J. Basic. Appl. Sci. Res., 2(2)1742-1748, 2012

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

The Effect of Data Mining Based on Association Rules in Strategic Management

MOHSEN RASOULIAN[1] and SAEED AMIRIRAD[2]

[1]Faculty member of Islamic Azad University, Naragh branch, Naragh, IRAN
[2]M.Sc student of management, Islamic Azad University, Naragh branch, Naragh, Iran

ABSTRACT

Strategic planning in the organizations asks for high consideration in required data mining. The challenge with which the strategists confront is the interchangeable nature of the data. Applying data mining is a suitable mean for managers.
In this article, the associated rules for different fields of one organization are going to be recognized. In order to do this, the researchers are supposed to seek the required information for strategic decisions in an organization by using statistical methods. The main aim of this article is helping to recognize much faster and more reliable methods to codify the required strategies for the organizations.
KEY WORDS: data mining; strategic management; association rules; marketing; frequent item.

1. INTRODUCTION

Looking at the remarkable achievements of societies during different periods of history of human civilization indicates the issue that planning has always had an especially unique place as the backbone of development. In the changing world, the organizations and societies should change according to environmental changes not only to achieve a trying superiority, but also, to revive. However, suitable changing in human sets is much more complex due to dynamicity and continuity of changing, and as well due to collectivity of effects and interaction of more complex factors. In the recent decades, the importance ofhuman behavior study, specially the attitudes of buyers has won special importance. One issue that helps managers in providing required products for customers is recognizing the complementary or related product to satisfy the needs of customers. The data mining community has a well-established set of techniques available, which we are seeking to apply to an ever greater variety of data. Generally speaking, the actual data mining processes, in many cases, are readily available. [8]. In data mining, most of the current textbooks show a strong bias towards one of its finding disciplines, like database management [11], machine learning [12] or statistics [13].Whereas economic theory and strategic choice theory were the dominant conceptual perspectives through the 1980s, since then the resource-based view, knowledge-based view, agency theory, and institutional theory all have attracted adherents. [10]

Data mining is one of the latest methods that can help to show the importance of relations between needs in this field. Today, in the context of tabular data, we have a well-established range of data mining techniques available [8].

2. Problem definition

One marketing information system (MIS) should consider the marketing environment and provide the required information to achieve the key decision-making for the managers [6]. Typically a customer can be served only in case it's corresponding ‘quantity currently demanded’ is more than zero [9].The best method to collect the descriptive information is inductive research that presents some information about awareness, opinions, preference, or purchasing attitude of people through questionnaire (direct questioning) [6]. Data Mining requires a set of system appearances to derive information about system structure [9].

The most important advantage of inductive research is high flexibility in collecting information from the main sources. However, it is limited in some ways. Sometimes people cannot answer the questions because they have not thought about that work even they cannot remember the reason of that at all. Moreover, they may are not intended to answer some questions that are regarded as personal ones to strangers. Some audiences present false information without being aware to present themselves or to win the attention of the interviewer [6].

Therefore, in order to prevent occurring these issues, one can present modern methods and data mining algorithms in this field. Through mining on the questionnaires of the similar organizations and mining of laws and hidden knowledge in them, one can achieve a criterion on relation between different parts in strategic planning. Mining is the final goal to discover the hidden information or knowledge from either the original or the transformed time series data [7].

***Corresponding Author:** SAEED AMIRIRAD, M.Sc student of management, Islamic Azad University, Naragh branch, Naragh, Iran.
Email: Amiriradsaeed.edu@gmail.com

Association rules:

Frequent item set plays a vital role in many data analysis. Frequent item set, are methods by which one tries to extract interesting patterns from data bank. Among these methods, one can refer to association rules, correlation, sequence, episodes, etc [1].

Association rulesin fact indicate how to buy multiple products together. For instance, the association rule "soft drink-cake (%80)" indicates that among 100 soft drink buyers 80 persons will buy cake as well. Such rules are very useful including one can refer to strategic decision-making.

Formulating the problem:

In order to find the solution and the answer of each research, it is of great importance to analyze the definitions and the way of formulating the problem, and test hypotheses that are raised as probable responses. Therefore, the applied definitions and symbols in this research and the way of formulating the problem [2] are all described as follows:

Definition1. Suppose I= $\{i_1, i_2,\ldots,i_n\}$ is a set of items with **n** members. Set $X=\{x_1,x_2,\ldots,x_n\}$ is called a **K** item set.

