International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:2, No:4, 2008

Application of Association Rule Mining in Supplier Selection Criteria

A. Haery, N. Salmasi, M. Modarres Yazdi, and H. Iranmanesh

***Abstract*—**In this paper the application of rule mining in order to review the effective factors on supplier selection is reviewed in the following three sections 1) criteria selecting and information gathering 2) performing association rule mining 3) validation and constituting rule base. Afterwards a few of applications of rule base is explained. Then, a numerical example is presented and analyzed by Clementine software. Some of extracted rules as well as the results are presented at the end.

***Keywords*—**Association rule mining, data mining, supplier selection criteria.

I. Introduction

SUPPLIER selection is an important subject in supply chain management (SCM). Low accuracy in that leads to financial loses for buyers.

Supplier selecting methods consist of two major categories: quantitative and qualitative. Many multi criteria techniques have been used for this purpose. Dickson [1] reviewed 23 criteria and claims that quality, lead time, and efficiency are the most important criteria. Weber et al. [2] perform another study and conclude that the price, lead time, and quality are the most important criteria. Weber et al. [3] did another study about this subject by emphasizing on geographic position criterion. There are also many articles about supplier selection criteria in numerous industries and countries. After 1995 by fast growth of e-commerce, some of the criteria have been changed. De Boer et al. [4] run a complete survey on decision making method in supplier selection process. Their major results are:

- DEA (Data Employment Analysis), Cluster Analysis and Case-Base Reasoning have been used in initial steps of selection and qualification audit.
- Decision models for ultimate selection include linear weighting, total cost ownership, mathematical programming, statistical techniques and artificial intelligence based models.

Morlacchi [5] by combining fuzzy sets and AHP develops a model to evaluate small suppliers in engineering field. Ghodsypour et al. [6], present a model on the basis of AHP and liner programming integration. The objectives of the model are selecting the best suppliers and determining order size that maximizes the revenue.

Weber et al. [7] by combining MOP (Multi-Objective Programming) and DEA, develop a tool for negotiation among buyers and suppliers. Wang et al. [8] by using AHP and goal programming integration suggest a technique that involve both quality and quantity criteria. Ramayya Krishnan and Sung Ho Ha [9] present a combined method on the basis of AHP, DEA, and neural network that uses of a Combined Supplier Score (CSS). This model is based on both qualitative and quantitative criteria. The model enables buyers to comparison among single and multi resource options. The model attains supplier map (SM) by cluster analysis. After clustering, suppliers are divided into different segments. Every segment is different from other segments in basic criteria.

In most of the past research the variables and the criteria were considered independently which is not practical. For instance, there exist the following dependencies among several selection criteria:

- Lead time and information system
- Manufacturing processes efficiency and price
- Quality and human resource skills

In this research we introduce the application of association of rule mining (one of data mining techniques) in supplier selection. Data mining is a process that finds valuable information from a huge amount of data in order to be used in decision making for achieving to business goals.

Valuable information is hidden patterns and rules that are not obvious. Especially, efficiency of data mining techniques is apparent when the data set is huge.

Certainly, there is rich information about suppliers in any company. The purpose is to find the best suppliers that maximize the buyer’s profit. So far data mining can enables the managers to making better decisions.

Association rule mining (ARM) is a useful data mining technique to detect patterns and rules. Apriori is an applied ARM algorithm to find rules and relations among variables in a data set.

A. Haery is with Socio-Economic Systems Engineering, Sharif University of Technology, Tehran, Iran (e-mail: a_haery@ie.sharif.edu).
N. Salmasi, Assistant Professor, Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran (nsalmasi@sharif.edu).
M. Modarres Yazdi, Professor, Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran (e-mail: m.modarres@sharif.edu).
H. Iranmanesh, Assistant Professor, "University of Tehran" & "Institute for Trade Studies & Research" ,Tehran, Iran (corresponding author , phone: +9821-88021067, Fax: +9821- 88013102 , e-mail: hiranmanesh@ut.ac.ir).

