Retail Inventory Management with Purchase Dependencies

Pradip Kumar Bala

***Abstract*—'Purchase dependency' refers to the dependency of purchase or non-purchase of one item or itemset on the purchase or non-purchase of another item or itemset. The research in this paper models for incorporating purchase dependencies in retail multi-item inventory management. One illustrative example has been discussed with data mining in retail sale data. Various types of purchase dependencies including negative dependency have been identified and discussed. These purchase dependencies are relevant for inventory management in retail stores. The model with an illustration in this research paper will be motivational to researchers and inventory managers for incorporating purchase dependencies while managing inventory.**

***Index Terms*—Data Mining, Inventory Management, Purchase Dependency, Multi-Item Inventory, Negative Dependency, Retail Inventory.**

I. INTRODUCTION

Customer relationship management (CRM) aims at stronger loyalty of customers with greater market share. With competition for shelf space intensifying, there is a pressing need to provide shoppers with a highly differentiated value proposition through "right product mix in right amount at right time". Mining or extracting consumer insight from structured and unstructured data, call-center records, customer complaints, and other sources will be of tremendous importance for inventory management in retail stores. Purchase dependency is one type of consumer which has not been addressed or used properly for managing inventory in retail sale. Purchase or non-purchase of item or items by a customer may depend on the purchase or non-purchase of some other item or items made by him or her. This is referred as purchase dependency in this paper.

In most of the real situations, we do not find a one-item inventory, but multi-item inventory with different periods of replenishment. These inventories are usually managed in aggregate because of the complexity of handling each individual item. Intuition and commonsense rules with the aid of historical data have been useful to some extent, but they are not often efficient or cost effective. Some examples of the multi-item inventory are retail sale store, spare parts for maintenance, medicine store etc.

Out of thousands of items held in an inventory of a typical organization, only a small percentage of them deserve management's closest attention and tightest control. It is not economical to exercise same degree of inventory control on all the items. Using selective inventory control, varying degree of control are exercised on different items. Most widely-used technique used to classify items for the purpose of selective inventory control is ABC classification. Inventory replenishment policy deals with 'how-much-to-order' and 'when-to-order' aspects. Purchase pattern is an important consumer insight. It is obvious that the knowledge of purchase pattern will be an important input for designing inventory replenishment policy.

Data mining is used to find new, hidden or unexpected patterns from a very large volume of historical data, typically stored in a data warehouse. Knowledge or insight discovered using data mining helps in more effective individual and group decision making. Irrespective of the specific technique, data mining methods may be classified by the function they perform or by their class of application. Using this approach, some major categories of data mining tools, techniques and methods can be identified as given below.

(i) Association Rule: Association rule is a type of data mining that correlates one set of items or events with another set of items or events. Strength of association rule is measured in the framework of support, confidence and lift [1].

(ii)Classification: Classification techniques include mining processes intended to discover rules that define whether an item or event belongs to a particular predefined subset or class of data. This category of techniques is probably the most broadly applicable to different types of business problems. Methods based on 'Decision Tree' and 'Neural Network' are used for classification.

(iii) Clustering: In some cases, it is difficult or impossible to define the parameters of a class of data to be analyzed. When parameters are elusive, clustering methods can be used to create partitions so that all members of each set are similar according to a specified set of metrics. Various algorithms like k-means, CLARA, CLARANS are used for clustering. Kohonen's map is also used for clustering.

(iv)Sequence Mining: Sequencing pattern analysis or time series analysis methods relate events in time, such as prediction of interest rate fluctuations or stock performance, based on a series of preceding events. Through this analysis, various hidden trends, often highly predictive of future events, can be discovered. GSP algorithm is used to mine sequence rules [2].

(v) Summarization: Summarization describes a set of data in compact form.

(vi) Regression: Regression techniques are used to predict a continuous value. The regression can be linear or non-linear

The author is working as Assistant Professor in Xavier Institute of Management, Bhubaneswar, PIN-751013, INDIA (phone: +91-674-3983857; fax: +91-674-2300995; e-mail: p_k_bala@rediffmail.com).

with one predictor variable or more than one predictor variables, known as 'multiple regression'.

