Customized Bundle Recommendation by Association Rules of Product Categories for Online Supermarkets

Yan Fang
School of Business
Dalian University of Technology
Panjin, Liaoning, 124221, China
yfang@dlut.edu.cn

Xinyue Xiao
School of Business
Dalian University of Technology
Panjin, Liaoning, 124221, China
xiaoxinyue0606@mail.dlut.edu.cn

Xiaoyu Wang
School of Business
Dalian University of Technology
Panjin, Liaoning, 124221, China
wxyu970114@mail.dlut.edu.cn

Huiqing Lan
School of Business
Dalian University of Technology
Panjin, Liaoning, 124221, China
861194659ctt@mail.dlut.edu.cn

Abstract—**A customized bundle is a list of products recommended to consumers among them they can choose his/her favorite products according to his/her preference. It is an efficient way to not only simplify the customer's shopping process, but also reduce the order fulfillment cost for the online supermarkets. A customized bundle recommendation method is proposed for online supermarket in this research. It is realized by combinational using association rule mining, customer segmentation and recommendation techniques. The association rules of product category are used to avoid mass unnecessary association rules of product. The product lists recommended within each category are generated by product ranking on each customer segmentation. Numerical experiments are conducted to verify the effectiveness of the proposed method. The method can be easily extended**

Keywords— Customized Bundle, e-Commerce, Association Rules, Customer Segmentation.

I. INTRODUCTION

The recent 10 years has witnessed a booming development of online supermarkets in China. It grows at a rate of 40% every year [1] and some typical online supermarkets with a large market share has been founded, such as DJ.com, Tmall and Suning in China. However, the online supermarkets in China are currently in a state of losing profit and facing with severe survival problems. For example, the first nationwide online supermarket in China, Yihaodian, has been acquired [2]. The JD.com has been operating at a low profit level for many years [3]. It is mainly because that the higher operational cost for order fulfillment, such as order picking, packing and delivery, reduces the profit margin sharply and consequently limits the long-term development of the online supermarket [4]. Therefore, how to increase profit and reduce operating costs in the meantime is of vital important and has become a major challenge for online supermarkets [5]. Bundling is an efficient way to solve this problem. It means to sell more than one goods or services by package at a more attractive united price compared with selling separately. By bundling goods into several packages for sale, it can simplify the order picking, packing and delivery progress, and consequently reduce the associated operating cost. On the other hand, it also simplifies the customer's shopping process for order padding to qualify for free shipping, which enables the customer to have a better shopping experience through online supermarkets.

However, due to the large scale of SKUs (Stock Keeping Unit) and online consumers' high personalization demand, the fixed bundle can hardly satisfy consumer preference. Therefore, a customized product bundling approach has been proposed in this research. As shown in figure 1, when a consumer reviewing information for some products, for example, toothpastes, a customized product bundle will be recommended. There are three type of products within the bundle, namely toothpastes, toothbrushes and laundry detergents. For each product category, more than one similar products with different brands, prices and other features are provided. The consumer can choose one of his/her favorite product in each category. Finally, a customized bundle is formulated and an overall price will be given. Under this customized bundle, the customer can obtain a level of personal preference on different products. It can reduce the time or energy spent on choosing products from a long list of similar products, and on order padding due the threshold value for free shipping which is required by most online supermarkets. Meanwhile, it is convenient for the online supermarkets to pre-optimize the storage assignment of these products in the warehouse, and consequently reduce the cost for order fulfillment.

