

## EOQ estimation for imperfect quality items using association rule mining with clustering

Mandeep Mittal<sup>a\*</sup>, Sarla Pareek<sup>b</sup> and Reshu Agarwal<sup>b</sup>

<sup>a</sup>Department of Computer Science Engineering, Amity School of Engineering and Technology, Bijwasan, New Delhi – 110061, India

<sup>b</sup>Apaji Institute of Mathematics and Applied Computer Technology, Banasthali University, Rajasthan-304022, India

### CHRONICLE

*Article history:*  
Received February 9, 2015  
Received in revised format:  
May 12, 2015  
Accepted May 28, 2015  
Available online  
May 30 2015

*Keywords:*  
Data mining  
Apriori algorithm  
Clustering  
EOQ  
Imperfect quality items

### ABSTRACT

Timely identification of newly emerging trends is needed in business process. Data mining techniques like clustering, association rule mining, classification, etc. are very important for business support and decision making. This paper presents a method for redesigning the ordering policy by including cross-selling effect. Initially, association rules are mined on the transactional database and EOQ is estimated with revenue earned. Then, transactions are clustered to obtain homogeneous clusters and association rules are mined in each cluster to estimate EOQ with revenue earned for each cluster. Further, this paper compares ordering policy for imperfect quality items which is developed by applying rules derived from apriori algorithm viz. a) without clustering the transactions, and b) after clustering the transactions. A numerical example is illustrated to validate the results.

© 2015 Growing Science Ltd. All rights reserved.

## 1. Introduction

As information technology (IT) progresses rapidly, its capacity to store and manage data in databases is becoming important. Though IT development facilitates data processing and eases demands on storage media, extraction of available implicit information to aid decision making has become a new and challenging task. Data mining has thus been emerging as a powerful new technology to analyze and extract hidden potential information from huge volume of data. Data mining techniques, such as clustering, association rule mining, and classification, have attracted remarkable attention during the past few years (Zhao & Bhowmick, 2003).

Further, association rule mining is an important component of data mining. It helps in finding regularities/patterns in data. Extracting association rules is the core of data mining (AL-Zawaidah et al., 2011). It is mining for association rules in database of sales transactions between items which is important field of the research in dataset (Han & Kamber, 2000). The benefits of these rules are detecting unknown relationships, producing results which can perform basis for decision making and

\* Corresponding author.

E-mail address: [mittal\\_mandeep@yahoo.com](mailto:mittal_mandeep@yahoo.com) (M. Mittal)

prediction. Apriori algorithm is a very popular algorithm for mining association rules. The goal of apriori algorithm is to find all rules satisfying some basic requirement such as the minimum support and the minimum confidence. Moreover, clustering is a technique for grouping objects on the basis of similarity. Broder et al. (1997) defined clusters as maximal connected components of some pair-wise similarity of transactions, thus suffers from the breakdown of the transitivity of pair-wise similarity. Guha et al. (2000) proposed the common neighbors of two transactions as a measure of pair-wise similarity. Wang et al. (1999) did not use any notion of pair-wise similarity. They clustered transactions that contain similar items. The difference is that clustering emphasizes the dissimilarity of clusters. The rationale behind clustering transactions prior to mining association rules is that the latter is performed on partitions that are essentially distinct from each other. Both association rule mining and clustering techniques helps in effective inventory management.

Inventory management is a system used to oversee the flow of products and services in and out of an organization. It usually involves monitoring the transfer of units in a company to prevent the inventory from increasing too high, or to dwindling to levels that could put the operation into jeopardy. In other words, effective inventory management is to keep a product in its right place at the right time and in the right quantity. However, sometimes there are defective items that should also be considered for more realistic situation. Defective items can arise due to man-handling or due to technical fault. As a result, to properly ascertain the role of defective items in inventory management, researchers have devoted a great amount of efforts in developing EOQ models. Porteus (1986) incorporated the effect of imperfect quality items into the basic economic order quantity model. He assumed that there was some probability that the process would go out of control while producing one unit of the product. Rosenblatt and Lee (1986) proposed an EOQ model for a production system which contains defective production and concluded that the presence of defective products motivates smaller lot sizes. Later, Lee and Rosenblatt (1987) considered using process inspection during the production run so that the shift to out-of-control state can be detected and restoration made earlier. Salameh and Jaber (2000) developed an economic order quantity model where each order contains a random fraction of imperfect quality items with a known probability distribution. Papachristos and Konstantaras (2006) examined the work accomplished by Salameh and Jaber (2000) and rectified the proposed conditions to ensure that shortages would not occur. Maddah and Jaber (2008) corrected Salameh and Jaber (2000) work associated with the method of evaluating the expected profit per unit time. Jaggi et al. (2011, 2012, 2013) formulated an inventory model for deteriorating items. They assumed that the screening rate is more than the demand rate. This assumption helps one fulfill the demand, out of the products which are found to be of perfect quality, along with the screening process. Jaggi and Mittal (2011, 2012) developed an inventory model with joint effect of inspection, deterioration, time-dependent demand, inflation and time value of money. However, for some inventory items, the criteria (such as the price of an item) are derived not only from themselves, but also from their influence on the criteria of other items, usually called the “cross-selling effect” (Anand et al., 1997). Thus, items should be classified while considering such relationships. Kaku (2004, 2008) extended economic order quantity model for perfect items considering cross-selling effect. Mittal et al. (2014) extended economic order quantity model for imperfect quality items considering time expressions into association rules.