Definition2. A transaction on I includes one T=(tid, I)I pair that **tid** is the number of the transaction and **I** is a set of items on **I**.

If item $x \subseteq I$ in **I**, i.e. $x \subseteq I$ supports **T** transaction from item set X.

Definition3. Bank data **D** and transactions on **I**, is a set of transactions on **I**.

Definition4. The cover of item sets in bank **D** defines as follows:

Cover (X, D) = {tid| (tid,I) ϵ D, X⊆ I}

Definition5. The support of one set of items in **D** means, the number of transactions that covers **X,** in **D.** support(X, D) =|Cover(X,D)|.

Definition6. Frequency includes the probability of **X** happening in a **T** transaction that belongs to **D**.

$$\text{Frequency}(X,D)=p(x)=\frac{Support(X,D)}{|D|}$$

That |D| includes the total number of transactions and is equal to |D|=Support ({}, D)

Definition7. One calls a set, a frequent item if the value of support will not be less that threshold σ_{abs} which means $0\leq \sigma_{abs}\leq|D|$

When we talk about frequency, this value can be as the against; $0\leq \sigma_{abs}\leq 1$ and naturally the relationship $\sigma_{abs}=[\sigma_{rel}\times|D|]$ will exist.

Definition8. Suppose that **D** is a transaction bank on items **I** and σ is a threshold, therefore the set of frequent items will be shown as follows;

F (D, σ)={X ⊆ I| Support (X, D)≥ σ }Sometimes F (D, σ) is shown as **F.**Ex 1. (Item set mining) if σ , **D, I** are given, find F (D, σ)

In practice, F (D, σ) just is not important, but the number of real frequency of existed items is moreimportant in that. One association rule is X $\Rightarrow$ Y when **X** and **Y** are the item set and X $\cap$ Y= ∅. This rule indicates that if transaction includes **X** it should include **Y** as well. The support value ruleX $\Rightarrow$ Y equals the support value of X $\cup$ Y in **D** and the frequency equals frequency of X $\cup$ Y in **D**.

Definition9. Confidence or accuracy of a law X $\Rightarrow$ Yin **D** includes the possibility of **Y** in a transaction if **X** exists in that transaction.

$$\text{Confidence}(X \Rightarrow Y, D) = P(Y|X) = \frac{Support(X \cup Y, D)}{Suppor(X, D)}$$

Definition10. One rule is regarded "confident" if P(Y|X) is greater than a threshold; λ for example that equals $0\leq \lambda \leq 1$.

Definition11. Suppose **D** is a transaction bank on **I** and σ is a confidence threshold, and λ is a confident threshold, the frequent and confident item set are regarded as follows regarding σ and λ :

R (D, σ, λ)={X $\Rightarrow$ Y |X, Y $\subseteq$ I, X $\cap$ Y= $\emptyset$, X $\cup$ Y $\in$ F(D, σ), Confidence (X $\Rightarrow$ Y, D)$\geq$ λ }

Example2. (Association rule mining) suppose that we have a set **I,** transactions set **D** on that we have and σ, and λ , now find R(D, σ, λ).

Along **R,** support and confidence coefficient of the extracted rules are important as well. As you can see the problem of frequency items set mining is a special case of the mining association rulesproblem. Itmeans that each frequent item indicates that one rule however unnecessary is as the opposite; X $\Rightarrow$ { } is 100% based on the confidence coefficient of this rule. It is clear that the number of rule frequency and **X** is the same.

Therefore, it is clear that for each set the frequent item **I**, there are rule as X $\Rightarrow$ Y that X $\cup$ Y=I and all of them have a confidence coefficient equal to σ_{rel}.

Hypotheses:

One of the most important stages in the procedure of strategic management is the stage of converting strategic theory to the strategic practice. Performing this stage requires three related steps:

1. Recognizing and setting annual measurable objectives.
2. Planning and developing functional strategies.
3. Creating and connecting between politics in order to guide the decisions.

Setting annual measurable objectives, mean turning long-term expectations to annual budget. The role of functional strategies is converting main strategies to practical sketches and finally policies are specific guidelines for operation managers and their staffs.