With focus on how to generate this kind of customized product bundle, a method that combining association rules of category and recommendations is proposed. The left of this paper is organized as follows. In Section 2, literature related to our research is reviewed. The proposed research method is presented in Section 3. Reference cases for application of this method is given in Section 4. Finally, we conclude our research in Section 5.
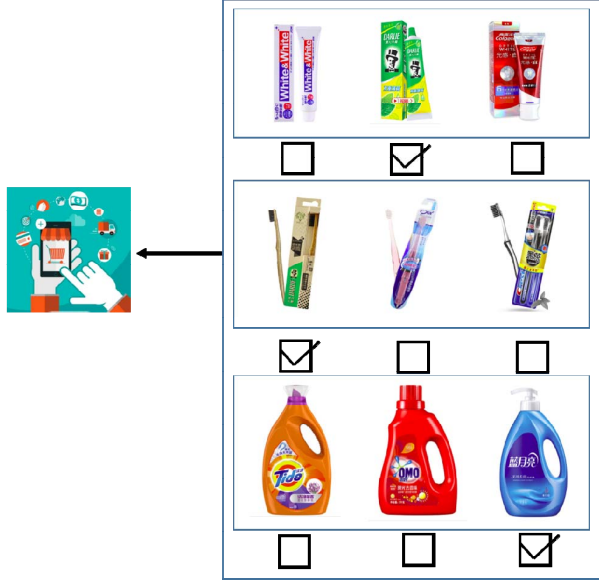
Fig. 1.   An example of customized bundle

II.   LITERATURE REVIEW

Data mining techniques are widely used for analyzing customers' preference in order to develop product bundles [6] to exact more consumer surplus. Yang and Lai [7] have investigated the method of product bundling for an online book store with different data sources. They pointed out that by integrating the customer orders, browsing data and data of shopping-cart, more interesting product bundles can be provided to online consumers. A cluster algorithm for exterminating the customers' lifestyles has been proposed by Miguéis et al. [8]. Then each customer was classified into different lifestyle cluster according to the rule of closet similarity of the customer past purchase history. The algorithms of K-means and association rules are proposed by Liao et al. [9] as a bundling solution for marketing. Liao et al. [10] have classified customers into three clusters, which are low, medium, and high according to their consumption frequency. And in order to determine the categories of product preferred by each customer group, the Apriori algorithm is used.

In conclusion, the existing data mining methods provide us with popular product combinations and consumption patterns. However, due to the large amount of SKUs in the online supermarkets, and various preferences of a large number of customers, it also suffers from a very large number of unneccessary frequent itemsets or consumption patterns. Our approach incorporates two aspects: (i) a mechanism to reduce the large number of frequent itemsets, and (2) a novel bundling pattern to balance between satisfying consumer preferences and reducing cost.

III.   THE PROPOSED METHOD

Based on association rules and recommendation technique, the method to recommend a customized bundle is proposed. The basic principles of the method is to combine association rules of product categories and personal recommendation techniques. The association rules are used to generate a serious correlated product categories for customers. For example, in figure 1, three product categories are provided for the consumer, which is based on association rules of product categories. Personal recommendation techniques is used to offer a number of products in each category instead of too many products. For example, in figure 1, there are three products available in each category.   The procedures are shown in figure 2 and explained as follows.
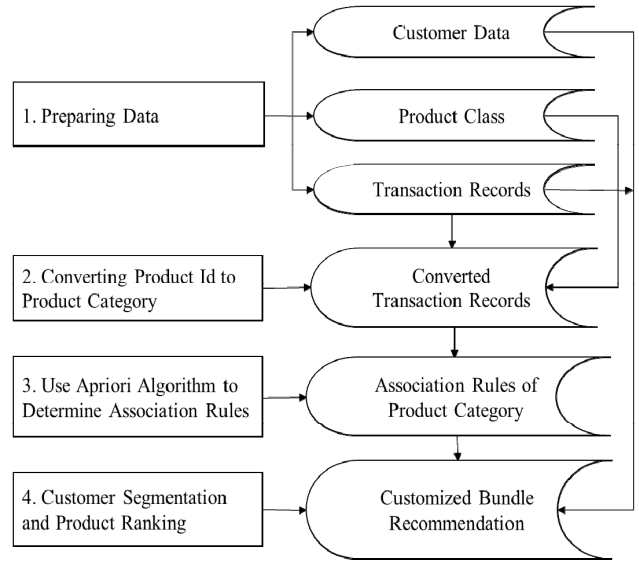

Fig. 2.   The Proposed Method

A.   Data Preparing

The data used for customized bundle is mainly from the transaction records, product class and customer data. The transaction are recorded as an order consists of serious of products bought by a customer.   The information of product class represents for a hierarchy structure of products and product class. Customer data covers basic profiles of customers. Three sources of data are prepared for further data mining and analysis.