However, none of the researcher has explored so far the joint effect of cross-selling effect, association rule mining and clustering. Motivated with this aspect, this paper focuses on EOQ estimation using apriori algorithm with clustering and without clustering the transactions. The study shows that the combined use of association rules and clustering methods was more relevant. This approach brings an important increase in the number of rules produced, which eventually results in higher expected profit to the retailer’s side. The results have been validated with the help of a numerical example.

The remaining parts of this paper are organized as follows. Notations used in this paper are defined in Section 2. EOQ estimation considering imperfect quality items and cross-selling effect is proposed in Section 3. An example to illustrate the proposed approach is given in Section 4. Observations are stated in Section 5. Conclusions are finally given in Section 6.

## 2. Notations

The following notations are used:

I	set of items,
$C_i$	the $i^{\text{th}}$ cluster,
s	the support of an item,
$\zeta$	cost due to large and small items,
$OC_r$	opportunity cost,
$Z_r$	index,
y	order size,
w	purchasing price per unit,
K	ordering cost per order,
$f(\alpha)$	probability density function of $\alpha$ ,
$E[\alpha]$	expected value of $\alpha$ , which is equal to $\int_a^b \alpha f(\alpha) d\alpha$ $0 < a < b < 1$ ,
p	selling price of good quality item,
$c_s$	selling price of imperfect quality item,
c	purchase cost per unit,
h	holding cost per unit per unit time,
$H_r$	holding cost of item r per unit,
T	duration of a cycle,
D	demand rate,
$\beta$	unit screening cost,
$\lambda$	screening rate,
t	screening time,
Rev	revenue earned,
TC	total cost,
$TC(y)$	total cost per cycle of length t,
$P(y)$	present worth of total profit,
$E[P(y)]$	present worth of expected profit,
TID	inventory transaction set.

## 3. Proposed work

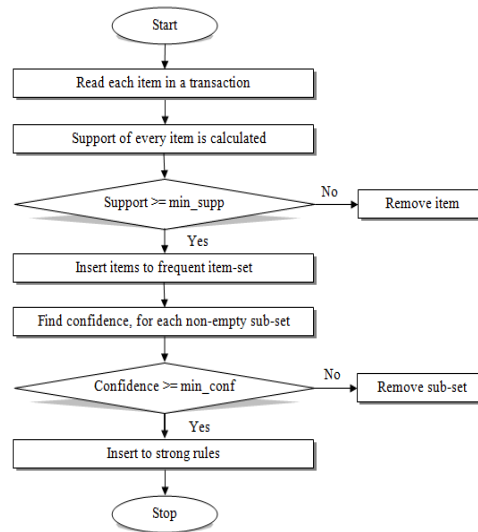
This paper proposes a comparison between order quantity for imperfect quality items in frequent item-sets, considering cross-selling effect, with and without clustering the transactions.

Cross-selling is a technique of suggesting related products or services to a customer. The profit of a product does not only come from itself but also from other products that influence its sale. Hence, there are more chances of losing sale if cross-selling among items is more. The cross-selling effect among items can be determined by using association rules. Association rule mining finds interesting associations and/or correlation among large set of data items. Let  $I = \{i_1, i_2, i_3, i_4, \dots, i_m\}$  be a set of items. Now, support of item  $i_1$  is defined as the frequency of its occurrences in total transactions and confidence is defined as conditional probability of purchasing  $i_2$  when  $i_1$  is purchased and is given by formula:

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

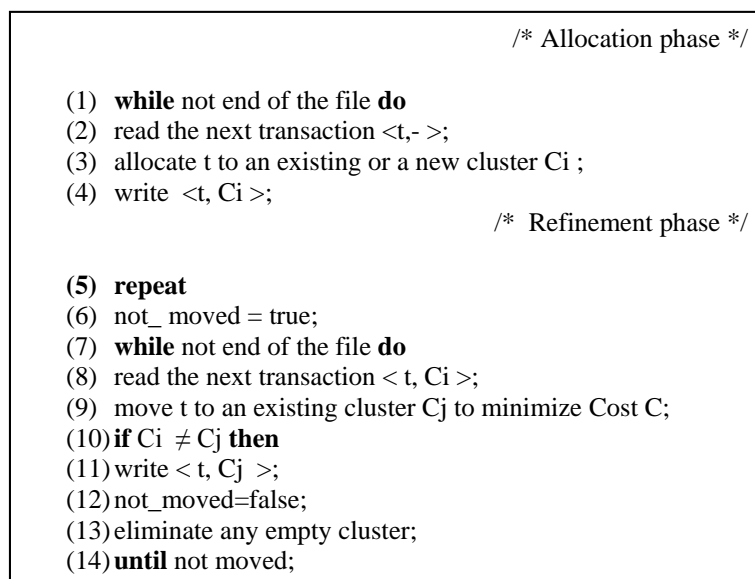
$$\text{Confidence}(i_1 \rightarrow i_2) = \frac{\text{Support of } i_1 \cup i_2}{\text{Support of } i_1} \quad (2)$$

