Modeling Retail Transaction Data for Personalized Shopping Recommendation

Pengfei Wang,Jiangfeng Guo,Yanyan Lan,Xueqi Cheng
Key lab of Network Data Science and Technology in ICT
Beijing, China
wangpengfei@software.ict.ac.cn,{guojiafeng,lanyanyan},cxq@ict.ac.cn

ABSTRACT

Retail transaction data conveys rich preference information on brands and goods from customers. How to mine the transaction data to provide personalized recommendation to customers becomes a critical task for retailers. Previous recommendation methods either focus on the user-product matrix and ignore the transactions, or only use the partial information of transactions, leading to inferior performance in recommendation. Inspired by association rule mining, we introduce association pattern as a basic unit to capture the correlation between products from both intra- and inter-transactions. A Probabilistic model over the Association Patterns (PAP for short) is then employed to learn the potential shopping interests and also to provide personalized recommendations. Experimental results on two real world retail data sets show that our proposed method can outperform the state-of-the-art recommendation methods.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining

Keywords

recommendation, association pattern, probabilistic model

1. INTRODUCTION

Massive transaction data sets have been routinely recorded in offline retails, which convey rich preference information on brands and goods from customers. It becomes a critical task for retailers to mine these valuable transaction data to provide personalized recommendation to customers, so that they can stimulate consumption and compete with e-commerce business where recommender systems have already been widely employed. In fact, in the past decades data mining technologies, like association rule mining, have been applied on transaction data to discover interesting relations between products. For example, some useful rules are found in the sales data of a supermarket indicating that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat. The obtained rules can help retailers in promotional pricing or product placements, but usually lack personalization which is important to customers. Therefore, how to model the transaction data to provide personalized shopping recommendations becomes a key challenge.

Previous methods on recommender systems, like collaborative filtering methods [3], mainly focus on directly modeling the user-product matrix. In this way, the transaction information is usually ignored in these methods. However, transactions indicate the strong correlation between products and are very prominent in retail as compared with the correlation in e-commerce. For example, we compared the sales data from a large retailer BeiRen with data from the biggest e-commerce website Taobao in China. We found that in the retail data set there is 33.6% transactions containing more than two products, while in the e-commerce data set there is less than 12%. Some recent work on basket recommendation did take transactions into account . However, in their work, only partial transaction information: either patterns across transactions [4], or patterns within transactions [7] has been utilized.

In this paper, we present a novel approach on modeling retail transaction data for personalized shopping recommendation. Inspired by association rules, we introduce association patterns as basic units to capture the correlation between products and summarize the dataset. Here an association pattern is defined as a weighted pair of products from either intra- and inter- transactions of a user. The weight of a pattern describes correlations strength between the two products in the pattern, which is defined according to the time span between the two products. In this way, the original transaction data can been turned into a collection of association patterns, which preserves the important correlation information within and across transactions.

By assuming the association patterns are generated from some low-dimensional latent shopping interests, we propose a Probabilistic model over the Association Patterns (PAP for short) to model the generation process and learn the representation of shopping interests. With the learned model, we can then inference the shopping interests of each individual and provide personalized shopping recommendations. Experimental results on two real world retail datasets show that our proposed method can outperform the state-of-the-art recommendation methods.

The remainder of this paper is organized as follows. We first discuss related work in section 2. Section 3 introduce

CIKM'14, November 3–7, 2014, Shanghai, China.

http://dx.doi.org/10.1145/2661829.2662020 .

our model. We present our experiments and discussion in section 4, then make a conclusion in section 5.

2. RELATED WORK

In this section, we will briefly review the related work on general recommender systems, and the work on transaction data mining for recommendations.

The content-based method recommends products based on a comparison between contents of products [1]. However, content information may not always available in many cases, which limits the usage of this model. Collaborative filtering is a popular recommendation method, which tries to predict the utility of products for a particular user based on the products rated by other users [1, 5]. Algorithms of collaborative filtering essentially can be grouped into two general classes: memory-based methods, e.g. KNN, and model-based methods, e.g. matrix factorization model.

KNN makes prediction based on the entire collection of previous rated products [6]. The rate to an unknown product for a given user can be calculated as an aggregation of similar users' behaviors. However this kind of algorithm only concerns local information in choosing similar users, and it leads to a inferior performance in recommendation.

Matrix factorization models map both products and users into a low-dimensional latent space. For example, BPR (Bayesian Personalized Ranking) is a popular factorization model[5, 8]. The model tries to obtain the representation of users and products by optimizing a personalized ranking for all products.