B.   Data Converting

The second step to generate customized bundle is to convert the original transaction records consists of products into product classes. By basic techniques of databases, the converted transaction records can be obtained from data on original transaction records and product class.

C. Association Rules of Product Categories

Association rules, in the form of X ⇒ Y, show the relationships between products. Here X is the antecedent while Y is the consequent in the rule. There are two minimum thresholds values the association rules must satisfy. One is the support, and the other is confidence. The support is the ratio of the transactions including X and Y to the total transaction. It represents for the frequency of the products in the rule that be brought together occurring in all transactions. The confidence is the percentage of the transactions including X and Y to the transactions including X. The confidence is a conditional probability that indicates the implication of connections of the association rule. There are a lot of algorithms used to extract the association rules, out of which the Apriori algorithm is one of the widely used method[11].

D. Recommendation Product Choice in Each Category

There are numerous recommendation algorithms that have been adopted, including traditional collaborative filtering, cluster models, search-based methods, and item-to-item collaborative filtering, among others. In this paper, we use the sales ranking data within each product category for recommendation. However, due to distinct characteristics of customers, the top ranking product may not be interesting for each customer. Consequently, we employ the idea of customer segmentation to distinguish customer types. Based on the customer segmentation, the ranking of sales for each product category can be obtained. Finally the top *n* products are recommended for each product category according to the segmentation of each customer. Finally, the consumer can choose products among the list of customized bundle.

In this research, the RFM model (Recency, Frequency and Monetary Value) is used for customer segmentation. RFM model is one of the most widely used model to distinguish the value of customers. In RFM, R shows the recency of a customer's last purchase, F represents for the frequency of the purchases, and M signifies the monetary value of the purchases of a customer [12]. Through RFM, attributes of customers can be obtained. Then, using k-means, the segmentation of customers can be generated.

IV. EMPIRICAL ANALYSIS AND RESULTS

In this section, all the steps in each phase of the research methodology are investigated in detail. Python software is employed to implement these step. In the present research, the company selected is an electronic retailer selling foods, drinks, proteins and junk foods. The company aims to recommend customized product bundles to its customers.

A. Data Understanding

The data for this research were collected from the electronic transactions of the company with its customers. These data include 10281 customers, 251396 records, 1560 products and 110 product categories. The information used in the present study were obtained from the four databases of customers, transactions, products and products classed. The attributes of these databases are presented in Table 1.

B. Data preparation and conversion

The primary keys of the databases of transactions, products, and product classes are used to merge them and achieve an integrated converted data set. The converted data set is shown in Table 2.

TABLE 1
ATTRIBUTES OF THE DATABASES

Database	Attributes
Customers	Customer ID, Customer Name, Gender, Address, Phone number
Transactions	Invoice ID, Product code, Customer ID, Quantity, Transaction Date
Products	Product ID, Product Name, Price, Category
Product Classes	Product Class Id, Product Category, Product Department, Product Family

TABLE 2
EXAMPLE OF CONVERTED TRANSACTION RECORDS

Invoice ID	Customer Id	Product Category
1	1185	Canned Soup, Starchy Foods
2	0976	Canned Soup, Plastic Products, Dairy
3	9872	Candy,Dairy, Jams and Jellies, Cleaning Supplies
4	3542	Snack Foods, Jams and Jellies, Cleaning Supplies

C. Association rules of product category

The Apriori algorithm is applied to extract the association rules from the transactions data. The program is conducted by Python. The minimum confidence and support is set to 0.01 and 0.1 respectively. The top three association rules ranked by value of confidence related to category "Dairy" is listed in Table 3.

TABLE 3
EXAMPLE OF ASSOCIATION RULES

Antecedent	Consequent	Support	Confidence	Lift
Dairy	Hygiene	1.48%	45.87%	1.13
Dairy	Jams and Jellies	1.02%	27.21%	1.09
Dairy	Jams and Jellies, Snack foods	1.85%	26.24%	1,12

D. Customized Bundle Recommendation

After merging databases of customers and transactions, the recency, frequency, and monetary value are calculated from each customer's transactions in order to be used in clustering the customers into different segmentations. Then the values are normalized prior to being used by z-score formula. Finally the K-means clustering algorithm is used to determine each market segment. Through numerical testing, the number of clusters in K-means algorithm is set to 4. Results of the segmentation are shown in Table 4. According

to the characters of values on R, F and M, we define the four types of customer as high value customers, customers with potential purchasing power, customers who is sensitive to price, and low value customers.