Numerous techniques are available to assist in mining the data, along with numerous technologies for building the mining models. Data mining is used by business intelligence organizations to learn consumer behavior for the purpose of customer relationship management, fraud detection etc.

II. LITERATURE REVIEW

Inventory items have been *classified* or *clustered* in groups for the purpose of joint replenishment policy in [3], [4], [5] and [6]. Generic inventory stock control policies are derived using multi-item classification [3]. Clustering of items has also been done in production and inventory systems [4]. Multi-criteria inventory classification has been done by parameter optimization using genetic algorithm [5]. Artificial neural networks (ANNs) have been used for ABC classification of stock keeping units (SKUs). ANN has been used in a pharmaceutical company [6].

The interacting items showing the mutual increase in the demand of one commodity due to the presence of the other has been considered in the model given in [7]. Correlation between the demands of two items has been considered. The article gives a new approach towards a two-item inventory model for deteriorating items with a linear stock-dependent demand rate. It has also assumed a linear demand rate, that is, more is the inventory, more is the demand. The model has made an attempt to capture demand interdependency to some extent.

Market segmentation has been done using clustering through neural network and genetic algorithm [8]. Clustering of customers on the basis of their demand attributes, rather than the static geographic property has been done in [9]. However, the findings have not been used for the purpose of inventory management.

As part of the customer relationship management (CRM) strategy, many researchers have been analyzing 'why' customers decide to switch. However, despite its practical relevance, few studies have investigated how companies can react to defection prone customers by offering the right set of products [10]. Consumer insight has been captured in the work, but not used for inventory replenishment. For cross-selling consideration, a method to select inventory items from the association rules has been developed in [11] which gives a methodology to choose a subset of items which can give the maximal profit with the consideration of cross-selling effect. However, this does not talk about any inventory replenishment policy.

In the present research paper, a data mining model has been proposed which can be used for multi-item inventory management in retail sale stores. The model has been illustrated with an example database.

III. PURCHASE DEPENDENCIES IN RETAIL SALE IN THE CONTEXT OF INVENTORY MANAGEMENT

In the context of data mining, various aspects of purchase dependencies which may be useful for the purpose of retail inventory management have been discussed as given below.

Demand Interdependency: The problem of multi-item inventory is more challenging when there is association in the demand or usage pattern amongst the items or item-sets (Item-set refers to a set of one or more items, hence if we say item-set, it may refer to one item or a number of items.). The correlation in the demand amongst the items can be one to one, one to many, many to one or many to many. In many situations, a customer buys an item or item-set only when another item or item-set is also in stock.

To explain the above situation further, say, in a store, item B is in stock and item A is out of stock. One customer is interested in purchasing B, provided A is also available in the store, so that he can purchase both A and B. As the demand for B depends on demand for A, he will not purchase B, if A is not available. Under this situation, we can say that stock of B is as good as a stock-out situation for that customer. Hence, if A is not in stock, there will be no sale of B also in many cases. This example depicts a case of interdependency in demand. Purchase dependency between two items or itemsets can be captured in various ways and can be used for inventory management. Various aspects of purchase dependency have been discussed below.

Customer Profile and Demand Pattern: Customer profile is the detailed features of a customer. The profile may contain income, age, marital status, gender, education, number of cars, family size etc. The profile may also contain frequency of shopping, credit rating, loyalty index etc. In retail sale, demand pattern depends a lot on the customer profile along with the other factors. Hence, customer profile is an important parameter which may be used for learning the purchase pattern of a customer and this may be useful for inventory modeling. Classification and clustering can be used for learning the impact of customer profile on demand pattern. Customer profile will be a useful input for forecasting the demand of the items.

Sequence of Purchase: Many a times, a sequence of purchase gets repeated most of the times with a time gap between two purchases. A sequence may be of with two or more events of purchases. In each event of purchase, certain itemset is purchased. Once a repetitive sequence rule is identified, it can be used as an input for inventory modeling. Sequence rules are mined using GSP algorithm.

Time-Dependent Purchase Pattern: On different days of the week, there may be different purchase pattern of the customers. Purchase patterns on the weekdays and the weekends are generally different. Similarly, in different months or seasons, different patterns may be observed. Time-dependent purchase pattern may also be observed at different hours of the day. Purchase pattern in the evening hours happens to be different from the day hours. Segregated data with respect to time can be used to learn time-dependent purchase patterns. For example, data of weekend and weekdays can be segregated for this purpose.