Apriori algorithm is used to generate association rules whose support and confidence is greater than user-defined minimum support and minimum confidence (Agrawal & Srikant, 1994). The frequent item-set is determined on the basis of minimum support and association rules are generated on the basis of minimum confidence. The flowchart of apriori algorithm is depicted in Fig. 1.



**Fig. 1.** Flowchart of apriori algorithm

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  $C_i$  is the number of transactions in  $C_i$ . Thus, for a user-specified minimum support  $s$ , an item is large in cluster  $C_i$  if its support is at least equal to  $s \times C_i$ , otherwise item is small. Thus, large items contribute to similarity in a cluster while small items contribute to dissimilarity. The cost  $C$  to be minimized consists of two components: the intra-cluster cost and the inter-cluster cost. The intra-cluster cost is measured by the total number of small items and the inter-cluster cost measures the duplication of large items in different clusters. This clustering algorithm aims to minimize the cost due to large items and small items (Wang et al., 1999). The overview of the clustering algorithm is described in Fig. 2.



**Fig. 2.** Overview of clustering algorithm

Further, the opportunity lost can be calculated as the sum of the possibilities that related items lose their sales, when one of them is out of stock. Define  $G_{r,i}$  is the possibility that when item  $r$  is out of stock and it influences on other item  $i$ ,  $G_{r,i}$  can be written as follows:

$$G_{r,i} = \sum_{i \in I_{r,i}} \text{confidence}(r \rightarrow I_{r,i}), \quad (3)$$

where,  $r = 1, 2, \dots, n$  are the items in a frequent item-set.  $n$  is the number of items in a frequent item-set.  $I(r, i)$  is the subset of item  $i$  except  $r$  item in a frequent item-set. In the case of  $i = r$  in the formula,  $I(i, i) = i$  and  $\text{confidence}(i \rightarrow i) = 1$ . We can define the opportunity cost ( $OC_r$ ) by the formula as:

$$OC_r = \sum_{i=1}^n c_i \cdot G_{r,i} \quad (4)$$

where,  $c =$  cost of unit item  $i$ .

Index,  $Z_r$  is used in both deterministic and probabilistic classical inventory policy which is given below (Kaku, 2004):

$$Z_r = \frac{OC_r + H_r}{OC_r}, \quad (5)$$

where,  $H_r =$  holding cost of item  $r$  per unit.

Further,  $Z_r$  is used to implement the concept of opportunity cost. Now, we have considered the model for EOQ described by Maddah and Jaber (2008). In this model, items are delivered with lot size  $y$ , with a purchasing price of  $w$  per unit, and an ordering cost of  $K$ . Here, the assumption is that each lot contains percentage defectives,  $\alpha$  with a known probability density function,  $f(\alpha)$ . The selling price of good-quality item is  $p$  per unit. A 100% percent screening process of the lot is conducted at a rate of  $\lambda$  units per unit time. Further, defective items are sold after inspection at discounted price,  $c_s$  per unit. Now, after inspection,  $\alpha y$  represents the number of defective items. Then, numbers of perfect items are  $(1-\alpha)y$ . To avoid shortages, it is assumed that number of perfect quality items is at least equal to or greater than the demand during screening time  $t$ , that is

$$(1 - \alpha)y \geq Dt, \quad (6)$$

where,  $D$  is the demand per year. Now, replacing screening time  $t$  by  $y/\lambda$  in equation (6), the value of  $\alpha$  is restricted to

$$\alpha \leq 1 - \frac{D}{\lambda}. \quad (7)$$

Revenue (Rev) earned by selling all the items is sum of selling perfect quality items and imperfect quality items.

$$\text{Rev} = p(1 - \alpha)y + c_s \alpha y \quad (8)$$

Total cost (TC) is the sum of cost of all items, ordering cost, screening cost, and holding cost of all items.

$TC =$  cost of items + ordering cost + screening cost + holding cost.

