Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 1575–1582
Expert Systems with Applications
www.elsevier.com/locate/eswa

A hybrid mining approach for optimizing returns policies in e-retailing

Chien-Chih Yu *, Chen-Shu Wang

Department of Management Information Systems, National Cheng-Chi University, 64 Section 2, Chihnan Road, Mucha Dist., 116 Taipei, Taiwan

Abstract

The returns policy has long been considered as a critical yet controversial issue in the development of supply chain and marketing strategies. Up-stream manufacturers or distributors may offer returns policies to the down-stream retailers or customers to increase order and sales quantities. There are trade-offs between returns policies and customer satisfaction, product sales, and operating costs. The goal of this paper is to use a hybrid mining approach for analyzing return patterns from both the customer and product perspectives, classifying customers and products into levels, and then for adopting proper returns policies and marketing strategies to these customer classes for sustaining better profits. A multi-dimensional framework and an associated model for the hybrid mining approach are provided with a demonstrated example for validation. It is expected that by adopting suitable returns policies, benefits can be created and shared by both e-retailers and customers.

Keywords: Returns policies; e-Retailing; Hybrid data mining approach

1. Introduction

Returns policies have been adopted across various industries such as computer, publication, and pharmaceutical industries (Davis, Hagerty, & Gerstner, 1998; Eduardo & Andres, 2004; Hoffman, Keedy, & Roberts, 2002; Longo, 1995; Padmanabhan & Png, 1997). 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 (Lau, Lau, & Willett, 2000; Mantrala & Raman, 1999; Padmanabhan & Png, 1995; Yao, Yue, Wang, & Liu, 2005). Up-stream manufacturers or distributors may offer returns policies to the down-stream retailers or customers to increase order and sales quantities. 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 (Hahn, Hwang, & Shinn, 2004; Mukhopadhyay & Setoputro, 2005; Webster & Weng, 2000). It is noted that the adoption of returns policies may substantially affect product sales and operating costs. A loose returns policy can stimulate customers' buying decisions to leverage sales volume; nevertheless, it can also increase the number of return transactions that incur more handling and logistic costs. In the e-commerce era with more direct channels, customized products, and online shoppings, the returns policy has become an even 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 in which sales profits is the objective function to be optimized and the buyback price for returned products is the major decision variable. Among these researches, Padmanabhan and Png (1997) concern the effect of a returns policy on pricing and stocking in a competitive retail sector. They show that manufacturers should accept returns if production costs are sufficiently low and demand uncertainty is not too great. Choi, Li, and Yan (2004) investigate the optimal returns policy (also called a buyback policy) for supply chain with e-marketplace in which returned product can be sold with a higher price.

* Corresponding author. Tel.: +886 2 29387349; fax: +886 2 29393754.
E-mail addresses: ccyu@nccu.edu.tw (C.-C. Yu), 93356506@nccu.edu.tw (C.-S. Wang).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.08.099

Mukhopadhyay and Setoputro (2005) develop a manufacturer's profit maximization model to jointly consider level of returns policy (buyback price for returned products) and level of modularity in product design for build-to-order products. Although these optimization approaches for dealing with returns policies do provide partial solutions to the strategic problem, many key factors are still missing in the model set up. For instance, customers' demographic and transaction-based characteristics such as gender, income level, frequency of buying, average monetary of transactions, as well as return patterns are critical for properly classifying customers and selecting returns policies.

In addition, whether product types and complexity of operation are critical factors that would influence the likelihood of customers' return transactions? Are there any associations between customer classes, product types, and return patterns? These questions are no doubt crucial in making decisions related to the adoption and implementation of returns policies. Therefore, in order to make optimal decisions for returns policy, multiple factors from customer, product, and supply chain dimensions must be taken into account. It is reasonable to start from analyzing the return patterns of customers and products.

As the business realizes more about the return patterns, they can offer to their customers better returns policies that can not only increase product sales but also decrease the probability of returns as well as associated handling costs. As a result, adopting returns policies can eventually become a win–win strategic move to benefit the supply chain businesses and customers as well. 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 directing suitable returns policies and marketing strategies to associated customer and product classes (Hsieh, 2005; Kuo, Ho, & Hu, 2002). The rest of this paper is organized as follows. The framework and hybrid mining model for returns policies are provided in Section 2. Section 3 demonstrates the mining process using an example with simulated data. Relating returns policies and marketing strategies is also discussed in this section. The final section is a conclusion and directions of future works.