II. PROPOSED APPROACH

The suggested approach for performing ARM includes three phases:

phase1: criteria selection and information gathering

In this phase, a few criteria (depend on the industry and the country) is selected. Then, the available information about the chosen criteria is gathered.

Phase 2: performing ARM

In this phase, the collected information is analyzed and a few relevant rulecriteria are detected. One of the applied algorithm is Apriori. Apriori finds rules and relations among variables in a data set. The algorithm finds frequent patterns and calculates the rule's several indices. Support and confidence, are the two major indices, which have useful applications to evaluate the rules. For instance, consider rule X: if A then B. Suppose that this rule has 80% confidence and 30% support. It expresses that 30% of records contain A and B. This means that in 30% of total records, rule X is valid. Additionally, it expresses that 80% of records that contain A, contain B as well. We explain this algorithm by a simple example. This example is adopted from [10].

Consider the data illustrated in Table I, depicting the items purchased by customers in four transactions.

TABLE I
SAMPLE TRANSACTIONS TO DEMONSTRATE ASSOCIATION RULE MINING

Transaction ID	Purchased items
1	{1,2,3}
2	{1,3}
3	{1,4}
4	{2,5,6}

For a minimum support of 50% (here, two transactions) and a minimum confidence of 50%, we have the following rules:

(i) 1 => 3 with 50% support and 66% confidence;
(ii) 3 => 1 with 50% support and 100% confidence.

TABLE II
COMPUTATION OF FREQUENT ITEMSETS

Frequent itemset	Support (%)
{1}	75
{2}	50
{3}	50
{1,3}	50

The objective is to generate confident rules, having at least the minimum confidence. The problem decomposition proceeds as follows:

- Find all sets of items that have minimum support, typically using the Apriori algorithm. This is the most time consuming phase of the search, and involves lots of research for reducing the complexity.
- Use the frequent itemsets to generate the desired rules. Given m items there can be potentially 2^m frequent itemsets.

Consider Table 2. For the rule 1 => 3, we have
Support = Support{l,3} = 50% and
Confidence = Support{l,3}/ Support{l} = 66%.

The Apriori algorithm is outlined as follows. Let F_k be the set of frequent itemsets of size k, let C_k be the set of candidate itemsets of size k, and let F_1 be the set of large items. We start from $k = 1$.

1. for all items in frequent itemset F_k repeat steps 2-4.
2. Generate new candidates C_{k+1} from F_k.
3. for each transaction, let's say transaction T, in the database, increment the count of all candidates in C_{k+1} that are contained in T.
4. Generate the frequent itemsets F_{k+1} of size k from candidates in C_{k+1} with minimum support.

The final solution is F_k.. The data set for the example is illustrated in Table III and the procedure is shown in Table IV, respectively.

TABLE III
EXAMPLE TRANSACTIONS DATABASE FOR FREQUENT ITEMSET GENERATION

Transaction ID	Purchased items
1	{1,3,4}
2	{2,3,5}
3	{1,2,3,5}
4	{2,5}

TABLE IV
STAGES OF APRIORI ALGORITHM DEMONSTRATING FREQUENT ITEMSET GENERATION

C_1	Count	Suppor	F_1		C_2	Count	Suppor	F_2
{1}	2	50	_		{2,3}	2	50	_
{2}	3	75	{2}		{2,5}	3	75	{2,5}
{3}	3	75	{3}		{3,5}	4	50	_
{4}	1	25	_		_	_	_	_
{5}	3	75	{5}		_	_	_	_

A key observation is that every subset of a frequent itemset is also frequent. This implies that a candidate itemset in C_{k+1} can be pruned if even one of its subsets is not contained in F_k..

This process is continued for every subset. There are also other methods that can help to Apriori to perform investigation and increase the speed of the calculation.