Total cost per cycle of length  $T$  is given as:

$$TC(y) = wy + K + \beta y + h\left(\frac{y(1-\alpha)T}{2} + \frac{\alpha y^2}{\lambda}\right), \quad (9)$$

where,  $\beta$  is the screening cost per unit and  $h$  is holding cost per unit per unit time. The total profit  $P(y)$  is given as total revenue (Rev) – total cost (TC),

$$P(y) = p(1 - \alpha)y + c_s \alpha y - [wy + K + \beta y + h \left( \frac{y(1 - \alpha)T}{2} + \frac{\alpha y^2}{\lambda} \right)] \quad (10)$$

The expected profit per unit time is given by dividing profit per cycle by cycle length  $T$ , i.e.

$$E[P(y)] = E[P(y)]/E[T], \text{ where } T = \frac{(1-E[\alpha])y}{D}$$

$$E[P(y)] = py(1 - E[\alpha]) + c_s y E[\alpha] - [wy + K + \beta y - h \left( \frac{y^2 E[(1 - \alpha)^2]}{2D} + \frac{E[\alpha] y^2}{\lambda} \right)] \quad (11)$$

Since  $\alpha$  is a random variable with known probability density function,  $f(\alpha)$ , then expected profit per unit time,  $E[PU(y)]$  is given as:

$$E[PU(y)] = [p(1 - E[\alpha]) + c_s E[\alpha] - w - \beta]D - \frac{KD}{y} - \frac{hy(E[(1 - \alpha)^2]/2 + E[\alpha]D/\lambda)}{1 - E[\alpha]} \quad (12)$$

Now, to maximize the total expected profit with respect to  $y$ , we will determine first derivative and second derivative of Eq. (12) which are given below:

$$EPU'(y) = \frac{KD}{y^2} - \frac{hE[(1 - \alpha)^2]}{2(1 - E[\alpha])} - \frac{hE[\alpha]D/\lambda}{1 - E[\alpha]}$$

and

$$EPU''(y) = \frac{-2KD}{y^3} \leq 0$$

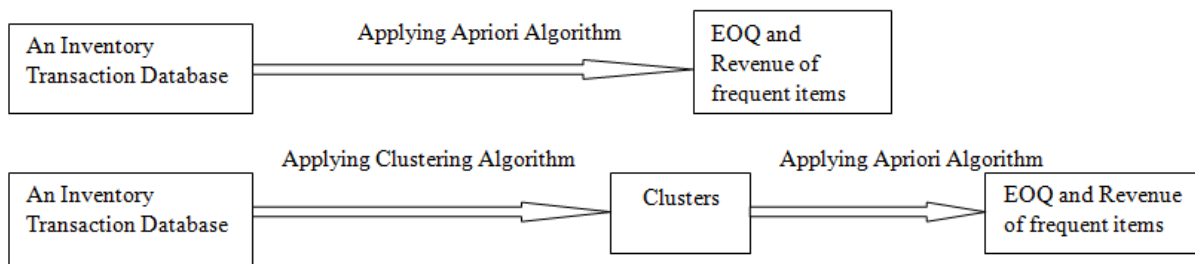
Second order derivative is negative for all values, which implies that there exists a unique value of  $y$  i.e.  $y^*$  that maximizes the profit and it is given as:

$$EOQ(y) = y^* = \sqrt{\frac{2KD}{h[E[(1-\alpha)^2] + 2E[\alpha]D/\lambda]}} \quad (13)$$

The value of  $y^*$  gives the optimal order quantity for item-set. Further, the order quantity is modified to get the optimum order quantity considering cross-selling effects by Mittal et al. (2014). Eq. (13) multiplied with square root of Eq. (5) to get modified order quantity for imperfect quality frequent item-set with cross-selling effects, which is given as:

$$EOQ = y^* \cdot \sqrt{Z_r} \quad (14)$$

Finally, the proposed work can be summarized in Fig. 3.



**Fig. 3.** Proposed research work

Fig. 3 explains the ordering policy for two distinct cases:

- Apriori algorithm is applied on inventory transaction database to calculate opportunity cost of frequent items. Further, opportunity cost is used to estimate the EOQ and revenue of frequent items.
- Clustering algorithm is applied on inventory transaction database to obtain homogeneous clusters. Further, apriori algorithm is applied on each cluster to calculate opportunity cost of frequent items. This opportunity cost is used to estimate the EOQ and revenue of frequent items in each cluster.

#### 4. Numerical Example

A numerical example is solved to calculate EOQ by using rules obtained by apriori algorithm without clustering the data. First, opportunity cost is calculated and then it is used to determine EOQ for imperfect quality items.

The parameters required are given below:

Minimum support, $\text{min\_sup} = 60\%$ ,	Ordering cost, $K = 100/\text{cycle}$ ,
Minimum confidence, $\text{min\_conf} = 75\%$ ,	Screening rate, $\lambda = 2 \text{ unit/min}$ ,
Demand rate, $D = 50,000 \text{ units/year}$ ,	Screening cost, $\beta = \$1/\text{unit}$ .

