A Hybrid Mining Approach for Optimizing Returns Policies in e Retailing

Chien-Chih Yu and Cheng-Su Wang

Dept. Of Management Information Systems,
National ChengChi University

Abstract

This paper aims at proposing a hybrid mining approach for e-retailers to analyze returns patterns from both the customer and product perspectives, classify customers and products by returns ratios, and then adopt proper returns policies and marketing strategies to the identified classes for reducing returns transactions and sustaining profits. A multi-dimensional data framework and a process model for the hybrid mining approach are presented with a simulated example for validation.

Keywords: Returns policy, data mining, marketing strategy

1. Introduction

Returns policies have been adopted in various retailing industries including the e-shops of computer, book, and healthcare products. [2,3,6,10,14]. Researchers have pointed out that the returns policy is a critical yet controversial issue in the planning and implementation of supply chain and marketing strategies [13,16]. The most generous returns policy offers unconditional refund of wholesale/retail price for returned products, while the less generous returns policy accepts no returns at all or imposes some types of restrictions for returning [5,12]. It is noted that the adoption of returns policies may substantially affect product sales and operating costs. In the e-commerce era, the returns policy has increasingly become a more important strategic action for e-business to sustain competitiveness and profits. In the research literature, previous works regarding returns policy tend to formulate this problem as a mathematical model. Among these researches, Padmanabhan and Png (1997) concern the effect of a returns policy on pricing and stocking in a competitive retail sector [14]. Choi, Li, and Yan (2004) investigate the optimal returns policy for supply chain with an e-marketplace outlet [1]. Mukhopadhyay and Setoputro (2005), focusing on build-to-order products, develop a manufacturer's profit maximization model to jointly consider level of returns policy and level of modularity in product design [12]. However, many key factors are still missing in these previous works. For instance, customers' demographic data and transaction-based returns patterns, as well as potential associations between customer classes, product types, and marketing strategies are all crucial in the adoption and implementation of returns policies. Therefore, in order to make optimal decisions for returns policy, multiple factors from customer and product dimensions must be taken into account to ensure a win-win result for both the supply chain businesses and customers. The goal of this paper is to first propose a multi-dimensional framework for illustrating the key factors of the returns policies, and then to use a hybrid mining approach for analyzing return patterns, classifying customers and products with return ratios, as well as assigning suitable returns policies and marketing strategies to associated customer and product classes. In the following sections, section 2 provides a hybrid mining solution approach with a data framework and a process model. Section 3 demonstrates the hybrid mining process using an example with simulated data and embedded scenarios. The final section contains a conclusion and directions of future works.

2. The Framework and Process

To address the issues of returns policies, several data dimensions need to be considered. In the customer dimension, demographic data set include gender, age, education, and income level, etc. In the transaction dimension, data concerned include the detailed product purchasing records and associated returning status. In the product dimension, the product-related data mainly contain product type, price, size, level of customization, and ease of operation. Finally, in the marketing strategies dimension, promotion activity is considered. Taking transactions and related returns transactions as the facts to be monitored and analyzed, and recency, frequency, and monetary of transactions and returns transactions (marked as R, F, M and R-R, R-F, R-M) as the derived elements for measurement, we then have the customer, product, and marketing strategies as the associated dimensions.

The hybrid mining process for analyzing the returns patterns includes two stages. In the first stage, single dimensional clustering analyses are conducted

for the customer and product dimensions separately. The resulting clusters are then segmented in terms of return ratios (RR) and other significant data elements. Then, a decision tree function is activated to generate classification rules for properly classifying customers and products. In the second stage, a cross-dimensional analysis is performed to generate association rules with respect to customer and product classes, as well as marketing strategies. The objective of the hybrid mining approach is to classify customers and products for adopting suitable returns policies and marketing strategies in order to increase the monetary of transactions as well as to decrease the return ratios.

For processing the mining approach, a data framework for the input dataset is shown in Table 1. For a given time period, the recency of returns is the number of days from the latest returns transaction, the frequency of returns is the ratio of returns transactions and total buying transactions, and the monetary of returns is the total buyback prices of returned products. Values of the R-R, R-F, R-M are set to be High (H), Medium (M), or Low (L). Promotional activities for marketing include buy-one-get-one-free, double credits, and price discount.