Although developing and scientific planning of annual objectives are powerful tool to operate the strategies, doing this job individually is not enough for strategies success. The managers should define and devise the necessities of objectives realization or functional strategies in order to encourage the successful operation of strategy as clearly as possible. Therefore, the designers of functional strategies should recognize and coordinate the actions in every single unit and increase the probability of annual objectives by supporting the main strategies.

It is observed in practice that in order to extract the actions, large volumes of data are produced. For example, in an organization with 20 sub-units one should distribute some questionnaires at the level of each unit experts (supposedly 4 people in each unit) to extract functional strategies in 5 areas including marketing, financial, research and development, production and operations and field of human resources. If this questionnaire contains 10 questions in average and each question contains 5 choices, we will have:

The number of organizational unit: 20
The number of extracting strategy fields: 5
The number of experts= 4 (one manager and three experts)
The number of questions= 10
The number of choices= 5

The number of all answers= 20*5*3*10*5=20000

Of course, the above-mentioned hypothetical organization is regarded as a small organization. Organizations with 400 units are now regarded as average organizations. Therefore, in large organizations, the volume of information increases severely.

H1: it is possible that in this method the users answer the questionnaire questions wrongly due to lack of knowledge or not understanding the question.

H2: Through exploring the similar answer sheets, extracting rules, and hidden knowledge in them, we achieve suitable criterion in the field of relation between different parts in strategic programming.

H3: Regarding the criterion, we can extract the suspicious answers and we can reach to some conclusion by examining responses and ideas of related subjects. If this problem is because of users' lack of enough time and patience, then we have to give them enough time to complete the answer sheets, or it may resulted from incorrect understanding of questions, in this case one should train them. Yet finally, it may due to other reasons for which one should find out the reasons of this weakness by more examinations.

H4:another vital note in this method, is designing suitable questions comprehensively so that it can lead us to a suitable strategy for the organization. Using the data analysis techniques, one can analyze the designed questions in this field. Having analyzed the resulted data from the users' questionnaires and having extracted the answering pattern to the questions, we reached to this conclusion that some of the questions were not necessary.

Naturally, if the relationship between two questions is 100%, it may have some problems in designing the questions.

3. Problem solving method

The authors of the present paper, attempted to determine the strategy of data technology of an organization in order to plan the strategic program to convert the status quo to the suitable condition through scientific method. In order to do this, one needed to analyze the status quo. On one hand, a questionnaire is one of the common research tools and a direct method for collecting data. A questionnaire presents a set of questions, statements, and phrases by answering to which, respondents compose the data necessary for the researcher. One can consider questions as a type of stimulus- response. Through questionnaire questions, we can analyze the knowledge, interests, attitudes, and ideas of that person, we can also find out his/her previous experience and we can learn what s/he is doing now. Therefore, the executive team attempted to design and prepare the questionnaire.

In this context, there was the concern of preparing a standard questionnaire which is based on scientific principals [3].

Having done a lot examination, we chose Mosaic group questionnaire that has classic origin. Since this questionnaire states that the capacity of data technology in each organization indicates the capability of organizations and the subjects of that organization in line of development, publication and efficient application of inter organizational to determine the technological, social, and economical objectives [4].

This model is useful for beneficiaries who aim at using the data technology or aim at investing in it and it discusses for those policy makers who discuss the manner of positive and negative effects of using and developing data technology. Moreover, this model is useful for researchers who study the macro expansion of complicated technologies.

Mosaic group, categorizes the effective and determinative factors of data technology in four major dimensions:

- Technology
- Resources
- Landscape
- Dynamicity

After examinations and necessary studies, final questions determined as it is shown in table 1:

