell., vol. 33, 2019, pp. 61–68.
[22] Z. Yu, J. Lian, A. Mahmoody, G. Liu, and X. Xie, “Adaptive user modeling with long and short-term preferences for personalized recommendation,” in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 4213–4219.
[23] A. Vaswani, N. Shazeer, and N. Parmar, “Attention is all you need,” in Proc. Neural Inf. Process. Syst., 2017, pp. 5998–6008.
[24] I. Parmar, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Proc. Neural Inf. Process. Syst., 2014, pp. 3104–3112.
[25] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1724–1734.
[26] H. Sepp and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[27] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM networks,” in Proc. IEEE Int. Joint Conf. Neural Netw., 2005, pp. 2047–2052.
[28] F. Wang and D. M. J. Tax, “Survey on the attention based RNN model and its applications in computer vision,” 2016, arXiv:1601.06823. [Online]. Available: http://arxiv.org/abs/1601.06823
[29] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, “Show, attend and tell: Neural image caption generation with visual attention,” in Proc. Int. Conf. Mach. Learn., 2015, pp. 2048–2057.
[30] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2014, arXiv:1409.0473. [Online]. Available: http://arxiv.org/abs/1409.0473
[31] M.-T. Luong, H. Pham, and C. D. Manning, “Effective approaches to attention-based neural machine translation,” 2015, arXiv:1508.04025. [Online]. Available: http://arxiv.org/abs/1508.04025
[32] D. Lee and S. Seung, “Algorithms for non-negative matrix factorization,” in Proc. Neural Inf. Process. Syst., 2000, pp. 556–562.
[33] B. Che, P. Zhao, J. Fang, L. Zhao, V. S. Sheng, and Z. Cui, “Inter-basket and intra-basket adaptive attention network for next basket recommendation,” IEEE Access, vol. 7, pp. 80644–80650, 2019.
[34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, “PyTorch: An imperative style, high-performance deep learning library,” in Proc. Neural Inf. Process. Syst., 2019, pp. 8026–8037.

TONG LIU received the Ph.D. degree in computer software and theory from the Shandong University of Science and Technology, Qingdao, China, in 2012. She is currently an Associate Professor with the College of Computer Science and Engineering, Shandong University of Science and Technology. Her current research interests include natural language processing, machine learning, and social network analysis.

XIANRUI YIN is currently pursuing the master’s degree with the College of Computer Science and Engineering, Shandong University of Science and Technology. His current research interests include deep learning and recommendation systems.

WEIJIAN NI received the Ph.D. degree in computer application technology from Nankai University, Tianjin, China, in 2008. From 2008 to 2010, he did research work with the Business Search Department, Baidu Corporation. From 2015 to 2016, he was a Visiting Scholar with the Department of Computer Science, State University of New York at Binghamton, USA. He is currently an Associate Professor with the College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China. His current research interests include machine learning with feature interactions, recommendation systems for structural items, and business process optimization. He has published research papers in several major conferences and journals, such as CIKM, PRICAI, and KAIS.