Table 1 The Data Elements for Hybrid Mining Process

Dimension	Data Elements	Attributes/values
Customer	R-R, R-F, R-M	High, Medium, Low
	R, F, M of Trans.	High, Medium, Low
	Gender, Education, Age and Income	F/M, H/M/L,..
Product	R-F	High, Medium, Low
	Frequency of Trans.	High, Medium, Low
	Type	Consum. electronics, Furniture, Jewelry,etc.
	Price, Size, Ease of Operation	High, Medium, Low
Marketing Strategies	Promotional activities	Buy one get one free, Doubled-Credit, Price Discount…etc.

One other key factor to be investigated is the returns ratio that indicates the degree of returns patterns. Young (1988) evaluated returns ratio in transaction level. In this paper, we calculate the returns ratio in a different way. Using the data set in Table 2 as an example, the returns ratio is 2/10 that is obtained from dividing the two returned items by all ten items in three transactions. The detail hybrid mining process is illustrated in Figure 1 and described below.

2.1. Single Dimensional Analyses

In stage I of the hybrid mining process, we employ k-means technique for performing the clustering analyses. Clusters of both the customer and product dimensions are generated by analyzing the data set as provided in Table 1. For instance, using the R-R, R-F, R-M as the analyzed elements, there are possibly 27

clusters each has the high, medium, or low as values of these three chosen data elements. The resulting clusters are then segmented by return ratios and descriptive statistics related to demographic and transaction data. In the subsequent step, the decision tree classification algorithm (such as J48) is used to derive classification rules that reflect the returns patterns of the customer classes. Better returns policies can be chosen aiming at improving the returns ratios. New customers can be classified using these rules and connected to proper returns policies for optimizing both sales and returns. Similar clustering and classification processes can be carrying out for product dimension with the data elements including product type, size, price, and ease of operation.

Table 2 The Data Elements for Hybrid Mining Process

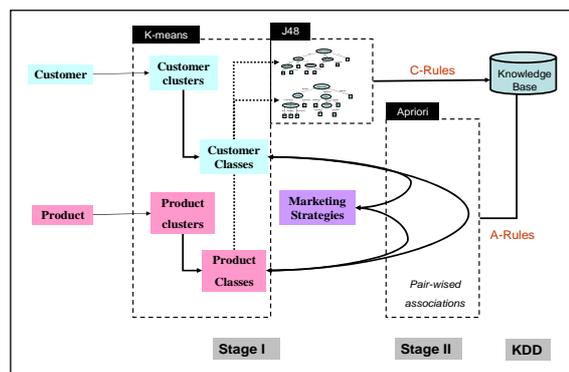| NO.	Transaction items	Returned items
T1	A,B,C	-
T2	D,B,F	D
T3	A,C,E,F	A

Fig. 1 The hybrid mining process

2.2. Cross Dimensional Analysis

In Stage   of the mining process, we conduct pair-wised association mining using Apriori algorithm. We took the customer and product classes, and marketing strategies as inputs of the association mining process. Each dimension would pair-wised associate with another dimension to find out more return patterns, as well as to get better ideas for adopting more specific returns policies and marketing strategies. For knowledge discovery from the mining process, both the validated classification rules (C-rules) and association rules (A-rules) are added into a knowledge base for further application.

3. An Experimental Example

For testing the proposed approach, we generate a set of simulated data that consists of 1000 transactions in association with 100 customers and 50 products based

on Faust's (1986) survey result [4]. Returns transactions are set to be 10% of the total transactions. In addition, we set up a test scenario in the customer dimension that the returns ratios are high for younger female with medium rating in education level. Also in the product dimension, we direct another scenario that the high price jewelry items have high returns ratios. Using this dataset with embedded scenarios, we conduct the hybrid mining process in two stages as described below.

3.1. The First Stage Mining

We first perform the clustering analyses for both the customer and product dimensions using K-means algorithm. In the customer dimension, we select the R-R, R-F, R-M (denoted as RFM-R) as the input data elements. We then examine the returns ratios (RR) and the RFM data of regular transactions. Totally 4 clusters (19%,64%,13%,4%) are generated and are segmented into 3 classes (19%, 68%, 13%) based on levels of the RR and the RFM. The clusters with H in returns ratio and medium to high in both RFM-R and RFM are grouped as a class that indicates a group of important customers with high returns to be watched. In other words, the clusters with (HHH, HHH, H), (HHH, HMM, H), (HMH, MMH, H), and the like as the value sets of (RFM-R, RFM, RR) are put into a class labeled as high-class-high-returns. Similarly, we have classes of medium-class-medium-returns, and low-class-low-returns, etc. In the product dimension, we use product price, type, and ease of operation as input data set for cluster mining. The 4 generated clusters (10%, 26%, 24%, 40%) are then segmented into 3 classes (36%, 24%, 40%) based on product returns ratios. In Table 3 (a) and (b), three segmented clusters and associated descriptive statistics generated in both the customer and product dimensions. In the next step, we conduct classification mining processes for both the customer and product dimensions based on the labeled segments. Two decision trees generated for the customer and product dimensions are illustrated in Figure 2 with associated accuracy rates. From the decision trees, 6 classification rules for the customer dimension and 5 classification rules for the product dimension are obtained. For derived customer classes, the total accuracy rate is 96%, and the precision/recall rates for class S3, S1, and S2 are 1/1, 0.826/1 and 1/0.941. For derived product classes, the accuracy rate is 80%, and the precision/recall rates for S1, S3, and S2 are 0.944/0.944, 0.727/0.8, and 0.7/0.583. The embedded scenarios such as the high returns ratios for younger female with high education level and for high price jewelry have all been successfully spotted. To be more specific, the classification rule for the first scenario is presented as "If the Education Level is High, and the Gender is