Phase 3: validation and constituting rule base

In this phase we should validate the discovered rules. The best procedure for validation is dividing the data set in two segments. The first segment is assigned for training and the remaining part is set for validation and testing Then, Apriori algorithm is applied for both sets. Afterwards, we proceed to analysis and do comparison among outputs of each section. The output of training phase, that is a set of rules, is used for validation and estimating the errors of the detected rules. Some of the rules should be selected. Key point for selecting rules is that the selected rules should have approximately closed support and confidence ratio. Then, the rules are filtered based on the management policy. For instance, suppose that the management's opinion is that the applied rules must have at least 30% confidence and 5% support. Ultimately the remained rules stored in a rule base for future applications.

III. The Advantages of Offered Approach

A. Help to Make Better Decisions when the Data Set is Incomplete

In practice it is possible that the information about some variables (variables are a selection of the supplier criteria) is not complete. For instance, suppose that there is deficiency about criterion '*C*'. In this case, estimating the value of criterion 'c' for a specific supplier is possible by using of rules that criterion 'C' is consequent in those. The estimate is as good as the confidence measure of those rules.

B. Increasing the Ability for Changing by Making Suitable Policies

Often supplier switching has a high expense for a company. Maybe there is not a better supplier to present supplier. Rule base, is the other term for rule set which enables the management to use of them to reach to the buyer business goals. Business goals maybe increase the quality of products or decrease the lead time. This subject depends on the authority of the buyer proportion to his suppliers.

Suppose that a rule is "if E >=3 then C <=2". E is the efficiency of supplier information system and C is the price of product. This rule reveals that if efficiency of supplier information system improves (greater or equal than a specific limit), the cost of activities performing reduces and the price will not increase more than a specific limit. Suppose that the value of E criterion about a supplier is equal to 2 and confidence of this rule is more than 70%. Thus, if any change is performed and the value of E criterion is increased to 3 or more, the price will be reduced (by 70% probability). For understanding the importance of that, suppose conditions that we should make decision without rule base. Besides we know that the cost depends to several causes such as efficiency of human resources, machines, manufacturing processes, and the cost of materials. Regarding to the budget constraint it is not also feasible to invest in improvement of all criteria. But on the basis of rule base, the managers can specify the priority of changes. Generally it is possible to develop a mathematical model for reaching to a specific objective. Objective function can be the cost for reaching to goal, probability of reaching to goal or both of them. If the objective function includes aforesaid cost, aforesaid probability considered as a constraint. This means that probability has an accepted minimum. If objective function includes probability, cost considered as a constraint. This means that cost has an accepted maximum.

C. Increase the Ability for Supplier Selecting

Consider a rule such: "if A=1 then B=1". Assume that the rule A is the speed of supplier's responsibility to buyer requirements (before contract ratification). B is the indicator of product on-time delivery. Suppose that this rule has a high confidence (like 80% or more) and B=1 is desirable value for the buyer. Hence, this rule helps to the buyer for selecting supplier. For instance, the buyer can initially send requests to different suppliers. After observing the feedbacks, the buyer can accomplish the survey on the suppliers that their score for A criterion is equal to1.

D. Information Gathering about Related Criteria with Less Time and Cost

We explain aforesaid title with an example. Suppose that the C criterion is always equal to 1 or 2 and two below rules have 80% confidence:

If A <=2 then C=1

If B >3 then C=2

Suppose that 80% is a sufficient accuracy for estimating C criterion. Hence, it is possible to estimate the C criterion instead of gathered information about the C criterion. In other words, C can be estimated on the basis of A and B criterion. Generally perception of rules among supplier evaluation criteria can help the suppliers to select suitable criteria.

In this part a mathematical model can be developed. The major input parameters are the importance of accurate estimation criteria, cost of information gathering about each criterion, and confidence of detected rules. The importance of parameters accurate estimation represents proportionate to cost of wrong estimation. The objective function is gathering information cost, probability of correct estimation (without information gathering) or both of them. If the objective function includes only the cost, the probability of correct estimation appears in constraints. In this status constraint expresses that probability of correct estimation about goal criteria is greater than a specific limit. Besides, if the objective function includes only probability, cost appears in constraints. In this case, the constraint expresses that the cost for reaching to goal is not greater than a specific limit.