Consider the database set  $D$  and the inventory item-set,  $I = \{a, b, c, d, e, f, g, h, i\}$ . The inventory transaction set,  $\text{TID} = \{\text{TID1}, \text{TID2}, \text{TID3}, \text{TID4}, \text{TID5}, \text{TID6}\}$  as shown in Table 1. Each row in Table 1 can be taken as an inventory transaction. The association rule can be identified from these inventory transactions using the apriori algorithm. Table 2, shows the inventory policy that will be used to calculate the opportunity cost of various items.

**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

**Table 2**

Inventory policy for various items

Item	Demand	Unit cost	Holding cost	Selling price of perfect quality items	Selling price of imperfect quality items
A	50,000	\$30.00	\$3.00	\$60.00	\$25.00
B	40,000	\$20.50	\$2.00	\$40.00	\$17.00
C	40,000	\$45.52	\$4.00	\$60.00	\$40.00
D	50,000	\$50.00	\$5.00	\$90.00	\$45.00
E	50,000	\$45.52	\$4.00	\$70.00	\$40.00
F	40,000	\$40.00	\$4.00	\$65.00	\$35.00
G	50,000	\$34.00	\$3.00	\$60.00	\$30.00
H	40,000	\$32.00	\$3.00	\$50.00	\$27.00
I	50,000	\$23.00	\$2.00	\$40.00	\$18.00

Further, apriori algorithm is applied on database of Table 1. The frequent item-set found by applying apriori algorithm is  $\{a, b\}$ . Hence, it should be treated as a special item-set in the ranked list of items.

Now, we calculate confidence and opportunity cost by using equation (2) and (4) as:

Confidence (a → b) = 100% and Confidence (b → a) = 100%.

Opportunity cost of item a ( $O_a$ ) =  $C_a \cdot \text{confidence}(a \rightarrow a) + C_b \cdot \text{confidence}(a \rightarrow b) = 30 \times 1 + 20.50 \times 1 \approx 50.50$

After substituting the values of opportunity cost of item 'a' in equation (5), we get Z for item 'a' as:

$$Z_a = \frac{10 + 50.50}{50.50} \approx 1.1980$$

Similarly, after applying rules and conditions described above, the opportunity cost of item 'b' is 50.50 and  $Z_b \approx 1.1980$

Consider the following parameters for item 'a':

Purchase cost,  $w = \$30/\text{unit}$ ,

Selling price of good quality items,  $p = \$60/\text{unit}$ ,

Selling price of imperfect quality items,  $c_s = \$25/\text{unit}$ .

It is assumed that the inventory operation operates on a 6 hours/day, for 365 days a year, then the annual screening rate,  $\lambda = 1 \times 60 \times 6 \times 365 = 1,31,400$  units/year.

Further, it is also assumed that the percentage effective random variable,  $\alpha$ , is uniformly distributed with its probability density function as:

$$f(\alpha) = \begin{cases} 25, & 0 \leq \alpha \leq 0.04, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Note: To avoid shortages condition, Eq. (7) must be satisfied.

$$\alpha \leq 1 - D/\lambda = 1 - 50,000 \times \frac{1}{131,400} \Rightarrow \alpha \leq 0.619482$$

From Eq. (15) we have  $0 \leq \alpha \leq 0.04$  with

$$E[\alpha] = \int_a^b \alpha f(\alpha) d\alpha = \int_0^{0.04} 25 \alpha d\alpha = 0.02$$

and

$$E[(1 - \alpha)^2] = \frac{1}{b-a} \int_a^b (1 - \alpha)^2 d\alpha = \frac{a^2 + ab + b^2}{3} + 1 - a - b = 0.960533$$

Then, the optimum value of item 'a' from Eq. (13) is:

$$EOQ(a) = a^* = \sqrt{\frac{2 \times 100 \times 50,000}{3 \times (0.960533 + 2 \times 0.02 \times 50,000 / 131,400)}} = 1848 \text{ units.}$$

Modified EOQ of item 'a' by Eq. (12),

$$EOQ(a) = a^* \cdot \sqrt{Z_a} = 1848 \times 1.0945 \approx 2023 \text{ units.}$$



Similarly, after applying rules and conditions described above, we can determine the modified EOQ of item ‘b’,

$$EOQ(b) = b^* \cdot \sqrt{Z_b} = 2028 \times 1.0945 \approx 2220 \text{ units.}$$

Table 3, shows the opportunity cost, EOQ and Revenue earned of various items.

**Table 3**

Values for opportunity cost, EOQ, and revenue for items a and b

Items	Opportunity cost	EOQ	Revenue
Item a	50.50	2023 units	\$1,19,964
Item b	50.50	2220 units	\$87,779

Now, we calculate EOQ by using rules obtained by apriori algorithm after clustering the data. First, the transaction database is clustered and then, opportunity cost is calculated, which is used to determine EOQ for imperfect quality items.