2. The framework and process

Generally speaking, there are several dimensions to be considered for fully addressing the issues of returns policies. For the customer dimension, different customer classes can be specified in different levels of the supply chain. The manufacturers have retailers and/or individual buyers as their target customers. The retailers sell products to general consumers. Customers can be further characterized by demographic data as well as transaction data. Data elements of the demographic data set include gender, age, education, income level, etc. Transaction-based data contain the detailed product purchasing and returning records, and is strongly related to the product dimension. The product dimension mainly considers product type, price, size, level of customization, and ease of operation as data elements that may affect customers' return patterns and businesses' adoption of returns policies. Product types can be categorized as seasonal, perishable, computer, jewellery, and many others. Build-to-order products, in contrast to mass products, have the highest level of customization. The third dimension identifies the types and restrictions of the returns policies. Types of returns policies include loose, tight, and partial policies. The loose or generous returns policy offers unconditional money back guarantee. The tight returns policy accepts no refund whatsoever. The partial returns policy provides store credit for future purchase or offers only percentage of the selling price for returned products. Restrictions impose on returns policies include unused product only, time limit for returning, return in original package, and other instructions. Restrictions on returns also indicate the level of difficulty in returning product items.

As for the marketing channels and strategies dimension, the selling channels can be specified into direct channel, e-marketplace, department store chain, specialty store chain, and single out-let specialty stores. The marketing strategies considered include buy-one-get-one-free, double credits, and price discount. Taking return transactions as the fact to be monitored and analyzed, and associated recency, frequency, and monetary of return transactions as the derived elements for measurement, we then have the customer, product, returns policies, and marketing strategies as the associated dimensions. An integrated framework with simplified scope for structurally representing dimensional factors of return patterns is depicted in Fig. 1.

The hybrid mining model and process for analyzing the return patterns includes three stages. In the first stage, single dimensional clustering analyses are conducted for the customer and product dimensions. In the second stage, resulting clusters of first stage are then segmented into a group of classes in terms of return ratios and other significant data elements. In the third stage, the first set of classification rules for classifying customers and products for selecting proper returns policies and marketing strategies are generated at this stage as intermediary results. In addition, the cross-dimensional analysis is performed to generate association rules with respect to customer and product classes, as well as return ratios. The classification scheme and the assignment of returns policies and marketing strategies are then adjusted to get the final mixed returns and marketing policies. The objective of the 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, an example input dataset is shown in Table 1. In the customer dimension, the recency of returns is the number of days from the latest

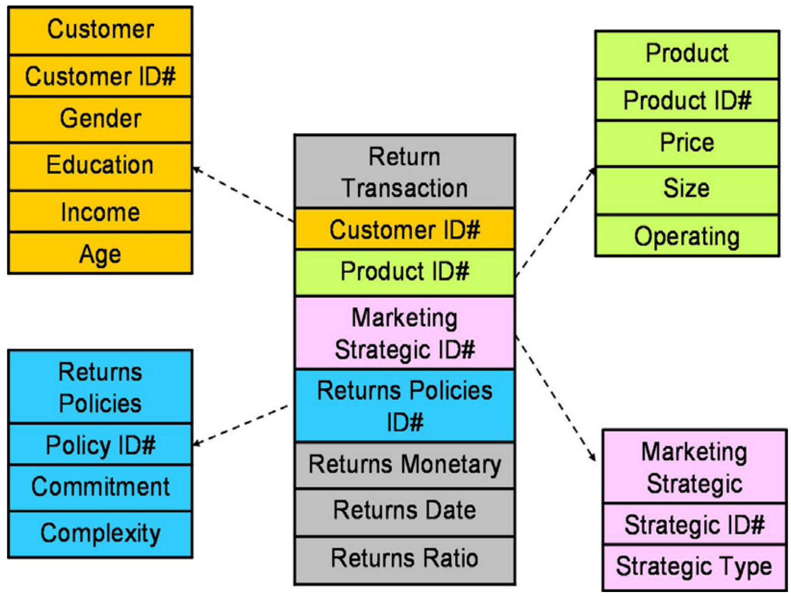

Fig. 1. The integrated framework for analyzing returns patterns.