Recently, some work in recommendation also take transaction information into account. For example, the rule-based models utilize data mining algorithms(i.e. Apriori and FP-growth)to recommend products by mining frequent itemsets from dataset. However the models tend to generate a large number of patterns once mining a large data set, most of which are spurious, not relevant to recommendations [4]. Rendle et al. [6] propose a factorization model by emphasizing correlations of products belonging to different transactions, while Xiang Wu et al. [7] utilize relations of products in the same transaction to recommendation songs for users.

3. OUR APPROACH

In this section we will describe our Probabilistic model over Association Pattern in detail. In the following, we first introduce the concept of association patterns, then describe the generative model (i.e. PAP) over the association patterns, we finally present how to infer customers' shopping interests with the learned model and provide personalized recommendations.

3.1 Association Pattern

Correlation between products is a basic factor in shopping behavior and critical for recommendation, which can be revealed by the co-occurrences between products. We introduce the concept of association patterns to describe the co-occurrences between products in user transaction data. According to different co-occurrence types, there are two kinds of association patterns, namely Intra Association Pattern and Inter Association Pattern, as shown in Figure 1.

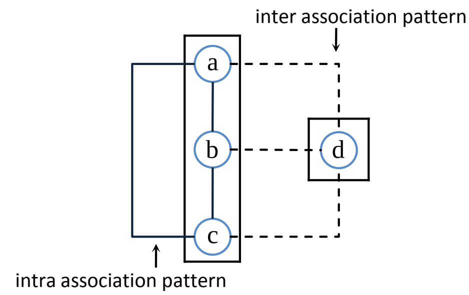


Figure 1: Association patterns between two transactions. The solid lines among product a, b, c stand for intra association patterns, and the dotted lines among a, b, c and d stand for inter association patterns.

Definition 1. ***Intra Association Pattern.*** *Given the transaction set* $T^u = \{t_1^u, t_2^u, \ldots, t_{|T|}^u\}$ *of user u, where* $|T|$ *is the count of transactions belonging to u, an intra association pattern is defined as a weighted pair of products* $< I_i, I_j, w_{ij} >$, *where* $I_i, I_j \in t_m^u$ *and* w_{ij} *denotes the weight of the pattern.*

Definition 2. ***Inter Association Pattern.*** *Given the transaction set* $T^u = \{t_1^u, t_2^u, \ldots, t_{|T|}^u\}$ *of user u, where* $|T|$ *is the count of transactions belonging to u, an inter association pattern is defined as a weighted pair of products* $< I_i, I_j, w_{ij} >$, *where* $I_i \in t_m^u$, $I_j \in t_n^u$, $m \neq n$, *and* w_{ij} *denotes the weight of the pattern.*

The weight of an association pattern represents the correlation strength between the pair of products. Obviously, the intra association pattern indicates strong correlations between products since customers prefer to buy them together, while the inter association pattern indicates weak correlation. To reflect this, here we define the weight based on the time-stamp of the transaction the products belong to. Let $I_i \in t_m^u$ and $I_j \in t_n^u$, the weight of an association pattern is defined in following unified form:

$$w_{ij} = \exp\{-\frac{|time(t_m^u) - time(t_n^u)|}{K}\} \tag{1}$$

where K is a normalization factor. As we can see, the weight of an intra association pattern is 1 by definition, while that of an inter association pattern is less than 1 and decayed with the time span between the two transactions.

Based on the definition of association patterns, the original transaction data set can be turned into a collection of association patterns. By this transformation, we can accumulate all the important co-occurrence information within and across transactions from different users.

3.2 The Generative Model

Although the association patterns reflect the correlations between products, they can be hardly directly utilized for recommendation due to the sparsity. To learn useful correlation information for recommendation, we assume that all the association patterns are generated from some latent low-dimension shopping interests. In this way, we introduce a generative model, namely PAP, to describe the generation of the association patterns and learn the latent shopping interests.

Specifically, PAP assumes that two products in an association pattern are drawn independently from a latent shopping interest. The key idea is that if two products co-occur more frequently, they are more likely to belong to the same interest.

Formally, let S denotes the whole collection of association patterns, S shares a n-dimensional latent shopping interests. Θ denotes a multinomial distribution of shopping interests, with $\Theta_k = p(z = k)$ standing for the proportion of the k-th shopping interest. Φ_k denotes a multinomial distribution of products, with $\Phi_{k,m}$ standing for the proportion of product I_m on the k-th shopping interest($\sum_m \Phi_{k,m} = 1$). The generative process of PAP is described as follows:

Algorithm 1 The generative process of PAP

1: sample a distribution of shopping interests $\Theta \sim$ Dirichlet(α)
2: for each shopping interest z
 draw a distribution $\Phi_z \sim$ Dirichlet(β)
3: for each pattern $< I_i, I_j > \in S$
 draw a latent shopping interest $z \sim$ Multinomial(Θ)
 draw a pattern $< I_i, I_j > \sim$ Multinomial(Φ_z)

where parameter α and β are Dirichlet priors. Figure 2 shows the probability graph of generative process.

