Next Basket Recommendation with Neural Networks

Shengxian Wan, Yanyan Lan, Pengfei Wang, Jiafeng Guo, Jun Xu, Xueqi Cheng
wanshengxian@software.ict.ac.cn, lanyanyan@ict.ac.cn
wangpengfei@software.ict.ac.cn, {guojiafeng, junxu, cxq}@ict.ac.cn

ABSTRACT

One crucial task in recommendation is to predict what a user will buy next given her shopping history. In this paper, we propose a novel neural network to complete this task. The model consists of an embedding layer, a hidden layer and an output layer. Firstly, the distributed representations of the user and the items bought before are obtained and used to form a feature vector by the embedding layer. Then the hidden layer transforms the feature vector to another space by a non-linear operator. Finally, the softmax operator is adopted to output the probabilities of next items. We can see that the model elegantly involves both the user's general interest and the sequential dependencies between items for prediction. Experimental results on two real datasets prove the effectiveness of our model.

1. INTRODUCTION

Next basket recommendation (NBP) is important in applications such as online electronic business and retail shopping market. Since a user's shopping history is typically represented as a sequence of baskets, it is natural to formalize this task to a sequential prediction problem. That is, given a sequence of baskets, we want to predict what the user will buy next. In this process, both the user's general interest (what items the user likes) and the sequential dependencies between items (the influence of items bought before to items in the next basket) are important for the prediction.

Existing approaches for this task are mainly based on matrix/tensor factorization and markov chains [5, 6]. In this paper, we propose to use neural networks to complete this task. The reason is that neural networks have been successfully applied to sequential prediction problems such as language model [2, 3] and click through rate prediction [7]. Besides, neural networks can learn richer representations than matrix factorization, and are more flexible and powerful in modelling complicated relationships [1].

Our model consists of three layers: embedding, hidden and output layer. The embedding layer firstly maps the user and the items in the user's shopping history to distributed representations, and then concatenates them together to obtain a feature vector. The hidden layer transforms the feature vector to another space non-linearly. Finally, the output layer gives the probabilities of next items by the softmax operator. From the above process, we can see that the two crucial factors of the user's general interest and the sequential dependencies between items are both elegantly involved in this model. Empirically, our experimental results on two real datasets show that our model significantly outperforms state-of-the-art methods.

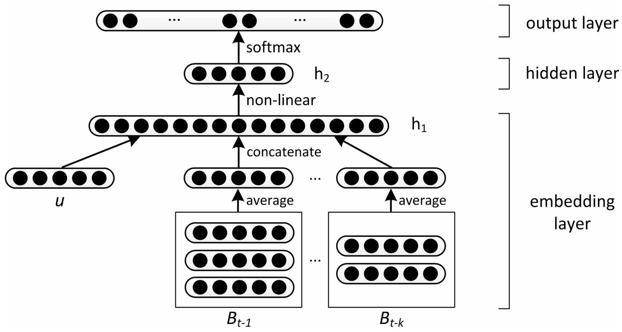

Figure 1: NN-Rec model

Our work is inspired by [2], where a neural network is proposed to use the local context to predict the next word in language model. However, [2] only models the dependences between next items and nearest items, while ignores the user's general interest if we apply it directly for NBP. The main novelty of our model lies in that the representation of the user is further introduced to model the influence of the user's general interest. In this way the model can well fit the characteristics of NBP. To the best of our knowledge, this is the first work to introduce neural networks to NBP.

2. OUR APPROACH

Next basket recommendation can be formalized as a sequential prediction problem, here we follow the notations of [5]. Let $\mathbb{U} = \{u_1, u_2, ..., u_{|\mathbb{U}|}\}$ be the user set and $\mathbb{I} = \{i_1, i_2, ..., i_{|\mathbb{I}|}\}$ be the item set. For each user u, there is a observed buying behavior sequence $S_u = (B_u^1, B_u^2, ..., B_u^{t-1})$, where B_u^t is a set of items which are bought by user u at time t. The sequential prediction problem is to predict B_u^t for each user u, given S_u.

