Decision making with association rule mining and clustering in supply chains

Reshu Agarwal[a*]

[a]G L bajaj Institute of Technology and Management, Greater Noida, India

CHRONICLE	ABSTRACT
	This paper deals with data mining applications for the supply chain inventory management. ABC characterization is typically utilized for stock items arrangement on the grounds that the quantity of stock items is large to the point that it is not computationally practical to set stock and administration control rules for every individual item. Moreover, in ABC classification, the inter-relationship between items is not considered. But practically, the sale of one item could influence the sale of other items (cross selling effect). Consequently, within each cluster, the inventories should be classified. In this paper, a modified approach is proposed considering both cross-selling effect and clusters to rank stock items. A numerical case is utilized to clarify the new approach. It is represented that by utilizing this modified approach; the ranking of items may get influenced bringing about higher profits.

1. Introduction

In numerous stock control frameworks, it has been viewed as that the quantity of items is large to the point that it is not computationally doable to set stock and administration control rules for every individual item. Subsequently, items are regularly gathered together and generic stock control strategies are connected to everything in a gathering. Such gathering techniques give administration more successful means for determining, checking, and controlling framework execution, since procedure destinations and association elements can frequently be spoken to all the more normally in the terms of item gatherings.

Further, the ABC classification scheme is the most utilized technique for items gathering. ABC classification is based on Pareto Analysis which says 20% of the items contribute to 80% of sales. It implies that a small portion of items in Inventory contribute to maximum sales. Typically less than 20% of items classified as A, contribute as much as 80% of the revenue. The next 15% (80% - 95%) contribution to revenue is done by B class Items. The last 5 % revenue is generated by items classified as C. However, for many items, ABC classification is not suitable for inventory control.

Managers need to move a few items among classifications for various reasons. A few analysts considered there might be other criteria that speak to critical contemplations for administration. The sureness of supply, the rate of obsolescence, and the effect of a stock out of the thing are generally cases of such contemplations. Some of these may even measure more vigorously than dollar utilization in the administration of the items. A few criteria have been recognized as critical in the administration of upkeep inventories (Chase et al., 1998). Likewise, a few scientists proposed the numerous criteria ought to be utilized as a part of the classification of inventories (Flores & Whybark, 1987; Cohen & Ernst, 1988; Lenard & Roy, 1995). Ramanathan (2006) solves a linear programming problem for each item in inventory to determine weights that maximize the weighted score for that item subject to constraints that the weighted sum for every item using this same set of weights is less than or equal to one. Thus, one immediate criticism of this model is that with more than a handful of items, the process will become cumbersome and time-consuming. Ng (2007) addresses this issue by proposing model like Ramanathan's, yet which is at that point changed into another arrangement of issues, the structure of which makes it simple to perceive the ideal arrangement. Zhou and Fan (2007) proposed an extended version of such weighted linear optimization for multi-criteria stock classification. Ozan and Mustafa (2008) proposed a stock grouping framework based on the fuzzy AHP to help a sensible multi-criteria stock arrangement. However, the issue is that the profit of one item originates from its own deals, as well as from its impact on the sales of other items i.e., the cross-selling effect (Anand et al., 1997). In such a circumstance, it ought to be clarified obviously whether the cross-selling effect would impact the ranking of items or not, and how to aggregate the items if such impacts existed, not concerning what and what number of criteria could be utilized.

Recently, Brijs et al. (1999, 2000) introduced a PROFSET model to account the impacts of cross-selling effects among items. The biggest commitment of the PROFSET model is that it demonstrates to figure the profit of a frequent item-set. However, the PROFSET model does not think about the strength of the relationship between items, so it gives no relative ranking of selected things, which is critical in arrangement of inventories. In addition, to compute the profit of a frequent item set the maximal frequent item set had been utilized. Unfortunately, the maximal frequent item-set regularly does not reflect "buy goals" since they may not happen as every now and again as their subsets. In this way, the PROFSET model can't be utilized to classify stock items on the grounds that frequent items as well as all of stock items ought to be ranked. That implies the rule of dollar uses of few frequent items must be utilized to choose some unique items and isn't fitting for all stock items classification. Cheng (2005) built up a model in which strength of association relationship is considered for bookkeeping the related benefit.