Table 1
The data elements for hybrid mining process

Dimensions	Data elements	Attributes/values
Customer	Recency of returns	High, Middle, Low
	Frequency of returns	High, Middle, Low
	Monetary of returns	High, Middle, Low
	Recency	High, Middle, Low
	Frequency	High, Middle, Low
	Monetary	High, Middle, Low
	Demographic data	Values of gender, education, age and income
Product	Frequency of returns	High, Middle, Low
	Frequency	High, Middle, Low
	Category	Consumable electronics, furniture, jewellery,. . .etc.
	Price	High, Middle, Low
	Size	Huge, Large, Middle, Small
	Ease of operation	Difficult, Normal, Easy
Marketing strategies	Promotional activities	Buy-one-get-one-free, doubled-credit, price discount,. . .etc.

return transaction. The lower the result indicates that the customer has return transaction more recently. The frequency of returns is the ratio of return transactions and total buying transactions within the span of a specified time horizon. The monetary of returns is the total buyback prices of returned products for a given time period. Other customer related data elements are demographic data such as gender, age, education, and income. In the product dimension, the product price, size and ease of operation are specified according to product categories. For example,

Table 2
Illustration transaction data

No.	Transaction items	Returned items
T1	A,B,C	–
T2	D,B,F	D
T3	A,C,E,F	A

the price of 3C electronic product is lower than Jewellery, but the size is bigger and the operation is harder. Finally, in the marketing strategies dimension, promotional activities include buy-one-get-one-free, double credits, and price discount.

One other key factor to be investigated is the return ratio that indicates the degree of return patterns. Young (1988) mentioned return ratio as transaction level. However, with different perspectives, the return ratio can be calculated in three different ways. Using the data set in Table 2 as the example for deriving return ratios, there are three transactions with ten purchased items and two returned transactions each with one returned item. The first return ratio is 2/3 that is equivalent to the frequency of returns. The second return ratio is 2/10 that is obtained from dividing the two returned items by all ten items in three transactions. And the third return ratio is 3/4 that comes from the average of item A's return ratio (1/2) and item D's return ratio (1/1). Without loss of generality, we choose the second format for generating return ratios. The detail hybrid mining process is illustrated in Fig. 2 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 (Hsieh, 2005). Clusters of both the customer and product

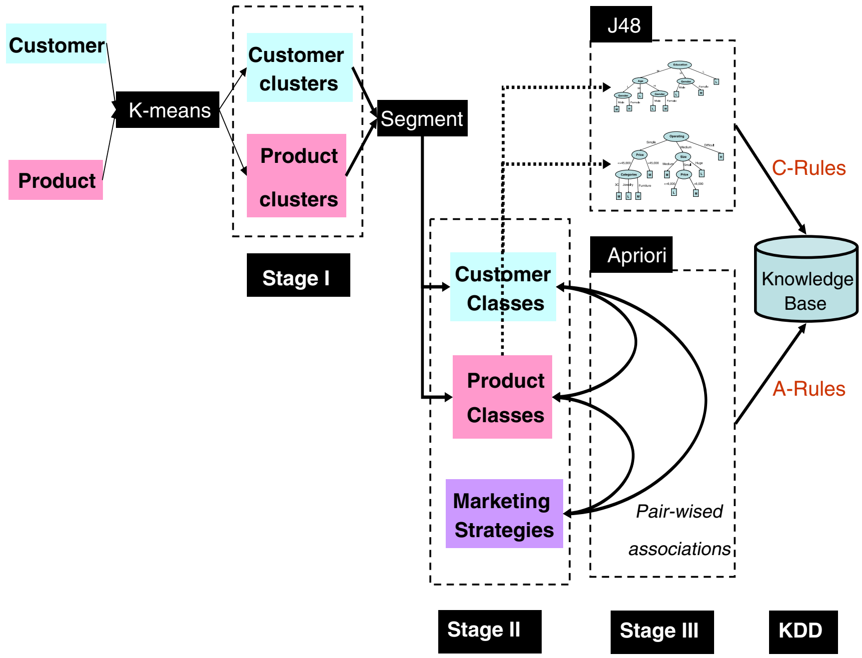

Fig. 2. The hybrid mining process.

dimensions are generated by analyzing the data set provided in Table 1. For instance, using the recency, frequency, and monetary of returns 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 into high, medium, and low classes. For these classes, descriptive statistics related to demographic and transaction data as well as data of returns policies and marketing strategies are then generated and analyzed.

In the subsequent step, the decision tree or other classification algorithms (such as J48) are used to derive classification rules that reflect the return patterns of the customer classes. A better returns policies as well as marketing strategies can be chosen aiming at improving the return 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 carried out for product dimension with the data elements including product type, size, price, and ease of operation.