Female, and the Age is Medium or Low, then the Returns Ratio is High". Using these classification rules, customers can be classified according to their demographic data to predict their returns patterns. More suitable returns policies for different customer classes can be adopted to decrease potential returns.

Table 3 (a). An example of classified Customer clusters

Cluster	Class	Gender (%)	Edu (%)	Age (%)	Income (%)
RFM-R = HHH	S3	F	H	M	76.92-H 23.08-M
RFM-R =LMM	S1	68.42-F 31.58-M	10.53-H 89.47-M/H	10.53 -H 89.47-M	31.58-H 68.42-M
RFM-R = LLL	S2	64.71-F 35.29-M	14.71-H 16.18-M/H 39.71-M 29.41-L	33.82-H 45.59-M 20.59-L	25.00-H 44.12-M 30.88-L

Gender: F=Female, M=Male

Table 3 (b). An example of classified Product clusters

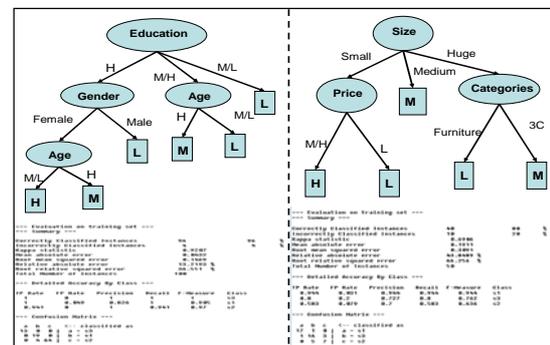| Cluster	Class	Type (%)	Size (%)	Oper (%)	Price (%)
F= M/L	S1	22.22-3C 77.78-Jew	77.78-L 22.22-M	66.67-H 27.78-M 5.56-L	66.67-H 27.78-M 5.56-L
F=H	S3	75.00-3C 5.00-Jew 20.00-Fur	55.00-H 40.00-M 5.00-L	35.00-H 45.00-M 20.00-L	40.00-H 55.00-M 5.00-L
F=M	S2	50.00-3C 50.00-Fur	50.00-H 33.33-M 16.67-L	25.00-H 16.67-M 58.33-L	25.00-H 50.00-M 25.00-L

Jew=Jewelry, Fur=Furniture

Fig. 2 Decision trees for customer and product classes

Similar results can be obtained for product classifications. The specified classes with candidate returns policies are listed in Table 4. For an example, for customers in class 3 with high returns ratio, we can adopt partial or tight returns policies. On the other hand, for products belong to class 3 with medium returns ratio, we can apply partial returns policies.

3.2. The Second Stage Mining

In Stage II, we perform the cross dimensional paired-wise association mining process across customer, product, and marketing strategy dimensions. There are

22 association rules obtained with minimum support and confidence set as 0.3 and 0.5.

Table 4 Candidate returns policies for the derived classes

Dimension	Classes	Returns policies
Customer	S3:RFM = HMM RR = 0.5109=H	Partial/Tight
	S1:RFM = LMM RR =0.1258= M	Partial
	S2:RFM = LLM RR =0.0021= L	Loose
Product	S1:RF = M/H RR =0.6710= H	Tight
	S3:RF =M/L RR = 0.0039=M	Partial
	S2:RF=L RR =0= L	Loose

Table 5 displays three of selected association rules. The first rule shows that customer class 3 and product class 1 imply potential returns transactions. We can assure that we should adopt tight returns policy for customers in class 3 buying products of class 1. The second/third rule indicates that the customer class 3/product class 1 and the marketing strategy A (Buy 1 Get Free 1) imply possibly no returns. We can then use this result to refine the returns policies and marketing strategies assigned to the customer class 3/product class 1. For instance, we can offer customers in class 3 more generous return policies when the promotion activity is 'buy 1 get free 1' to leverage sales while maintaining low returns.