However, there is little research on the best way to boost benefit when the environment changes. By utilizing the association rule, which is a standout amongst the most mainstream procedures in KDD, another approach of ranking items with the thought of the cross-selling effect has been introduced (Kaku, 2004; Kaku & Xiao, 2008). In any case, they have not considered that whether and how the quality of association with related items impacts such ranking approach. Bala (2008, 2012) recommended a model for making utilization of customer understanding data for stock administration in retail locations. Later an examination on purchase dependence association rules for retail items was recommended by Bala et al. (2010) to take stock replenishment decisions. According to their perception, in a multi-item retail stock of huge number of items, purchase dependence among the things is watched much of the time and when there is stock out of one item, it might come about the decrease in buy of another item. Moreover, to forecast demand and evaluating ordering policy, cross-selling effect ought to likewise be considered. Xiao et al. (2011) characterized stock items which are related each other utilizing the idea of cross-selling effect together with ABC classification and loss profit. They introduce a calculation for ranking all of stock items, to help stock manager in really perceiving most beneficial items. They grouped items on basis of loss rule (Wong et al. 2003, 2005). The loss profit of item/item set is characterized as the basis for evaluating the significance of item, in view of which stock items are grouped. They disclosed that to judge the significance of an item (set), it is not

just by taking a gander at the benefit it gets when it is on the rack, yet in addition the loss profit it might take away when it is missing or stock out. To represent the strength of the relationship between items, a profit ranking approach of items based on a ''Hub-Authority'' analogy was exploited in Wang and Su (2002).

Such similarity is introduced as the relationship of hubs and authorities received in the web pages ranking calculation HITS (Kleinberg, 1998), which is utilized as a part of the outstanding search engine Google. Further, Mittal et al. (2014, 2015a, 2015b, 2016) determines ordering policy using association rule mining. However, because of the complexity of interrelationships among items, there is small amount of research that treats decision making with correlation of inventory items. Hence, how to treat correlation is a challenge when developing inventory strategies.

In this paper, how to classify inventory items which are correlated each other is discussed by using the concept of cross-selling and clustering. A numerical example is used to explain the new approach. Result suggests that many items that traditionally belonged to the B or C group were moved into the A group by the cross-selling effect within each cluster.

2. Proposed Approach

This paper proposes to classify inventory items using association rule mining, cross-selling effect and clustering.

Cross-selling is a method of proposing related items or services to a customer. The benefit of an item does originate from itself as well as from different items that impact its deal. Consequently, there are more odds of losing deal if cross-selling among items is more. The cross-selling effect among items can be dictated by utilizing association rules. Association rule mining finds fascinating associations and/or correlation among large set of data items. Let $I = \{i_1, i_2, i_3, i_4, \ldots .. i_m\}$ be an arrangement of items. Further, support of item i1 is characterized as the recurrence of its occurrences in total transactions and confidence is characterized as conditional probability of acquiring i_2 when i_1 is obtained and is given by formula:

$$\text{Support}(i_1) = \frac{\text{Frequency of } i_1}{\text{Total number of Transactions}}, \tag{1}$$

$$\text{Confidence}(i_1 \rightarrow i_2) = \frac{\text{Support of } i_1 \cup i_2}{\text{Support of } i_1}. \tag{2}$$

Apriori algorithm is utilized to produce association rules whose support and confidence is greater than user-defined minimum support and minimum confidence (Srikant & Agrawal, 1995). The frequent item-set is determined based on minimum support and association rules are created based on minimum confidence.