E. Rule Base and Gathered Information Validation

Even if there is tendency to gathering complete information, the rule base can be useful for gathered information validation. Besides, the rule base is revised and updated on the basis of new information. For example, suppose that there is a rule Z "if X=2 then Y>=5". Also suppose that this rule has 60% confidence and 40% support. Now suppose that in gathered information, support of rule Z has a considerable difference from 60% or confidence of rule X has a considerable difference from 40%. In this case, there are 2 feasible conditions:

The difference derived from error in gathered information. Error depends on to small sample size, unreliability about information source and etc

The difference derived from rules changing among criteria. In this condition there is a need to gathering more accurate information. If changing is confirmed, the rule base revising is necessary.

IV. Numeric Example

In this section an ARM analysis is performed by using the data from [9]. Selection supplier criteria are QSO (Quality System Outcome), Claims (CL), Quality Improvement (QI), Response to Claims (RC), On-time Delivery (OD), Internal

Audit (IA), and Data Administration (DA). Data set is mentioned in appendix1. In each step of ARM, we should consider one field as consequent variable and other fields as antecedent variables. ARM is performed by using of Climentine 8.1. Some of detected rules are described in appendix2.

V. Offers for Future Researches

- Use of suggested approach for suppliers in different industries (by enough data) to detect valuable and applied rules
- developing the mathematical models that suggested in section B and section D

VI. Conclusion

After solving the numeric example, it is clear that the ARM can detect the rules with high support and confidence. The rule base detects valuable patterns about supplier behavior and prevents from making wrong decision repeatedly. The main advantage of the suggested approach is the possibility of revising and updating the rule base permanently.

Appendix I

Quantities of 7 Criteria about 27 Suppliers

Index	QSO	CL	QI	RC	OD	IA	DA
1	4.5	5	8	1	1	2	22.6
2	5	1	10	2	2	2.5	22.5
3	5	5	10	3	3	1.5	27.5
4	4.5	5	8	1	1	3.5	22
5	5	5	9	1	1	5	28.5
6	4	1	7	2	2	2.5	26
7	5	1	10	2	2	4	24.5
8	5	1	10	2	2	4	24.5
9	4	5	7	3	3	4.5	26.5
10	4.3	5	7.7	3	3	4	25.4
11	5	5	10	3	3	3	28
12	5	5	10	3	3	3.5	27.5
13	5	5	10	3	3	4	28
14	5	5	10	3	3	4	28
15	5	5	10	3	3	4	30
16	5	5	10	3	3	4.5	29
17	4	5	8	2	2	3	22.5
18	4	1	8	2	2	4	24.5
19	4	1	8	2	2	2.5	27.9
20	4.5	5	9.5	3	3	3	27.5
21	3.5	5	7.5	2.5	2.5	3.5	24
22	4	5	8.5	1	1	3.5	26.3
23	4	5	9	3	3	4	29
24	4	5	10	3	3	4.5	27.5
25	4	5	10	3	3	5	29.5
26	3	5	10	3	3	3	22.5
27	2.5	5	8	1	1	3	23.7