Fig.1 shows our proposed model, namely NN-Rec. It is a three-layer neural network:

(1) The first layer is the embedding layer. The inputs are the user ID and the item IDs in the user's last k baskets. We first transform the inputs to distributed representations, where each user is represented as a vector $u \in \mathbb{R}^{d_u}$, and each item is represented as $v \in \mathbb{R}^{d_i}$. We obtain user matrix $U \in \mathbb{R}^{d_u * |\mathbb{U}|}$ and item matrix $V \in \mathbb{R}^{d_i * |\mathbb{I}|}$ by putting all user and item vectors together, respectively. Both U and V are learned during the training process. Then, each basket is represented as the mean of all items included in it. The output of the embedding layer is to concatenate the user's and the baskets' representations together to obtain a feature vector $h_1 \in \mathbb{R}^{d_u + k * d_i}$, which can be viewed as the representation of both the user's general interest and local context.

(2) The second layer is a non-linear hidden layer, which

Table 1: Datasets

Dataset	# of users	# of items	# of baskets
Tafeng	9,238	7,973	77,202
Beiren	13,736	5,920	522,963

transforms h_1 to a hidden representation h_2 with dimension l. Here we use *tanh* as the activation function, which is commonly used in neural networks. h_2 is obtained as follows.

$$h_2 = tanh(W_1 h_1 + b_1),$$

where $W_1 \in \mathbb{R}^{l*|h_1|}, b_1 \in \mathbb{R}^{l*1}$ are parameters to be learned.
(3) The output layer is a *softmax* layer, which outputs the probabilities of next items,

$$s = W_2 h_2 + b_2,\ P(i_j \in B_t|u, B_{t-1}, ..., B_{t-k}) = \frac{e^{s_{i_j}}}{\sum_{p=1}^{|\mathbb{I}|} e^{s_{i_p}}}$$

where $W_2 \in \mathbb{R}^{|\mathbb{I}|*l}, b_2 \in \mathbb{R}^{|\mathbb{I}|*1}$ are parameters to be learned.
For training, we use negative log-likelihood as the loss function and stochastic gradient decent with back propagation for optimization. *Weight decay* is also used as regularization, therefore the optimization problem becomes

$$\arg\min_{U,V,W_1,W_2,b_1,b_2} -\sum_u \sum_t \sum_{i \in B_t} logP(i \in B_t|u, B_{t-1}, ..., B_{t-k})$$
$$+ \lambda_U ||U||_2^2 + \lambda_V ||V||_2^2 + \lambda_1 ||W_1||_2^2 + \lambda_2 ||W_2||_2^2$$

NN-Rec has some superiorities compared to previous state-of-the-art methods such as FPMC. Compared to FPMC which is based on tensor factorization, NN-Rec is neural network based and is more flexible and powerful. Firstly, the model can easily capture longer dependencies by varying the window size k of the embedding layer, while FPMC only captures the influence of the nearest one basket. Secondly, the embedding layer is flexible and we can add other features such as user profiles and item attributes to this layer without modifying the model's framework. Finally, the hidden layer gives the power to model more complicated interactions between user and items.

3. EXPERIMENTS

We conduct experiments on two real retail datasets (*Tafeng* and *Beiren*) to evaluate the effectiveness of our model. *Tafeng* is a benchmark dataset released by *RecSys*[1]. While *Beiren* is collected by a large retail store in China, which contains users' shopping history during 2011 to 2013. We preprocess the two datasets by removing the items bought less than 10 times and the users who bought less than 10 items or 4 baskets. The detailed information of the obtained datasets are shown in Table 1. For each user, we hold out the last basket as the test set and keep other data as the training set. The last basket of all users in the training set are used for validation. The final models are trained on the whole training set. For our model in all experiments, the dimensions of h_2 (i.e., l), user vector (i.e., d_u) and item vector (i.e., d_i) are set to the same value, denoted as d. k is set to 1 and 2 since that most users only have a few baskets in the two datasets.