Further, clustering refers to the partitioning of a collection of transactions into clusters such that similar transactions are in the same cluster and dissimilar transactions are in different clusters. Here, the term "large items" refers to the items contained in some minimum fraction of transactions in a cluster. Large items are used as a similarity measure of a cluster of transactions. The support of an item in cluster Ci is the number of transactions in Ci. Thus, for a user-specified minimum support s, an item is large in cluster Ci if its support is at least equal to s × Ci, otherwise item is small. Consequently, large items add to likeness in a group while small items add to disparity. The cost Ç to be minimized consists of two segments: the intra-cluster cost and the inter-cluster cost. The intra-cluster cost is estimated by the aggregate number of small items and the inter-cluster cost measures the duplication of large items in various clusters. This clustering algorithm expects to minimize the cost because of large items and small items (Wang et al., 1999). The outline of the clustering algorithm is depicted in Fig. 1.

```
/* Allocation phase */
   (1) while not end of the file do
   (2) read the next transaction <t,- >;
   (3) allocate t to an existing or a new cluster Ci ;
   (4) write  <t, Ci >;
/*  Refinement phase */
   (5) repeat
   (6) not_ moved = true;
   (7) while not end of the file do
   (8) read the next transaction < t, Ci >;
   (9) move t to an existing cluster Cj to minimize Cost C;
   (10) if Ci ≠ Cj then
   (11) write < t, Cj  >;
   (12) not_moved=false;
   (13) eliminate any empty cluster;
   (14) until not moved;
```

Fig. 1. Outline of Clustering Algorithm

Further, Yin et al. (2011) recommend another foundation of expected dollar utilization (EDU) to judge the significance of item(s), and serve as the evaluating index to rank inventory items in inventory classification. To figure the EDU of an item, frequent item-set is dealt with as a unique item, the dollar utilization of which can be computed like a normal item. Then, all frequent item-sets will be considered in the ranking process together with individual items.

Consider an item-set X in an inventory transaction database Db with support sup(X), then

$$sup(X) = \frac{|X(t)|}{D}$$

where X(t) = {t in D/t contains X} and t is a transaction.

Now, EDU can be calculated as follows:

$$Cx =   Sup(X) \sum_{i \in X} p_i$$

where Cx is the EDU of an item-set X, pi is the price of single item in X and $\sum_{i \in X} p_i$ reflects the set's price.

The working of calculation can be clarified in five stages:

Stage 1: Firstly, the database is partitioned into different clusters.

Stage 2: Calculate all frequent item-sets in a transaction database by utilizing frequent item-sets searching algorithm.

Stage3: Calculate the EDU of all frequent item sets and of all single items. Then, rank them in descending order beginning with the largest value.

Stage4: Replace all frequent item-sets in the ranking list with their contained items, the interior order of which must remain. From that point onward, scan the ranking list from the earliest starting point to the end and pull back the repeated items that have showed up for the second time to make each of the

items unique in the list.

Stage 5: Make the ABC classification of inventory in light of the new ranking list. The new A group is acquired by choosing items of the new ranking list from the earliest starting point to the end till the aggregate EDU of selected items achieves 80% of aggregate dollar usage occupancy.

3. Numerical Example

In this section, Yin et al. (2011) model parameters have been considered to calculate the EDU of items. Further, EDU has been calculated for inventory items in each cluster which was not considered by Yin et al. (2011).

Consider the database set D and the inventory item-set, I = {a, b, c, d, e, f, g, h, i}. The inventory transaction set, TID = {TID1, TID2, TID3, TID4, TID5, TID6} as shown in Table 1. Each row in Table 1 can be taken as an inventory transaction.