Location-Dependent Purchase Pattern: There may be an impact of culture, climatic condition and other factors on the purchase pattern. Segregated data with respect to location or space can be used to learn location-dependent purchase patterns.

Negative Dependency: Sometimes, purchase of one item or itemset results in non-purchase of another item or itemset. Similarly, non-purchase of one item or itemset may result in purchase of another item or itemset. This kind of purchase dependency may be observed within a transaction made by a customer or in successive transactions made by the customer with a time gap.

Various purchase patterns can be mined using appropriate data mining tool as discussed in section-I.

IV. MODEL DEVELOPMENT

The model proposed has been described in figure-I. In the present research, it has been proposed to discover purchase patterns using data mining. For this purpose, sale transaction data of the inventories contained in the 'Materials' module can be used for mining association rules describing demand interdependencies. Further, time-dependent and location or space-dependent association rules can be mined by proper segregation of the past sale transaction data from the "Materials' module. Using the modules of 'Materials' and 'CRM (Customer Relationship Management)' containing the demographical and other profiles of the customers, 'classification' technique of data mining can be applied to learn the impact of customer profiles on purchase pattern. Clustering of customers can also be done. The model can integrate various transactions based on the decision and action taken in other domains of finance, maintenance, human resource management, marketing, supply chain etc. In such cases, we find association rules with a mixture of real items with 'virtual items'. Virtual items are basically events, decisions and attributes etc. representing fields in a database or a data warehouse.

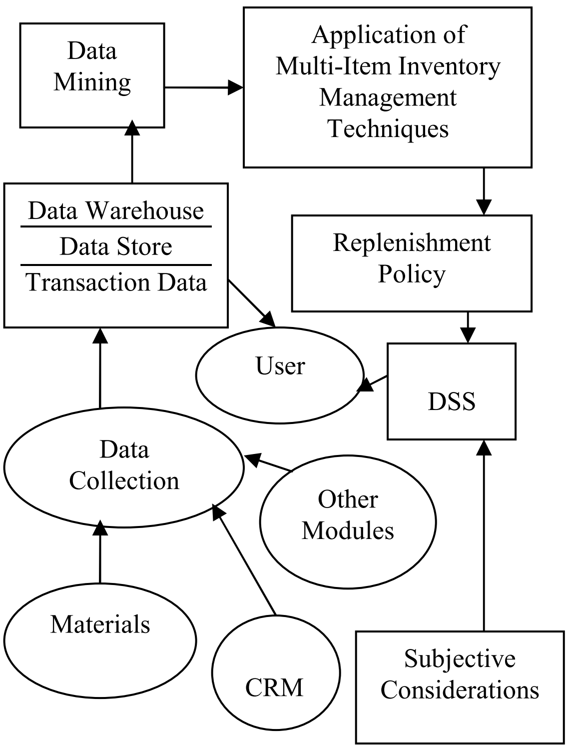

Figure – I: Block Diagram of the Model

Transaction data is stored in database which is also known as data store and useful for operational decisions. These can not address the requirements of decision support system (DSS) for inventory management. Data Mining, facilitated by data warehousing, addresses the need of strategic decision-making. Knowledge discovered from data mining is used in Decision Support System (DSS). In this model, various types of consumer insights discovered using data mining can be used in multi-item inventory modeling. The consumer insights mined act as additional input in designing the inventory replenishment policies.

The model provides scope for subjective considerations also as an input to DSS along with the objective findings obtained using data mining. The focus in this model has been to reduce the subjectivity by objective decision-making.

One example in section-V has been illustrated where association rules have been mined and classification has been done by building decision tree. A case has also been discussed in section-VI which shows purchase dependencies in the form of positive and negative association rules.