We recommend top c items for each user, denoted as $\hat{B}_u^t$, and use F1-measure for evaluation. c is set to 5 in Fig. 2.

$$F1 = \frac{2pr}{p+r}, p = \frac{\sum_u |B_u^t \cap \hat{B}_u^t|}{|U| * c}, r = \frac{\sum_u |B_u^t \cap \hat{B}_u^t|}{\sum_u |B_u^t|}$$

We compare our model with several existing methods, such as Top popular (TOP) which recommend items according to global popularity, NMF, BPR [4] and FPMC [5]. Notice that, when $k = 1$, our model uses the same information as FPMC. From the results (Fig. 2), we can see that

[1] http://recsyswiki.com/wiki/Grocery_shopping_datasets

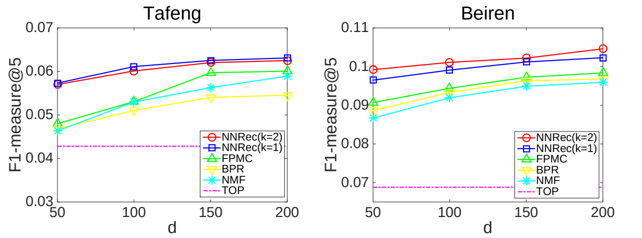

Figure 2: Experimental results on two datasets, d is the feature dimension

the performances of our model are consistently better than all baselines on both datasets, whenever $k = 1$ or 2. Therefore, we can conclude that neural networks are suitable for this task. Furthermore, on *Beiren* the performance when $k = 2$ is consistently better than $k = 1$. This indicates that longer dependences between items can be captured by NN-Rec and are important for this task. However, on *Tafeng*, larger k does not show benefits. This is because *Tafeng* is much smaller than *Beiren*, as shown in Table 1, and thus complex models are more easily to overfit the training set. Therefore, we recommend to choose appropriate k according to the scale of data.

4. CONCLUSIONS

In this paper, we propose to use neural networks for next basket recommendation. Our model consists three layers and can elegantly incorporate both a user's general interest and sequential dependencies between items for recommendation. Experimental results show that our model significantly outperforms existing approaches. To the best of our knowledge, this is the first time to introduce neural networks to this task. We believe neural network is a more flexible framework to model complicated interactions and will impact this area further. In future work we will investigate other neural networks such as recurrent neural network which can capture longer term sequential dependencies and try to incorporate other information such as time serials into the model.

5. ACKNOWLEDGMENTS

This research work was funded by the 973 Program of China under Grants No. 2014CB340401, No. 2012CB316303, the 863 Program of China under Grants No. 2014AA015204, the National Natural Science Foundation of China under Grant No. 61232010, No. 61425016, No. 61173064, No. 61472401, No. 61203298, and No. 61202214.

6. REFERENCES

[1] Y. Bengio. Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127, Jan. 2009.

[2] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A Neural Probabilistic Language Model. *JMLR*, 3, 2003.

[3] T. Mikolov, M. Karafiát, L. Burget, J. Cernocky, and S. Khudanpur. Recurrent neural network based language model. In *INTERSPEECH*, pages 1045–1048, 2010.

[4] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, pages 452–461, 2009.

[5] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *WWW*, pages 811–820. ACM, 2010.

[6] X. Wu, Q. Liu, E. Chen, L. He, J. Lv, C. Cao, and G. Hu. Personalized next-song recommendation in online karaokes. In *RecSys*, pages 137–140. ACM, 2013.

[7] Y. Zhang, H. Dai, C. Xu, J. Feng, T. Wang, J. Bian, B. Wang, and T. Liu. Sequential click prediction for sponsored search with recurrent neural networks. In *AAAI*, pages 1369–1375, 2014.