Table 5 Three example association rules

IF	Then
Customer Class S3 and Product Class S1 (393)	Returns=Yes (233) Support: 0.37 Conf:0.6
Customer Class S3 and Marketing Strategy A (Buy 1 Get Free 1) (273)	Returns=No (204) Support: 0.32 Conf: 0.75
Product Class 1 (in Table 3) and Marketing Strategy A (291)	Returns=No (204) Support: 0.32 Conf: 0.70

4. Conclusion

Adopting proper returns policies has been a common but critical issue for retailers to gain higher customer satisfaction and profits. In this paper, we propose a multi-dimensional data framework and a hybrid data mining approach to deal with this problem in two stages. By using the generated classification and association rules, better returns policies and marketing strategies can be adopted for labeled classes to increase sales as well as decrease returns. Future research works will focus on extending the proposed hybrid data mining process to e-supply chain, as well as on conducting the knowledge discovery process based on some real world data to validate the efficiency and effectiveness of this approach.

5. References

[1] T. M. Choi, D.Li, and H. Yan, "Optimal Returns Policy for Supply Chain with e-Marketplace," *International Journal of Production Economics*, 88(1), pp.205-227, 2004.

[2] S. Davis, M. Hagerty, and E. Gerstner, "Return Policies and the Optimal Level of Hassle," *Journal of Economics and Business*, 50(5), pp.445-460, 1998.

[3] R. O. Eduardo, and R. P. Andres, "The Regional Return of Public Investment Policies in Mexico," *World Development*, 32(9), pp.1545-1562, 2004.

[4] T. Faust, "Merchandise Returns—Customer's Viewpoint," *Quarterly Review of Commerce*, 8(2), pp.152-167, 1986.

[5] K. H. Hahn, H. Hwang, and S.W. Shinn, "A Returns Policy for Distribution Channel Coordination of Perishable Items," *European Journal of Operational Research*, 152(3), pp.770-780, 2004.

[6] W. Hoffman, J. Keedy, and K. Roberts,"The Unexpected Return of B2B.," *The McKinsey Quarterly,* No. 3, pp. 97-105, 2002.

[7] N.C. Hsieh, "Hybrid Mining Approach in the Design of Credit Scoring Models," *Expert Systems with Applications*, 28(4), pp.655-665, 2005.

[8] R. J. Kuo, L. M. Ho, and C. M.Hu, "Integration of Self-Organizing Feature Map and K-means Algorithm for Marketing Segmentation," *Computer and Operations Research*, 29(11), pp.1475-1493, 2002.

[9] A. H. L. Lau, H. S. Lau, and D. K. Willett, "Demand Uncertainty and Returns Policies for a Seasonal Product: An Alternative Model," *International Journal of Production Economics*, 66(1), pp.1-12, 2000.

[10] T. Longo, "At Stores, Many Unhappy Returns," *Kiplinger's Personal Finance Magazine*, 49(6), pp. 103, 1995.

[11] M. K Mantrala, and K. Raman, "Demand Uncertainty and Supplier's Returns Policies for a Multi-Store Style-Good Retailer," *European Journal of Operational Research,* 115 (2), pp.270-284, 1999.

[12] S. K. Mukhopadhyay, and R. Setoputro, "Optimal Return Policy and Modular Design for Build-To-Order Products," *Journal of Operations Management*, 23(5), pp.496-506, 2005.

[13] V. Padmanabhan, and I. P. L. Png, "Returns Policies: Make Money by Making Good," *Sloan Management Review,* Fall, 37(1), pp. 65-72, 1995.

[14] V. Padmanabhan, and I. P. L Png, "Manufacturer's Returns Policies and Retail Competition," *Marketing Science*, 16(1), pp.81-94, 1997.

[15] S. Webster, and Z.K. Weng, "A Risk-free Perishable Item Returns Policy," *Manufacturing & Service Operations Management,* 2(1), pp.100-106,2000.

[16] D. Q. Yao, X. Yue, X. Wang, and J.J. Liu, "The Impact of Information Sharing on a Returns Policy with the Addition of a Direct Channel," *International Journal of Production Economics*, 97(2), pp.196-209, 2005.

[17] K.B. Young, "Product Return Rates," *Quality,*27(5), pp. 38-41,1988.