Table 1
An inventory transaction database

TID	ITEMS						
TID1	a	b	c				
TID2	a	b	c	d			
TID3	a	b	c		e		
TID4	a	b			f		
TID5				d		g	h
TID6				d		g	i

Now, we determine inventory classification by using rules obtained by apriori algorithm after clustering the data. Consider the transaction database of Table 1. It is assumed that the user-specified minimum support is 60%. A large item must be contained in at least 4 transactions (i.e., 6 *60%). Further, consider the clustering $Ç_1 = \{C_1 = \{tid1, tid2, tid3, tid4, tid5, tid6\}\}$. We have $Large_1 = \{a, b\}$, $Small_1 = \{c, d, e, f, g, h, i\}$. Intra $(Ç_1) = 7$, and Inter $(Ç_1) = 0$. So Cost $(Ç_1) = 7$.

Again, consider the clustering $Ç_2 = \{C_1 = \{tid1, tid2, tid3, tid4\}, C_2 = \{tid5, tid6\}\}$. For C_1, a large item should be contained in at least 3 transactions in C_1. Now, $Large_1 = \{a, b, c\}$ and $Small_1 = \{d, e, f\}$. Similarly, $Large_2 = \{d, g\}$ and $Small_2 = \{h, i\}$. Hence, Intra $(Ç_2) = 5$, Inter $(Ç_2) = 0$, and Cost $(Ç_2) = 5$. Thus $Ç_2$ has less cost as compared to $Ç_1$.

Further, consider the clustering $Ç_3 = \{C_1 = \{tid1, tid2\}, C_2 = \{tid3, tid4\}, C_3 = \{tid5, tid6\}\}$. We have $Large_1 = \{a, b, c\}$, $Small_1 = \{d\}$, $Large_2 = \{a, b\}$, $Small_2 = \{c, e, f\}$, $Large_3 = \{d, g\}$, $Small_3 = \{h, i\}$. Intra $(Ç_3) = 6$, and Inter $(Ç_3) = 2$. Hence Cost $(Ç_3) = 8$, which is larger than $Ç_2$.

Thus, we will consider cluster $Ç_2$, as it has minimum cost as compared to cluster $Ç_1$ and $Ç_3$. Hence, the transaction database of table 1 is clustered into two clusters consisting of $C_1 = \{tid1, tid2, tid3, tid4\}$ and $C_2 = \{tid5, tid6\}$. We apply apriori algorithm on both clusters. We find item-set {a, b, c} is the most frequent item-set in cluster C_1 and item-set {d, g} is the most frequent item-set in cluster C_2.

Let us consider the prices of items, a = $4, b=$5, c=$3, d=$5, e=$1, f=$1, g=$8, h=$2, and i=$1. By using traditional ABC classification, let the dollar usage of items be a=$16, b=$20, c=$9, d=$15, e=$1, f=$1, g=$16, h=$2, i=$1. Then rank items in descending order starting with the largest value of dollar usage; the ranking list is (bagdchefi).

Consider cluster C1. Let minimum support =2. Table 2 shows the generation of all large frequent item-sets in cluster C1.

Table 2

The progress of finding frequent item sets in cluster C1

Id	{Items}	Length	Support	EDU	Is it Frequent?
One-large item-set					
1	{a}	1	4	4*4=16	Y
2	{b}	1	4	4*5=20	Y
3	{c}	1	3	3*3=9	Y
4	{d}	1	1	1*2=2	N
5	{e}	1	1	1*1=1	N
6	{f}	1	1	1*1=1	N
Two-large item-set					
7	{ab}	2	4	4*(4+5)=36	Y
8	{ac}	2	3	3*(4+3)=21	Y
9	{bc}	2	3	3*(5+3)=24	Y
Three-large item-set					
10	{abc}	3	3	3*(4+5+3)=24	Y

By ranking these individual items and frequent item-sets of Table 2 in descending order starting with the largest value of EDU, the following list of item-sets can be obtained:

{ab},{ac}, {abc}, {bc}, {a}, {b}, {c}

According to the proposed algorithm, replace the frequent item-sets with their elements to get the ranking list of items as follows:

{abacabcbcabc}.

Finally, the ranking list is scanned from the beginning to the end. All repeated items are withdrawn. The ranking list as below:

{abc}.