Based on the generative process mentioned above, we can obtain the joint probability of pattern $< I_i, I_j >$:

$$\begin{aligned} P(< I_i, I_j > |\Theta, \Phi) &= \sum_z P(z)P(I_i|z)P(I_j|z) \\ &= \sum_k \theta_k \Phi_{k,i} \Phi_{k,j} \end{aligned}$$

the marginal distribution of $< I_i, I_j >$ can be calculated through integrated Θ and Φ:

$$P(< I_i, I_j > |\alpha, \beta) = \iint \sum_k \theta_k \Phi_{k,i} \Phi_{k,j} d\Theta d\Phi$$

and the likelihood of the whole pattern-set S is:

$$P(S|\alpha, \beta) = \prod_{< I_i, I_j > \in S} \iint \sum_k \theta_k \Phi_{k,i} \Phi_{k,j} d\Theta d\Phi$$

We use Gibbs sampling to approximate inference. In our model there are three parameters need to be estimated: z, Θ, and Φ. Concerning that we can integrate out parameters Θ, Φ because of conjugate prior α, β. Given association pattern $< I_i, I_j >$, we just need to sample parameter z:

$$P(z = k|\mathbf{z}_{-< I_i, I_j >}, S, \alpha, \beta) \propto (n_k + \alpha) \frac{(n_{k,i} + \beta)(n_{k,j} + \beta)}{(\Sigma_m n_{k,m} + M\beta)^2}$$

where $\mathbf{z}_{-< I_i, I_j >}$ denotes the interest assignments for all patterns, except $< I_i, I_j >$. Θ_k, $\Phi_{k,m}$ can be calculated as:

$$\Theta_k = \frac{n_k + \alpha}{|S| + K\alpha}$$

$$\Phi_{k,m} = \frac{n_{k,m}}{\sum_m n_{k,m} + M\beta}$$

where n_k is the number of pattern $< I_i, I_j >$ assigned to the k-th interest, $n_{k,i}$ is the number of I_i assigned to the k-th interest, and $|S|$ is the number of patterns in pattern-set S.

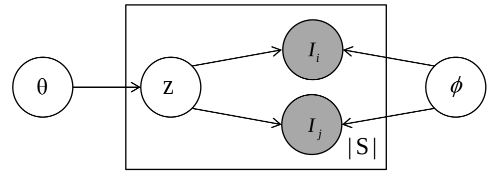


Figure 2: probabilistic model over association patterns

3.3 Inference of User Preference

With the learned shopping interests, we now aim to infer individual shopping interests for each user for recommendation. Through Bayesian theory, given association pattern $< I_i, I_j >$, we can get the probability of the k-th shopping interest:

$$\begin{aligned} P(z = k| < I_i, I_j >) &= \frac{P(< I_i, I_j > |z = k)P(z = k)}{\sum_z P(< I_i, I_j > |z)P(z)} \\ &= \frac{P(z = k)P(I_i|z = k)P(I_j|z = k)}{\sum_z P(I_i|z)P(I_j|z)P(z)} \\ &= \frac{\theta_k \phi_{k,i} \phi_{k,j}}{\sum_k \theta_k \phi_{k,i} \phi_{k,j}} \end{aligned}$$

Then the proportion on k-th shopping interest for user u can be calculated as:

$$\begin{aligned} \theta_k^u &= P(z = k|u) \\ &= \sum_{< I_i, I_j > \in S^u} P(z = k| < I_i, I_j >)P(< I_i, I_j > |u) \end{aligned}$$

where S^u denotes the collection of association patterns mined from transaction set T^u, and $P(< I_i, I_j > |u)$ can be obtained as follows:

$$P(< I_i, I_j > |u) = \frac{w_{ij}}{\sum_{< I_i, I_j > \in S^u} w_{ij}}$$

To conduct personalized shopping recommendation to users, we calculate users' preference to products with respect to their shopping interests as follows:

$$P(I_i|u) = \sum_z P(I_i|z)P(z|u) = \sum_k \theta_k^u \phi_{k,i}$$

By sorting the products according to $P(I_i|u)$, we can recommend top-k products to uesrs.