2.2. Cross-dimensional analysis

In Stage II of the mining process, we conduct pair-wised association mining using Apriori algorithm. We took the clusters and/or classes of each dimension as inputs of the association mining process. Each dimension would pair-wised associate with another dimension to find out more return patterns. In other words, this stage probes the impacts across customer and product dimensions. Hence, the return patterns and rules obtained from Stage II can be refined in Stage III to get better ideas for adopting more specific returns policies and marketing strategies. Some interesting findings can be expected in this stage. For instance, there may show no significant return patterns with respect to the adoption of marketing strategies in stage II. However, associating customer classes with marketing strategies may illustrate some return patterns such as that certain customer types with specific marketing strategies may imply certain degree of return ratios that is to be watched more carefully. The proposed process can be regarded as knowledge discovery; the classification rules (C-rules) and association rules (A-rules) can be added into knowledge base for further application.

3. A demonstrated 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. As Fig. 1 shows, a transaction record consists of TID#, CID#, PID#, purchase date, marketing strategies and returns status. A customer record contains CID#, gender, age, education degree and income levels. A product record includes PID#, price, product type, size and operator. The RFM data of a customer can be obtained by joint-operation from customer and transaction DBs. In addition, the RFM data of product can also be acquired by joint-operation from transaction and product DBs. Both RFM information of customer and product are analyzed by the proposed approaches.

Referring to Faust's (1986) survey result, 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 return ratios of younger female with middle level in education are high. Also in the product dimension, we direct the high price jewelry to have high return ratios. Based on this dataset and assumed return patterns, we conduct the hybrid mining process in the following two stages.

3.1. Single dimension analyses

We first performed the clustering analyses for both the customer and product dimensions. In the customer dimension, we selected the recency (R), frequency (F), and monetary (M) of return transactions (denoted as RFM-R) as the input data for cluster mining. We then examined the return ratios (denoted as RR) and the RFM data of regular transactions. Totally 27 potential clusters were segmented using return ratios and levels of the RFM in regular transactions. The clusters with H in the return 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 high-class-low-returns, medium-class-medium-returns, low-class-low-returns, etc. In the product dimension, we employ product's price, type, and ease of operation as input data set for clustering. Product types are then examined for each derived clusters.

In Tables 3a and 3b, three example classified clusters and descriptive statistics obtained in both the customer and product dimensions are listed with candidate returns policies and marketing strategies assigned in Table 4.

Then, we can conduct classification mining processes for both the customer and product dimensions based on the labeled return ratios. Two decision trees generated for the customer and product dimensions are illustrated in Fig. 3.

From the decision trees, six classification rules for the customer dimension and five classification rules for the product dimension can be generated. The embedded scenarios such as the high return ratios for younger female with middle level education and the high return ratios for high price jewelry have all been successfully spotted. To be more specific, two of these rules are presented as "If the Education Level is High, and the Gender is Female, and the Age is between Medium and Low, then the Return Ratio is High", and "If the Size of product is Small, and the Price is between High and Medium, then the Return Ratio is High".

Using these classification rules, customers can be classified according to their demographic data to predict their return patterns. More suitable returns policies and marketing strategies for different customer classes can be adopted to generate more sales while decrease potential returns. Similar results can be obtained for product classifications.

3.2. Cross-dimension analysis

In addition to the clusters and classes based on return patterns and other key factors in each of the customer and product dimensions found in Stage I, we then perform in stage II the cross-dimensional analysis. Paired-wise association mining process across different dimensions is executed. Table 5 displays two example association rules obtained from the process.

Table 3a
Example of classified clusters for customer dimension

Clusters	Classes	Gender (%)	Education (%)	Age (%)	Income (%)
RFM-R = HHH	RFM = HMM	100 (F)	100 (H)	100 (M)	76.92 (H)
	RR = H				
	(0.5109)				23.08 (M)
RFM-R = LMM	RFM = LMM	68.42 (F)	10.53 (H)	10.53 (H)	31.58 (H)
	RR = M				
	(0.1258)	31.58 (M)	89.47 (M/H)	89.47 (M)	68.42 (M)
RFM-R = LLL			14.71 (H)		
	RFM = LLM	64.71 (F)	16.18 (M/H)	33.82 (H)	25.00 (H)
	RR = L	35.29 (M)	39.71 (M)	45.59 (M)	44.12 (M)
	(0.0021)		29.41 (L)	20.59 (L)	30.88 (L)