V. ILLUSTRATION WITH EXAMPLE

Eleven (11) items have been considered in a retail sale store with one thousand (1,000) past sale transaction data along with the profiles of the customers. Eleven items are named as 'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j' and 'k'. A transaction contains the items purchased by a customer along with a small profile of the customers. The profile contains 'customer id', 'value' (as rated by the retailer), 'pmethod' (payment method by cheque/cash/card etc.), 'sex' (Male/Female expressed in M/F), 'hometown' (yes/no), 'income' and 'age'. Customer id is an identification code for the customer and it finds use for capturing the sequence of purchases made by the same customer and hence in mining sequence rules. Otherwise, customer id is not of any use for mining other patterns.
To understand the database, as it is not possible to show 1,000 records of the database in this paper, five records have been shown in table – I from which one can visualize various fields in the database.

Table – I: Sale Transaction Data with Customer Profile showing Five Transactions (Part of the Database with 1,000 rows)

CUSTOMER PROFILE						ITEMS										
VALUE	PAYMENT METHOD	SEX	HOMETOWN (Y/N)	INCOME (IN $ THOUSANDS)	AGE	a	b	c	d	e	f	g	h	i	J	k
43	Ch	M	N	27	46	F	T	T	F	F	F	F	F	F	F	T
25	Cs	F	N	30	28	F	T	F	F	F	F	F	F	F	F	T
21	cs	M	N	13	36	F	F	F	T	F	T	T	F	F	T	F
24	cd	F	N	12	26	F	F	T	F	F	F	F	T	F	F	F
19	cd	M	Y	11	24	F	F	F	F	F	F	F	F	F	F	F

In table – I, 'T' (under 'Items') implies purchase of the corresponding item in the corresponding transaction and 'F' (under 'Items') implies that the corresponding item has not been purchased in the corresponding transaction. Association rules have been mined from the database of 1,000 transactions with the threshold values of support and confidence to be 10% and 80% respectively. Association rules mined have been shown in table-II.

Table – II: Association Rules

Antecedent	Consequent	Association Rule	Support %	Confidence %	Rule Support %	Lift
d and g	f	(d, g) => (f)	17.8	87.4	15.5	2.7
f and g	d	(f, g) => (d)	18.1	85.9	15.5	2.7
d and f	g	(d, f) => (g)	18.4	84.4	15.5	2.7

Only three rules have qualified for the chosen threshold values of support and confidence. In each of these three rules, we observe that same three items (i.e., d, f and g) are involved. With respect to the simultaneous purchase of all three items (i.e., d, f and g) in the same transaction, classification of the customers has been done based on their profiles. Using data mining, decision tree has been built up from the database for the purpose of classification which is given in figure-II. Considering threshold value of 80% confidence for the decision rules, only one rule (at node 3) qualifies. The rule (with 84.242% confidence) is – IF "income is less than or equal to 16950 and sex is F (Female)", THEN "the customer purchases all three items, i.e., d, f, and g". The rule has been observed in 84.242% cases.

All three of d, f, g purchased

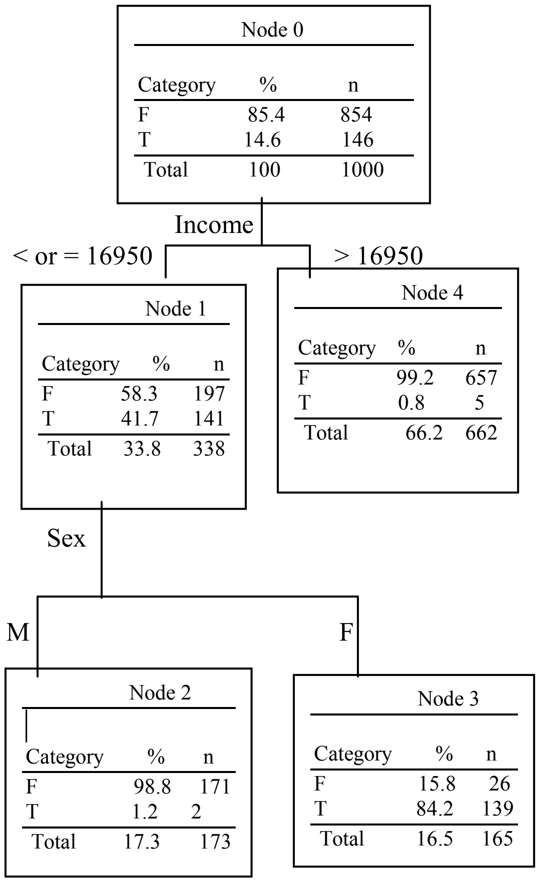

Figure-II: Decision Tree

VI. CASE DISCUSSION

The association rules and the decision rule mined from the database can be used as input for designing inventory replenishment policy. Inventory managers must consider these rules for placing the orders for replenishment.