Appendix2

Detected Rules and Their Measures

Ix	In	SP	CF	C	A1	A2	A3	A4
1	6	22.22	100	QSO>=4	CL < 3.000	-	-	-
2	7	25.93	57	QSO>=4	DA < 24.250	-	-	-
3	2	7.41	50	QSO>=4	QI < 7.600	CL > 3.000	-	-
4	5	18.52	40	QSO>=4	DA < 24.250	IA > 2.750	-	-
5	3	11.11	33	QSO>=4	DA < 24.250	OD > 1.500	CL > 3.000	-
6	3	11.11	33	QSO>=4	DA < 24.250	RC > 1.500	CL > 3.000	-
7	5	18.52	100	CL=1	OD < 2.250	OD > 1.500	DA > 24.25	-
8	5	18.52	100	CL=1	OD < 2.250	RC > 1.500	DA > 24.25	-
9	5	18.52	100	CL=1	RC < 2.250	OD > 1.500	DA > 24.25	-
10	5	18.52	100	CL=1	RC < 2.250	RC > 1.500	DA > 24.25	-
11	3	11.11	100	CL=1	IA < 2.750	IA > 2.250	-	-
12	3	11.11	100	CL=1	OD < 2.250	QI > 9.750	-	-
13	3	11.11	100	CL=1	RC < 2.250	QI > 9.750	-	-
14	3	11.11	100	CL=1	QSO < 4.150	QI < 8.250	DA > 24.25	RC < 2.250
15	2	7.41	100	CL=1	OD < 2.250	QSO <	IA < 2.750	
16	2	7.41	100	CL=1	OD < 2.250	QSO <	QI < 8.250	IA < 2.750
17	2	7.41	100	CL=1	RC < 2.250	QSO <	IA < 2.750	
18	2	7.41	100	CL=1	RC < 2.250	QSO <	QI < 8.250	IA < 2.750
19	2	7.41	100	CL=1	QSO < 4.150	IA < 2.750	-	-
20	2	7.41	100	CL=1	QSO < 4.150	OD < 2.250	IA < 2.750	-
21	2	7.41	100	CL=1	QSO < 4.150	OD < 2.250	QI < 8.250	IA < 2.750
22	2	7.41	100	CL=1	QSO < 4.150	RC < 2.250	IA < 2.750	-
23	2	7.41	100	CL=1	QSO < 4.150	RC < 2.250	QI < 8.250	IA < 2.750
24	2	7.41	100	CL=1	QSO < 4.150	QI < 8.250	IA < 2.750	-
25	7	25.93	86	CL=1	OD < 2.250	RC > 1.500	-	-
26	7	25.93	86	CL=1	RC < 2.250	RC > 1.500	-	-
27	15	55.56	100	CL=5	RC > 2.250	-	-	-
28	3	11.11	100	CL=5	QSO < 3.750	-	-	-

Abbreviations:
Ix = index, In = Instances, SP = Support, CF = Confidence,
C = Consequent, A1 = Antecedent 1, A2 = Antecedent 2,
A3 = Antecedent 3, A4 = Antecedent 4

ACKNOWLEDGMENT

The authors would like to appreciate Rohollah Ghasempour and A. Askary for their contribution in this research.

REFERENCES

[1] G.W. Dickson, An analysis of vendor selection systems and decisions, Journal of Purchasing 2 (1) (1966), pp. 5–17.
[2] C.A. Weber, J.R. Current and W.C. Benton, Vendor selection criteria and methods, European Journal of Operational Research 50 (1) (1991), pp. 2–18.
[3] C.A. Weber, J.R. Current and A. Desai, Non-cooperative negotiation strategies for vendor selection, European Journal of Operational Research 108 (1998), pp. 208–223.
[4] L. De Boer, E. Labro and P. Morlacchi, A review of methods supporting supplier selection, European Journal of Purchasing & Supply Management 7 (2) (2001), pp. 75–89.
[5] Morlacchi, P. (1999). Vendor evaluation and selection: the design process and a fuzzy-hierarchical model. In Proceedings of 8th International Annual IPSERA Conference, Belfast, Dublin.
[6] S.H. Ghodsypour and C. O'Brien, A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming, International Journal of Production Economics (1998), pp. 199–212.
[7] C.A. Weber, J.R. Current and A. Desai, An optimization approach to determining the number of vendors to employ, Supply Chain Management: an International Journal 5 (2) (2000), pp. 90–98.
[8] G. Wang, S.H. Huang and J.P. Dismukes, Product-driven supply chain selection using integrated multi-criteria decision-making methodology, International Journal of Production Economics 91 (1) (2004), pp. 1–15.
[9] Sung Ho Ha and Ramayya Krishnan, A hybrid approach to supplier selection for the maintenance of a competitive supply chain, Expert Systems with Applications, Volume 34,Issue 2, February 2008, Pages 1303-1311
[10] Sushmita Mitra, Tinku Acharya, Data Mining Multimedia, Soft Computing and Bioinformatics, John Wiley & Sons Publication, (2003) pp 269-271