Table 3b
Example of classified clusters for product dimension

Clusters	Classes	Type (%)	Size (%)	Operator (%)	Price (%)
F = M/L	RF = M/H	22.22 (3C)	77.78 (L)	66.67 (H)	66.67 (H)
	RR = H			27.78 (M)	27.78 (M)
	(0.6710)	77.78 (Jew)	22.22 (M)	5.56 (L)	5.56 (L)
F = H	RF = M/L	75.00 (3C)	55.00 (H)	35.00 (H)	40.00 (H)
	RR = M	5.00 (Jew)	40.00 (M)	45.00 (M)	55.00 (M)
	(0.0040)	20.00 (Fur)	5.00 (L)	20.00 (L)	5.00 (L)
F = M	RF = L	50.00 (3C)	50.00 (H)	25.00 (H)	25.00 (H)
			33.33 (M)	16.67 (M)	50.00 (M)
	RR = L (0)	50.00 (Fur)	16.67 (L)	58.33 (L)	25.00 (L)

Table 4
Returns policies recommendation

Dimensions	Classes	Marketing strategies	Returns policies
Customer	RFM = HMM	Buy-one-get-one-free	Partial,
	RR = H	Price discount	Tight
	RFM = LMM	Price discount	Partial
	RR = M		
	RFM = LLM	Double credit	Loose
	RR = L		
Product	RF = M/H	Buy-one-get-one-free	Partial,
	RR = H	Double credit	Tight
	RF = M/L	Double credit	Partial
	RR = M	Price discount	
	RF = L	Double credit	Loose
	RR = L	Price discount	

Table 5
Two example association rules

Dimensions	Data elements	Imply
Customer and Product	Customer Class 1 (in Table 3) and Product class 1 (in Table 3)	Return ratio = High Support: 0.37 Conf: 0.59
Product and marketing strategies	Product class 1 (in Table 3) and Marketing strategy (Buy-one-get-free-one)	Return ratio =; Low Support: 0.32 Conf: 0.70

The first rule indicates that the customer class 1 and the product class 2 (as listed in Table 3) imply low-return ratios. We can then use this result to refine the returns policies and marketing strategies assigned to the customer class 1. For instance, we can offer customer class 1 with product class 2 buying transactions more generous returns policies and more favorable promotion activities. According to the second association rule, which reveals a low-return ratio for product class 2 with double credit marketing act, a double credit marketing strategy will be chosen as the final promotion action for this cross-dimensional class. One interesting finding is that the return pattern of marketing strategy is not significant in Stage I; however, the return pattern shows off when it is associated with the product class.

4. Physical practice

The proposed approach tries to deal with the returns policies problem. The proposed approach provides appropriate returns policies recommendation by predicting return ratio. The experiment results demonstrate that businesses can provide looser returns policies to their customers with lower return ratios. The looser returns policies will not evoke mass