Based on the data volume of 8.418 sale transactions of 45 selected grocery items (names of the items have been written as commonly used by the Indian shoppers) as given in table-II, association rules were mined at threshold support of 20% and threshold confidence of 65%. Total of 14 association rules, as given in table-III, qualified for these threshold values. Two association rules mined with highest confidence are, (i) Noodles – Maggi (200 g) and Tomato Sauce (Maggi) (200 g) => MDH Chicken Masala with 92.63% confidence, (ii) Kismis (100 g) => Basmati (1 kg) with 92.45% confidence.

Moreover, negative association rules were also mined from the same transaction data. Some of these rules are – (i) Basmati (1 kg) having a negative impact on the sale of Sooji (1 kg) and vice-versa, (ii) Bournvita Cadbury (500 g) having a negative impact on the sale of Poha (Chura) (1 kg) and vice-versa, (iii) Ruchi Curry Powder (50 g) having a negative impact on the sale of Noodles – Maggi (200 g) and vice-versa. These negative rules depict negative dependencies in purchase, i.e., one item inhibiting the purchase of another item by the same customer in a transaction.

VII. CONCLUSIONS

Consumer insight extracted using data mining tools has been used for business decision making in marketing, shelf design, product management etc. There has been limited use in inventory management. Purchase dependency, as a type of consumer insight which is useful for inventory designing, has been discussed for the first time in this research work. The analysis and findings in this research throws light on various aspects of purchase dependencies which may be useful for inventory management in retail stores. The model, illustrative example and the case discussion in this research paper will motivate researchers and inventory managers to develop methodologies for using the results of data mining in multi-item inventory management.

Cost impact of the purchase dependencies in inventory management can be examined by the difference in profit or cost by considering and not considering a purchase dependency in inventory replenishment. Further, this can also be examined by simulation designed for these purchase dependencies. Simulation can also be done for designing inventory replenishment policy in the context of purchase dependencies.

Table – II: List of 45 Grocery Items

Item No.	Item-Name	Description
1	All Out Liquid (35 ml)	Mosquito Repellent
2	Atta - Ashirvad (5 kg)	Flour
3	Basmati (1 kg)	Special Rice
4	Bournvita - Cadbury (500 g)	Beverage
5	Cashew Nuts (100 g)	Cashew Nuts
6	Chilli Sauce – Kissan (200 g)	Chilli Sauce
7	Coffee – Nescafe (50 g)	Beverage
8	Hair Oil - Parachute Jasmine (100 ml)	Hair Oil
9	Hair Oil - Shalimar Coconut (100 g)	Hair Oil
10	Horlicks (500 g)	Beverage
11	Kismis (100 g)	Dry grapes
12	Lux - International Soap (100 g)	Bathing Soap
13	Maida – Rishta (1 kg)	Refined Flour
14	MDH Chicken Masala	Packaged Chicken Spices
15	MDH Meat Masala	Packaged Meat Spices
16	Noodles – Maggi (100 g)	Snacks Item
17	Noodles – Maggi (200 g)	Snacks Item
18	OK Powder (500 g)	Detergent Powder
19	Parle – Hide & Seek (100 g)	Biscuit
20	Parle – Hide & Seek - Orange Cream (100 g)	Biscuit with orange cream
21	Pears Soap – Green (100 g)	Bathing Soap
22	Pepsodent G (80 g)	Toothpaste
23	Poha (Chura) (1 kg)	Foodstuff
24	Pond's Dreamflower Powder (100 g)	Body Powder
25	Priya Pickles – Mango (200 g)	Pickles
26	Refined Oil – Safflower – fora (1 l)	Refined Oil
27	Rin Supreme (100 g)	Washing Soap
28	Ruchi Chilli Powder (100 g)	Chilli Powder
29	Ruchi Chilli Powder (50 g)	Chilli Powder
30	Ruchi Curry Powder (50 g)	Curry Powder
31	Ruchi Haldi powder (100 g)	Turmeric Powder
32	Salt - Tata (1 kg)	Salt
33	Shampoo - Clinic All Clear (50 ml)	Shampoo
34	Shampoo - Clinic Plus (50 ml)	Shampoo
35	Shaving Cream - Old Spice (100 g)	Shaving Cream
36	Sooji (1 kg)	Foodstuff
37	Sugar (1 kg)	Sugar
38	Sundrop Refined Oil (1 l)	Refined Oil
39	Surf Excel (500 g)	Detergent Powder
40	Tea - Tata Leaf (100 g)	Beverage
41	Tide Detergent (200 g)	Detergent Powder
42	Tide Detergent (500 g)	Detergent Powder
43	Tomato Sauce (Lalls) (200 g)	Tomato Sauce
44	Tomato Sauce (Maggi) (200 g)	Tomato Sauce
45	Vim Bar (100 g)	Utensil Cleaning Soap