TABLE 4
RESULTS OF SEGMENTATION

ID	Average Value of R	Average Value of F	Average Value of M	No. of Customers
1	0.439777	1.696267	1.751016	1366
2	0.235053	0.381812	0.346184	4208
3	0.392983	-0.840812	-0.842234	3210
4	-1.905330	-0.818785	-0.765553	1497

Then the product ranking for each category within each customer segmentation can be generated by basic operations on database. An example for the top 3 products of dairy is given in Table 5.

TABLE 5
AN EXAMPLE OF PRODUCT RANKING

Customer Segmentation ID	Product Category	Top 3 Products
1	Dairy	Gorilla 2% Milk Gorilla Whole Milk Carlson 2% Milk
2	Dairy	Gorilla 1% Milk Gorilla Buttermilk Booker 1% Milk
3	Dairy	Even Better 1% Milk Club Whole Milk Booker 1% Milk
4	Dairy	Booker Chocolate Milk Carlson Buttermilk Even Better Whole Milk

Finally, according to the association rules shown in Table 3 as an example, when it is perceived that a customer is interesting in a product of dairy, three customized bundle list in table 3 will be provided. Moreover, the most favorite products at each category will be included in the list.

V. CONCLUSIONS

In this paper, a customized bundle recommendation is proposed for online supermarket. The customized bundle could not only reduce the time spent on choosing products for customers, but also convenient for the online supermarkets to pre-optimize the storage assignment of these products in the warehouse, and consequently reduce the cost for order fulfillment. Based on association rule minning, customer segmentation and recommendation technique, a method to recommend a customized bundle is investigated. By using the association rules of product category, the mass unnecessary association rules of product is avoided. Products recommended within each category are generated by customer segmentation and product ranking date based on segmentation. It can be easily extended to practical problem.

As future works, how to increase the effectiveness of the proposed method to adapt to large scale problem is one direction. In addition, how to combine bundle pricing method can be proposed by considering the customers' characteristics and needs. We leave them for future research.

REFERENCES

[1] Bain Consulting. China Shoppers Report 2017 [R]. Beijing: Bein Consulting, 2017.

[2] Why the once-ambitious No. 1 store fell to such an end? [OL]. (2017,09,19) http://www.admin5.com/article/20170919/785116.shtml.

[3] Jingdong quarterly graphic: operating loss of 1.6 billion operating margin -1.4% [OL]. (2018,03,03) http://www.p5w.net/weyt/201803/t20180302_2086205.htm.

[4] Acimovic J, Graves S C. Making better fulfillment decisions on the fly in an online retail environment[J]. Manufacturing & Service Operations Management, 2014, 17(1): 34-51.

[5] Jasin S, Sinha A. An LP-based correlated rounding scheme for multi-item ecommerce order fulfillment[J]. Operations Research, 2015, 63(6): 1336-1351.

[6] Karimi-Majd, A.-M., & Fathian, M. (2017). Extracting new ideas from the behavior of social network users. Decision Science Letters, 6(3), 207–220.

[7] Yang, T. C., & Lai, H. (2006). Comparison of product bundling strategies on different online shopping behaviors. Electronic Commerce Research and Applications, 5(4), 295–304.

[8] Miguéis, V. L., Camanho, A. S., & Falcão e Cunha, J. (2012). Customer data mining for lifestyle segmentation. Expert Systems with Applications, 39(10), 9359–9366.

[9] Liao, S. H., Chen, Y. J., & Hsieh, H. H. (2011). Mining customer knowledge for direct selling and marketing. Expert Systems with Applications, 38(5), 6059–6069.

[10] Liao, S. H., Chen, Y. J., & Lin, Y. T. (2011). Mining customer knowledge to implement online shopping and home delivery for hypermarkets. Expert Systems with Applications, 38(4), 3982–3991.

[11] Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods, and Algorithms (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[12] Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing & Customer Strategy Management, 19(3), 197–208