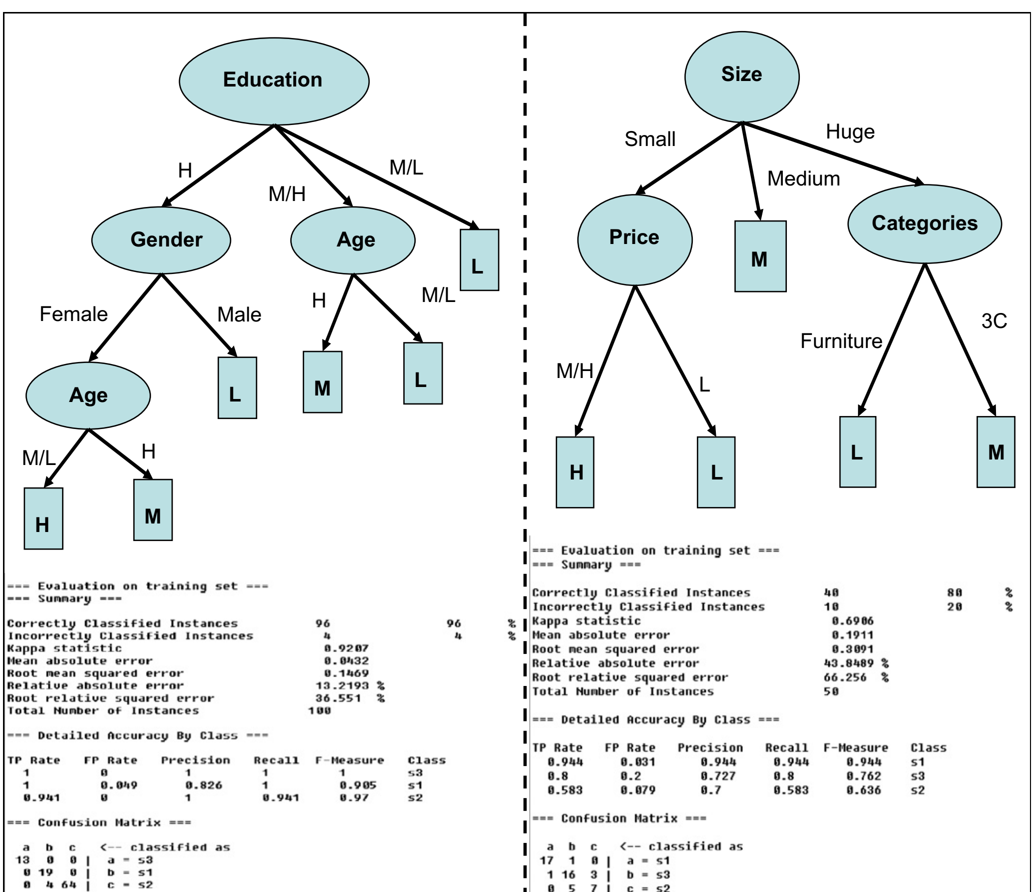

Fig. 3. Decision trees for the customer and product dimensions.

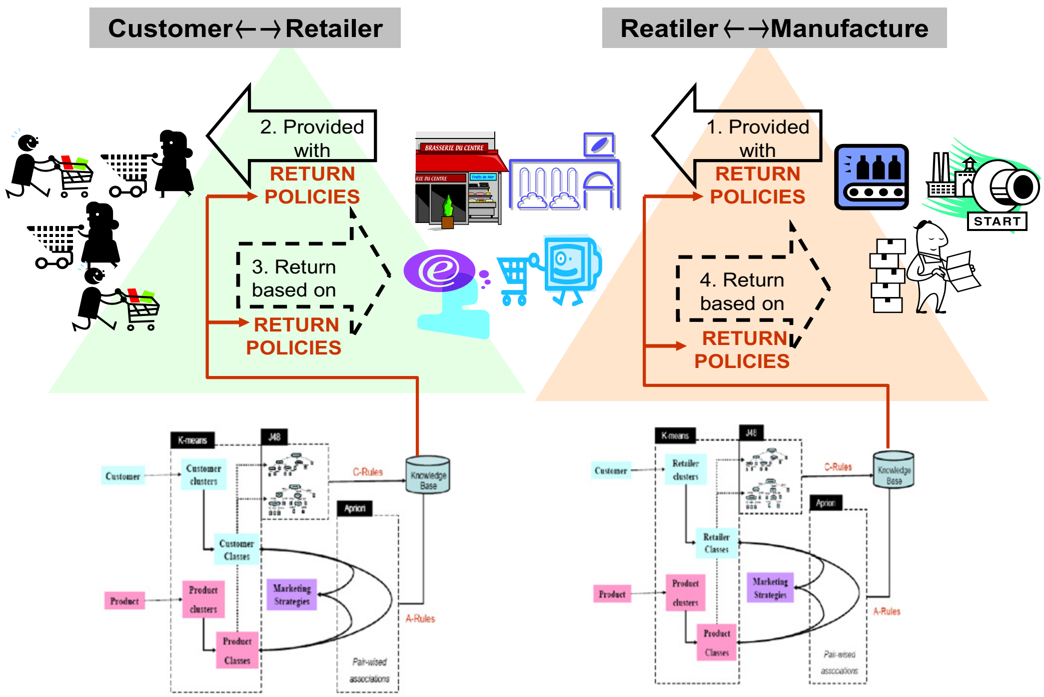

Fig. 4. Embedding proposed approach in entire supply chain.

returns due to lower return ratio. In addition, the looser returns policies can put an image to customers with the superior qualities of products. Thus, the customer may make purchase decision more comfortable. On the contrary, business will provide tighter returns policies to customer with higher return ratios. We expect to reduce irrational returns due to strict returns requirements and processes.