4. EVALUATION

To demonstrate the effectiveness of our model, we choose two real retail data sets: BeiRen dateset and Tafeng dataset. BeiRen dataset is collected by a large retail department store in China, recording brands of merchandise products from 2011 to 2013. Tafeng dataset[1] is offered by RecSys, which covers products from food, office supplies to furniture. The detail is showed in Table 1.

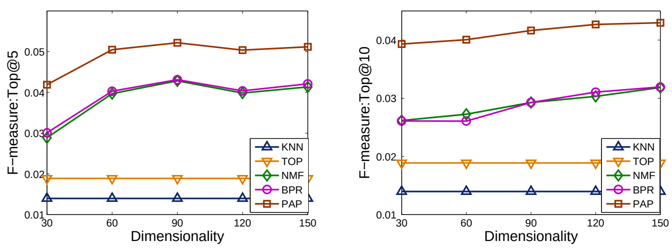
First we preprocess two datasets before evaluation. We reserve products and brands in datasets bought at least 10 times. We hold out 50% of the data set for training, with the remaining for test, The time of transactions in two datasets is recorded by the day, thus we assign $K = 365$ in Equation 1. We evaluate our model against four state-of-the-art methods in product recommendation:TOP(the most

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

Table 1: data set statistics

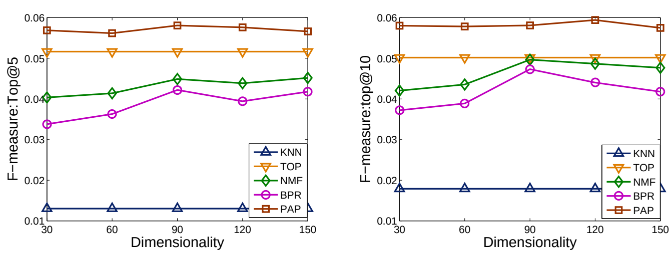
id	name	# users	# products	# transactions
1	$BeiRen$	18315	1442	242894
2	$Tafeng$	7141	6894	37269

popular products are recommended),KNN, NMF, and BPR method.
We compare the performance of different recommendation methods with the widely used F-measure [2, 6, 7] over top-5, top-10 products.



(a) (b)

Figure 3: Comparisons of TOP, KNN, NMF, BPR and our model PAP on BeiRen data set.



(a) (b)

Figure 4: Comparisons of TOP, KNN, NMF, BPR and our model PAP on Tafeng data set.

Figure 3 and Figure 4 show the results on BeiRen and Tafeng dataset. We can see that KNN performs worst in recommendation because it only utilizes local information. By simply recommending top popular products, Top method outperforms KNN slightly. Surprisingly, Top method performs the second best on Tafeng datset, indicating that the top method is very unstable. The BPR and NMF methods represent users and products into a low-dimensional latent space to avoid data sparsity, and the two methods show little difference in performance.

Comparing to other methods, our model outperforms all other methods, with F-score promoted at least 33% and 16% respectively. The improvement is statistically significant(p-$value < 0.01$)

5. CONCLUSIONS

In this paper we proposed a novel Probabilistic model over the Association Pattern for personalized recommendation. With a generative process to reduce association patterns into a n-dimensional latent shopping interests, we recommend user top-k products by user's preference. In the future we will try to consider sequential patterns, and make dynamic personalized recommendation.

6. ACKNOWLEDGMENTS

This research work has funded by 973 Program of China under Grants No.2014CB340401, 863 Program of China under Grants No.2014AA015204, No.2012AA011003, Project supported by the State Key Program of National Natural Science of China under Grant No.61232010, Project supported by the National Science Foundation for Young Scientists of China under Grant No.61203298, and The National Key Technology $R\&D$ Program under Grants No.2012BAH39B04.We would like to thank the anonymous reviewers for their helpful comments.

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, June 2005.

[2] K. Choi, D. Yoo, G. Kim, and Y. Suh. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electron. Commer. Rec. Appl.*, 11(4):309–317, July 2012.

[3] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, Aug. 2009.

[4] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa. Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd International Workshop on Web Information and Data Management*, WIDM '01, pages 9–15, New York, NY, USA, 2001. ACM.

[5] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.

[6] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 811–820, New York, NY, USA, 2010. ACM.

[7] X. Wu, Q. Liu, E. Chen, L. He, J. Lv, C. Cao, and G. Hu. Personalized next-song recommendation in online karaokes. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 137–140, New York, NY, USA, 2013. ACM.

[8] X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, and J. Han. Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 283–292, New York, NY, USA, 2014. ACM.