Consider cluster C2. Let minimum support = 2. Table 3. shows the generation of all large frequent item-sets in cluster C2.

Table 3

The progress of finding frequent item sets in cluster C2

Id	{Items}	Length	Support	EDU	Is it frequent?
One-large item-set					
1	{d}	1	2	2*5=10	Y
2	{g}	1	2	2*8=16	Y
3	{h}	1	1	1*2=2	N
4	{i}	1	1	1*1=1	N
Two-large item-set					
5	{dg}	2	2	2*(5+8)=26	Y

By ranking these individual items and frequent item-sets of Table 3 in descending order starting with the largest value of EDU, the following list of item-sets can be obtained:

{dg},{g},{d}

According to the proposed algorithm, replace the frequent item-sets with their elements to get the ranking list of items as follows:

{dggd}.

Finally, the ranking list is scanned from the beginning to the end. All repeated items are withdrawn. The ranking list as below:

{QRTSPU}.

4. Observations

1. Consider cluster C1, we obtain ranking list as {abc}. It is a very different order compared with the result {bac} of the traditional ABC classification. Item a is moved to higher position in the list, even though it has a smaller self-profits. That means the stock out of item a may result a larger loss than itself because the strong cross-selling relationship withother valuable items.

2. Consider cluster C2, we obtain ranking list as {dg}. It is a very different order compared with the result {gd} of the traditional ABC classification. Results indicate that a considerablelarge part of inventory items change their positions when they are ranked according to EDU as compared to traditional ABC classification in each cluster. Some items that traditionally do not belong to the A group in each cluster have been moved into the group A by the cross-selling effect to reconfigure their inventory policies, and also some items that traditionally belong to C group in each cluster have been promoted into higher group because of their high values of loss profits and should not be ignored as these were treated before. This approach assists inventory manager torecognize high profit items in each cluster, so that he earn profit and easily manage stocks.

5. Conclusion and Future Scope

Inventory Management is critical for most organizations, however is particularly significant for private ventures since when contrasted with huge organizations, they more often have constrained assets and haggling power, which effect sly affect the way inventory can be overseen. This paper introduces an approach of ranking inventory items. By utilizing the association rule mining and clustering, the cross-selling effect is brought into classification of inventories. A numerical example was presented to illustrate the new approach. Under this classification scheme, a few items with small value of itself might be thought to be a significant item since they can impact the sales of some other items with bigger dollar utilization in each cluster. Therefore, a few items being viewed as paltry in the old ABC characterization might be viewed as more essential and their positions are changed in this inventory ranking list. Thus, the manager can easily perceive ideal request amount of most gainful items in each one. Future investigations ought to create forecasting models that incorporate information of inventory policies in the frequent item-sets in light of the fact that the items in a frequent item-set might be from various classes and relate with each other.

References

Anand, S. S., Hughes, J. G., Bell, D. A., & Patrick, A. R. (1997).Tackling the cross- sales problem using data mining. *Proceedings of the 2nd Pacific-Asia Conference on Knowledge Discovery & Data Mining* (pp. 331-343). Hongkong.

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (1999).Using association rules for product assortment decisions: A case study. *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge discovery & data mining (pp. 254-260).* New York, USA.

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (2000). A data mining framework for optimal product selection in retail supermarket data: The generalized PROFSET model. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 300-304).* New York, USA.

Bala, P. K. (2008). Identification of purchase dependencies in retail sale. *International Journal of Strategic Management*, 8(2), 75–83.

Bala, P. K., Sural, S., & Banerjee, R. N. (2010). Association rule for purchase dependence in multi-item inventory. *Production Planning and Control*, *21*(3), 274–285.

Bala, P. K. (2012). An inventory replenishment model under purchase dependency in retail sale. *International Journal of Computer Applications*, *37*(10), 43–48.

Chase, R. B., Aquilano, N. J., & Jacobs, F. R. (1998). *Production & operations management manufacturing & services (8thed.).* McGraw-Hill companies.