The proposed approach can be applied to virtual as well as brick-and-mortar stores iff these businesses keep their transaction, customer and product DBs. Furthermore the proposed approach can be extended to the entire supply chain as shown in Fig. 4. In this research, we focus on the returns policies between customers and retailers. However, the proposed approach can also be applied to retailers and manufacturers. In detail, the same returns scenario also happens between retailers and manufacturers. Thus, manufacturers can provide levels of returns policies to retailers based on return patterns analyzed. For illustration, the manufacturer may analyze return patterns from retailer dimension (sales amount, types and so on) and product dimension (price, time and so on). Then, manufacturers can provide suitable returns policies to different retailers. The looser returns policies are great rewards for retailers due to lower stock loss. Thus, retailers will strive for looser returns policies, then try to reduce their returns. Eventually, we expect the returns policy to be regarded as a win–win solution within the entire SCM.

5. Conclusion

We should regard returns as happy returns and making money in reverse (Stock, Speh, & Shear, 2002). 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 framework and a hybrid data mining approach to deal with this problem. Through two stages of the mining process, we generate clusters and classes for the customer and product dimensions, and derive some cross-dimensional association rules. By using the classification and association rules, better returns policies and marketing strategies can be adopted for labeled classes to increase sales and decrease returns. This proposed hybrid mining approach can be extended and applied to the entire supply chain. In addition, when returned products can be sold through other channels such as the e-marketplace, products' lifecycles and values can be extended and preserved. Future research works will focus on extending the proposed model and approach to e-supply chain, as well as on conducting this hybrid mining process using some real world data to validate the effectiveness of the framework and process.

References

Choi, T. M., Li, D., & Yan, H. (2004). Optimal returns policy for supply chain with e-marketplace. *International Journal of Production Economics, 88*(1), 205–227.

Davis, S., Hagerty, M., & Gerstner, E. (1998). Return policies and the optimal level of Hassle. *Journal of Economics and Business, 50*(5), 445–460.

Eduardo, R. O., & Andres, R. P. (2004). The regional return of public investment policies in Mexico. *World Development, 32*(9), 1545–1562.

Faust, T. (1986). Merchandise returns—customer's viewpoint. *Quarterly Review of Commerce, 8*(2), 152–167.

Hahn, K. H., Hwang, H., & Shinn, S. W. (2004). A returns policy for distribution channel coordination of perishable items. *European Journal of Operational Research, 152*(3), 770–780.

Hoffman, W., Keedy, J., & Roberts, K. (2002). The unexpected return of B2B. *The McKinsey Quarterly*(3), 97–105.

Hsieh, N. C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications, 28*(4), 655–665.

Kuo, R. J., Ho, L. M., & Hu, C. M. (2002). Integration of self-organizing feature map and *K*-means algorithm for marketing segmentation. *Computer and Operations Research, 29*(11), 1475–1493.

Lau, A. H. L., Lau, H. S., & Willett, D. K. (2000). Demand uncertainty and returns policies for a seasonal product: An alternative model. *International Journal of Production Economics, 66*(1), 1–12.

Longo, T. (1995). At stores, many unhappy returns. *Kiplinger's Personal Finance Magazine, 49*(6), 103.

Mantrala, K. M., & Raman, K. (1999). Demand uncertainty and supplier's returns policies for a multi-store style-good retailer. *European Journal of Operational Research, 115*(2), 270–284.

Mukhopadhyay, S. K., & Setoputro, R. (2005). Optimal return policy and modular design for build-to-order products. *Journal of Operations Management, 23*(5), 496–506.

Padmanabhan, V., & Png, I. P. L. (1995). Returns policies: Make money by making good. *Sloan Management Review, 37*(1), 65–72.

Padmanabhan, V., & Png, I. P. L. (1997). Manufacturer's returns policies and retail competition. *Marketing Science, 16*(1), 81–94.

Stock, P., Speh, T., & Shear, H. (2002). Many happy (product) returns. *Harvard Business Review, 80*(7), 16–17.

Webster, S., & Weng, Z. K. (2000). A risk-free perishable item returns policy. *Manufacturing & Service Operations Management, 2*(1), 100–106.

Yao, D. Q., Yue, X., Wang, X., & Liu, J. J. (2005). The impact of information sharing on a returns policy with the addition of a direct channel. *International Journal of Production Economics, 97*(2), 196–209.

Young, K. B. (1988). Product return rates. *Quality, 27*(5), 38–41.