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.,  $6 \cdot 60\%$ ). Further, consider the clustering  $C_1 = \{C_1 = \{tid1, tid2, tid3, tid4, tid5, tid6\}\}$ . We have  $Large_1 = \{a, b\}$ ,  $Small_1 = \{c, d, e, f, g, h, i\}$ .  $Intra(C_1) = 7$ , and  $Inter(C_1) = 0$ . So  $Cost(C_1) = 7$ . Again, consider the clustering  $C_2 = \{C_1 = \{tid1, tid2, tid3, tid4\}, C_2 = \{tid5, tid6\}\}$ . 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,  $Intra(C_2) = 5$ ,  $Inter(C_2) = 0$ , and  $Cost(C_2) = 5$ . Thus  $C_2$  has less cost as compared to  $C_1$ . Further, consider the clustering  $C_3 = \{C_1 = \{tid1, tid2\}, C_2 = \{tid3, tid4\}, C_3 = \{tid5, tid6\}\}$ . We have  $Large_1 = \{a, b, c\}$ ,  $Small_1 = \{d\}$ ,  $Large_2 = \{a, b\}$ ,  $Small_2 = \{c, e, f\}$ ,  $Large_3 = \{d, g\}$ ,  $Small_3 = \{h, i\}$ .  $Intra(C_3) = 6$ , and  $Inter(C_3) = 2$ . Hence  $Cost(C_3) = 8$ , which is larger than  $C_2$ . Thus, we will consider cluster  $C_2$ , as it has minimum cost as compared to cluster  $C_1$  and  $C_3$ . 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  $C_2$ . Now, we calculate confidence of frequent item-set  $\{a, b, c\}$  of cluster  $C_1$  and  $\{d, g\}$  of cluster  $C_2$  by using equation (2), as shown in Table 4.

**Table 4**

Confidence of frequent item-set in Cluster  $C_1$  and Cluster  $C_2$

For Cluster $C_1$		For Cluster $C_2$	
Items	Confidence	Items	Confidence
(a → b)	100%	(d → g)	100%
(a → c)	75%	(g → d)	100%
(a → b ∪ c)	75%		
(b → c)	75%		
(b → a)	100%		
(b → c ∪ a)	75%		
(c → a)	100%		
(c → b)	100%		
(c → a ∪ b)	100%		

Opportunity cost of item a ( $O_a$ ) =  $C_a \cdot \text{confidence}(a \rightarrow a) + C_b \cdot (\text{confidence}(a \rightarrow b) + \text{confidence}(a \rightarrow b \cup c)) + C_c \cdot (\text{confidence}(a \rightarrow c) + \text{confidence}(a \rightarrow b \cup c)) = 30 \times 1 + 20.50 \times (1 + .75) + 45.52 \times (.75 + .75) \approx 134.55$ . After substituting the values of opportunity cost of item ‘a’ in equation (5), we get Z for item ‘a’ as:

$$Z_a = \frac{10 + 134.55}{134.55} \approx 1.0743$$

Similarly, after applying rules and conditions described above, the opportunity cost of item ‘b’ is 141.28 and  $Z_b \approx 1.0708$ . The optimum value of item ‘a’ is:

$$EOQ(a) = a^* = \sqrt{\frac{2 \times 100 \times 50\,000}{3 \times (.960533 + 2 \times .02 \times 50000 / 131400)}} = 1848 \text{ units.}$$

Modified EOQ of item ‘a’ by Eq. (12),

$$EOQ(a) = a^* \cdot \sqrt{Z_a} = 1848 \times 1.0364 \approx 1915 \text{ units.}$$

Similarly, after applying rules and conditions described above, we can determine the modified EOQ of item ‘b’,

$$EOQ(b) = b^* \cdot \sqrt{Z_b} = 2028 \times 1.0348 \approx 2099 \text{ units.}$$