Table – III: Association Rules

Rule #	Antecedent	Consequent
1	Noodles – Maggi (200 g). and Tomato Sauce (Maggi) (200 g).	MDH Chicken Masala.
2	MDH Chicken Masala.	Noodles – Maggi (200 g). and Tomato Sau
3	MDH Chicken Masala. and Noodles – Maggi (200 g).	Tomato Sauce (Maggi) (200 g).
4	Tomato Sauce (Maggi) (200 g).	MDH Chicken Masala. and Noodles – Maggi (200 g)
5	Kismis (100 g).	Basmati (1 kg).
6	Tomato Sauce (Maggi) (200 g).	MDH Chicken Masala.
7	MDH Chicken Masala.	Tomato Sauce (Maggi) (200 g).
8	Coffee – Nescafe (50 g).	Bournvita - Cadbury (500 g).
9	MDH Chicken Masala. and Tomato Sauce (Maggi) (200 g).	Noodles – Maggi (200 g).
10	Poha (Chura) (1 kg). and Sooji (1 kg).	Noodles – Maggi (200 g).
11	Poha (Chura) (1 kg).	Noodles – Maggi (200 g).
12	Sooji (1 kg).	Noodles – Maggi (200 g).
13	MDH Chicken Masala.	Noodles – Maggi (200 g).
14	Tomato Sauce (Maggi) (200 g).	Noodles – Maggi (200 g).

REFERENCES

[1] Han, J., Kamber, *"Data Mining: Concepts and Techniques"*, Morgan Kaufmann Publishers, 2006., pp. 227-284.

[2] Pujari, A.K., *"Data Mining Techniques"*, University Press, 2002, pp. 259-260.

[3] Cohen, M.A. & Ernst, R., "Multi-item classification and generic inventory stock control policies", Production and Inventory Management Journal, 29(3), 1988, pp. 6-8.

[4] Ernst, R. & Cohen, M.A., "A clustering procedure for production/inventory systems", Journal of Operations Management, 9(4), 1990, pp..574-598.

[5] Guvenir, H.A. & Erel, E., "Multicriteria inventory classification using a genetic algorithm", European Journal of Operations Research, 105(1), 1998, pp. 29-37.

[6] Partovi, F.Y. & Anandarajan, M., "Classifying inventory using an artificial neural network approach. Computers & Industrial Engineering", 41, 2002, pp. 389-404.

[7] Bhattacharya, D.K., "Production, Manufacturing and Logistics on Multi-Item Inventory", European Journal of Operational Research 162 (3), 2005, pp. 786–791

[8] Kuo, R.J., An, A.L., Wang, H.S., Chung, W.J., "Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation.", Expert Systems with Applications, 30, 2006, pp. 313–324.

[9] Hu, T. L., Sheu, J.B., "A fuzzy - based customer classification method for demand-responsive logistical distribution operations", Fuzzy Sets and Systems 139, 2003, pp. 431–450.

[10] Larivie`re, B., Piel, D.V.D., "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services", Expert Systems with Applications 27, 2004, pp. 277–285.

[11] Wong, W., Fu, A.W., and Wang, K., "Data Mining for Inventory Item Selection with Cross-Selling Considerations", Data Mining and Knowledge Discovery, 11 (1), 2005, pp. 81-112.