Cheng, Y. (2005). Genetic algorithm for item selection with cross-selling considerations. *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*. China. (pp. 3293-3298).

Cohen, M. A., & Ernst, R. (1988). Multi-item classification and generic inventory stock control Policies. *Production & Inventory Management Journal*, 29(3), 6–8.

Flores, B. E., & Whybark, D. C. (1987). Implementing multiple criteria ABC analysis. *Journal of Operations Management*, *7*(1 & 2), 79–85.

Kaku, I. (2004). A data mining framework for classification of inventories. *Proceedings of the 5thAsia Pacific Industrial Engineering & Management Systems (pp. 450-455).* Japan.

Kaku, I., & Xiao, Y. (2008). A new algorithm of inventory classification based on the association rules. *International Journal of Services Sciences*, *1*(2), 148–163.

Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms(pp. 668-677).*San Francisco, California, USA.

Lenard, J. D., & Roy, B. (1995). Multi-item inventory control: A multicriteria view. *European Journal of Operational Research*, *87*(3), 685–692.

Mittal, M., Pareek, S., & Agarwal, R. (2014). Efficient ordering policy for imperfect quality items using association rule mining. *In Encyclopedia of information science and technology (3rd ed., pp. 773–786).* United states: Information Science publishing.

Mittal, M., Pareek, S., & Agarwal, R. (2015a). Loss profit estimation using association rule mining with clustering. *Management Science Letters*, *5*(2), 167–174.

Mittal, M., Pareek, S., & Agarwal, R. (2015b). Ordering policy using temporal association rule mining. *International Journal of Data Science*, *1*(2), 157-171.

Mittal, M., Pareek, S., & Agarwal, R. (2016*). Loss profit estimation using temporal association rule mining. International Journal of Business Analytics*, *3*(1), 45-57.

Ng, W. L. (2007). A simple classifier for multiple criteria ABC analysis. *European Journal of Operational Research, 177*(1), 344-353.

Ozan, C., & Mustafa, S.C. (2008). A web-based decision support system for multi-criteria inventory classification using fuzzy AHP methodology. *Expert Systems with Applications*, *35*(3), 1367–1378.

Ramanathan, R. (2006). ABC inventory classification with multiple-criteria using weighted linear optimization. *Computer and Operations Research*, *33*(3), 695-700.

Srikant, R., & Agrawal, R. (1995). Mining generalized association rules. *Proceedings of the 21th International Conference on Very Large Data Bases (pp. 407–419)*. Zurich, Switzerland.

Wang, K. and Su, M.Y. (2002). Item Selection by "Hub-Authority" Profit Ranking. *Proceedings* of the 8th *ACM SIGKDD* International Conference on Knowledge Discovery and Data Mining *(pp. 652-657).* Ed- monton, Alberta, Canada

Wang, K., Xu, C. & Liu, B. (1999). Clustering transactions using large item. *ACM CIKM International Conference on Information and Knowledge Management*, New York (pp. 483-490).

Wong, R. C., Fu, A. W., & Wang, K. (2003).MPIS: Maximal-profit item selection with cross-selling considerations. *IEEE International Conference on Data Mining (pp. 371-378).* Florida, USA.

Wong, R. C., Fu, A. W., & Wang, K. (2005). Data mining for inventory item selection with cross-selling consideration. *Data Mining and Knowledge Discovery*, *11*(1), 81–112.

Xiao, Y., Zhang, R., & Kaku, I. (2011). A new approach of inventory classification based on loss profit. *Expert Systems with Applications*, *38*(8), 9382–9391.

Yin, Y., Kaku, I., Tang, J., & Zhu, J. M. (2011). *Data Mining Concepts, Methods and Applications in Management and Engineering Design.* London: Springer.

Zhou, P., and Fan, L. (2007). A note on multi-criteria ABC inventory classification using weighted linear optimization, *European Journal of Operational Research*, *182*(3), 1488-1491.