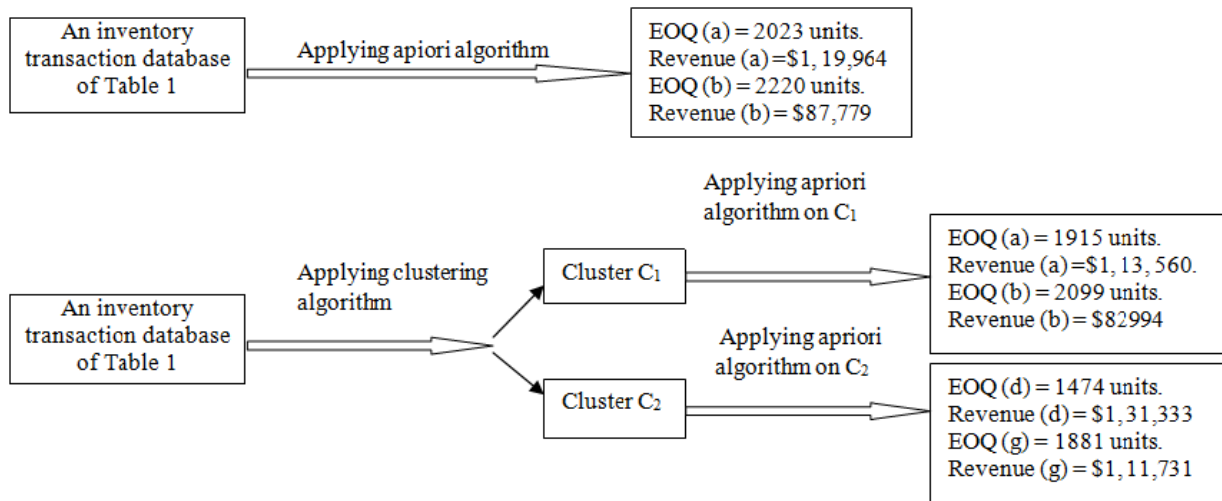
Table 5, shows the opportunity cost, EOQ and Revenue earned of various items after clustering the transactions.

**Table 5**

Values for opportunity cost, EOQ, and revenue for frequent items of cluster C<sub>1</sub> and C<sub>2</sub>

Items	Opportunity cost	EOQ	Revenue
Item a	135.55	1915 units	\$1,13,560
Item b	141.28	2099 units	\$82,994
Item d	84.00	1474 units	\$1,31,333
Item g	84.00	1881 units	\$1,11,731

Finally, the results are summarized in Fig. 4.



**Fig. 4.** Results summary

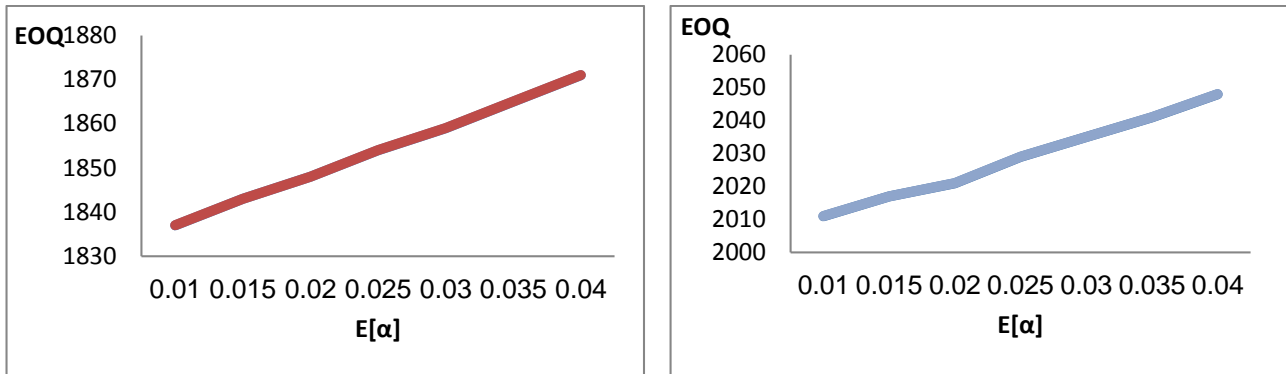
Fig. 4, summarizes the results with two distinct cases:

- Apriori algorithm is applied on inventory transaction database to calculate opportunity cost of frequent items. Further, opportunity cost is used to estimate EOQ and revenue of frequent items. Since the selected number of rules is less, therefore it results in lesser profit for the retailer.
- Clustering algorithm is applied on inventory transaction database to obtain homogeneous clusters. Further, apriori algorithm is applied on each cluster to calculate opportunity cost of frequent items. This opportunity cost is used to estimate EOQ and revenue earned from frequent items. Since the selected number of rules is more in each cluster due to the cross-selling effect among items, therefore it results in higher profit for the retailer.

## 5. Observations

The impact of defective parameter  $E[\alpha]$  on the EOQ have been determined for two cases for item a. Similarly, the graph for other frequent items can be drawn. Results are displayed in Fig. 5.

- Case 1: When items are defectives and without cross-selling effect.
- Case 2: When items are defectives and with cross-selling effect.



**Fig. 5.** Results summary of case 1 and case 2

On the basis of computation result, as shown in Figure 5, we obtain the following phenomenon:

- Fig. 5 explains that as the percentage of defective items decreases, the optimal order quantity increases, which results in higher expected profit for the retailer.
- Further, if items are of defective quality and cross-selling effect is taken into consideration, the optimal order quantity tends to increase more as compared to case 1, which eventually results in higher expected profit.

## 6. Conclusion

This paper determined the EOQ of imperfect quality items for two cases. Case 1 calculated EOQ using apriori algorithm without clustering the transactions while case 2 calculated it by clustering the transactions. In the light of the facts presented in this paper, it has become clear that the number of association rules selected was more when transactions were clustered, which eventually resulted in higher expected profit. Additionally, the impact of defective parameter  $E[\alpha]$  on the EOQ has been determined for both the cases, namely with defectives and without cross-selling effect, and with both defectives and cross-selling effect. This increase in percentage defective items alerts the retailer to look into the source of supply and take corrective measures in order to improve the quality of supply. A numerical example has been solved to validate the result. For future study, it is desirable to extend the proposed model for imperfect quality items by using combination of data mining techniques, such as association rule mining and classification, to obtain improved value of EOQ.

## References

- Agrawal, R. & Srikant, R. (1994). Fast algorithm for mining association rules, *Proceedings of the 20<sup>th</sup> International Conference on Very Large Data Bases*, Chile (pp. 487-499).
- AL-Zawaidah, F.H., Jbara, Y. H. & Marwan, A. A. (2011). An improved algorithm for mining association rules in large databases, *World of Computer Science and Information Technology Journal*, 1(7), 311-316.
- Anand, S.S., Hughes, J.G., Bell, D.A. & Patrick, A.R. (1997). Tackling the cross-sales problem using data mining, *Proceedings of the 2<sup>nd</sup> Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Hongkong (pp. 331-343).

- Broder, A.Z., Glassman, S.C., Manasse, M.S., & Zweig, G. (1997). Syntactic clustering of the web, *Journal of Computer Networks and ISDN Systems*, 29(8), 1157-1166.
- Guha, S., Rastogi, R. & Shim, K. (2000). A robust clustering algorithm for categorical attributes. *Information System*, 25(5), 345-366.
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers. Inc., USA.
- Jaggi, C.K., Mittal, M. & Khanna, A. (2012). Effects of inspection on retailer's ordering policy for deteriorating items with time-dependent demand under inflationary conditions. *International Journal of Systems and Science*, 44(9), 1774-1782.
- Jaggi, C.K. & Mittal, M. (2012). Retailer ordering policy for deteriorating items with initial inspection and allowable shortage under the condition of permissible delay in payments. *International Journal of Applied Industrial Engineering*, 1(1), 46-79.
- Jaggi, C.K. & Mittal, M. (2011). Economic order quantity model for deteriorating items with imperfect quality. *International Journal Revista Invetigacion Operacional*, 32(2), 107-113.
- Jaggi, C.K., Goel, S.K. & Mittal, M. (2013). Credit financing in economic ordering policies for defective items with allowable shortages. *International Journal of Applied Mathematics and Computation*, 219(10), 5268–5282.
- Jaggi, C.K., Goel, S.K. & Mittal, M. (2011). Economic order quantity model for deteriorating items with imperfect quality and permissible delay on payments. *International Journal of Industrial Engineering Computations*, 2(2), 237-248.
- Kaku, I. (2004). A data mining framework for classification of inventories. *Proceedings of the 5<sup>th</sup> International Conference on Management*, Australia (pp. 450-455).
- Lee, H.L. & Rosenblatt, M.J. (1987). Simultaneous determination of production cycles and inspection schedules in a production system. *Management Science*, 33(9), 1125-1137.
- Lindell, Y. & Pinkas, B. (2002). Privacy preserving data mining. *Journal of Cryptography*, 15(3), 177-206.
- Maddah, B. & Jaber, M. Y. (2008). Economic production quantity model for items with imperfect quality: Revisited. *International Journal of Production Economics*, 112(2), 808-815.
- Maddah, B., Salameh, M.K. & Moussawi- Haider, L. (2010). Order overlapping: A practical approach for preventing shortages during screening. *Computer and Industrial Engineering*, 58(4), 691-695.
- Mittal, M., Pareek, S., & Agarwal, R. (2014). Efficient ordering policy for imperfect quality items using association rule mining. *Encyclopedia of Information Science and Technology* 3<sup>rd</sup> ed., United States, Information Science Publishing, pp. 773-786.
- Porteus, E.L. (1986). Optimal lot sizing, process quality improvement and setup cost reduction. *Operations Research*, 34(1), 137-144.
- Papachristos, S. & Kontantaras, I. (2006). Economic ordering quantity models for items with imperfect quality. *International Journal of Production Economics*, 100(1), 148–154.
- Rosenblatt, M.J. & Lee, H.L. (1986). Economic production cycles with imperfect production processes. *IIE Transactions*, 18(1), 48-55.
- Salameh, M.K. & Jaber, M.Y. (2000). Economic production quantity model for item with imperfect quality. *International Journal of Production Economics*, 64(1), 59-64.
- 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).
- Wee H.M., Yu J. & Chen M.C. (2007). Optimal inventory model for items with imperfect quantity and shortage backordering. *OMEGA*, 35(1), 7-11.
- Zhao, Q. & Bhowmick, S.S. (2003). Association rule mining: A survey, (Report No. 2003118), *Center for advanced information systems, Nanyang Technological University, Singapore*.