Table 1, the finalized questionnaire in order to do the tests

Question title	Question number	Question field
Describe the status of the existing hardware.	1	The questions of the government policies in field of technology
Describe the status of the existing software.	2	
Describe the status of existing information systems.	3	
Describe the status of telephone line of communication.	4	
Describe the existing Backboned **communication status among (phone lines, network cables, and fiber channels).**	5	
Describe the status of existing database	6	
Describe the existing status of network security equipment	7	
Describe the existing amount of systems integration.	8	
Describe the knowledge rate of managers about IT.	9	The questions of resources field
Describe the staff information about IT.	10	
Describe the rate of effective use and targeted information.	11	
Describe the rate of knowledge production.	12	
Regarding the responsibilities in IT **field, describe the balance between applied experts and specialist.**	13	
Describe the rate of internet access.	14	
Describe the existing amount of standard codified in order to apply IT**.**	15	
Describe the rate of investment in IT **based on macro objectives.**	16	
Describe the strategic status of the existing IT.	17	
Describe the rate of comprehensive codifications of IT **processes.**	18	The questions of landscape field
Describe the status of the existing organizational structure.	19	
Describe the rate of managers' support of IT**.**	20	
Describe the rate of IT **managers' support in decision-making.**	21	
Describe the rate of IT **training courses.**	22	
Describe the rate of attention to IT **in processes codification.**	23	
Describe the role of informative systems in managers' decision-making.	24	
Describe the rate of LAN.	25	The questions of dynamicity field
Describe the flexibility rate of IT **substructures against the changes.**	26	
Describe the status of encouragement for innovation and entrepreneurship.	27	
Describe the storage volume and data recovery.	28	
Describe the rate of staff resistance against IT.	29	
Describe the rate of knowledge sharing in the status quo.	30	
Describe the participation rate of users in decision-making of IT **field.**	31	
Describe the status of connection level and organizational exchanges.	32	

In addition, after finalizing the questions titles, in order to score and how to fill the questions,**EFQM** excellence model **were** used [5]. Table 2 determines scoring method.

Statistical society

The questionnaire distributed in Iran manufacturing organizations in Tehran among all **IT** staff. Our statistical society was composed of 4000 people and after completing the forms in order to use DATA MINING ALGORITHM, we had to convert the statistics in a form that DATA MINING ALGORITHM of the association rulesis appropriate.

The questionnaires were as follows:

1.2 2.4 3.1 4.5 …

The interpretation of this line of questionnaire is as follows; the respondent has selected option 2, in the response of question 1, option 4, for question 2, and option 1, for question 3, option 5 for question 5, …. We cannot use the data directionally in association rule algorithm. Because at first, each line should be composed of integers that are the representatives of items, and at the second place, there should be no duplicate numbers (items) in each line, so we encoded the related data to questionnaire as follows;

(3366) 11

(2142) 131 11

(3366) 11 means that item 11 (selecting the option 1 for the question 2), has repeated for 3366. (2142) 131 11 means that simultaneous occurrence of items 11 and 131 (selecting the option 1 for the question 2 and selecting the option 1 for the question 13) has been 2142 times in one questionnaire.

In the next stage, we extract the association rulesthat determine the relation among items. A sample of the extracted rules is as follows;

(1, 1530) 131 $\Leftarrow$ 150 21

This rule means that when item 21 and 150 have come in one series of responses, then item 131 will probably 100% follow them. Moreover, this rule has been probably 100% achieved by 1530 backed in the questionnaires.

After extracting the association rules, we will decode the extracted rules by this algorithm and produce the understandable rules based on the questionnaires. As an example let's decode the previous rule ((1, 1530) 131 $\Leftarrow$ 150 21)

"If someone selects the option 1 for the question 1 and 0 for the question15, then s/he will probably 100% selects the option 1 for the question 13 (by 1530 backed)."

At the end, the data relating to 4000 questionnaires entered the related software and 182 rules were extracted at over 50% confidence coefficient. Table 4 shows a sample of the rules with over 50% confidence coefficient.

Table 2, EFQM scoring method

100%	75%	50%	25%	5%	0%
Option 5	Option 4	Option 3	Option 2	Option 1	Option 0
Comprehensive evidence	Clear evidence	Appropriate evidence	Little evidence	No evidence or word without backing	No action has been taken
100 95 90	85 80 75 70	65 60 55 50	45 40 35 30	25 20 15	10 5 0

Table 3, a sample of extracted rules

Row						
1	11	21	$\Rightarrow$	131	1530	0.71
2	11	131	$\Rightarrow$	21	1530	0.71
3	21	131	$\Rightarrow$	11	1530	0.56
4	21	131	$\Rightarrow$	150	1530	0.56
5	21	131	$\Rightarrow$	241	1530	0.56
6	21	150	$\Rightarrow$	131	1530	1.00
7	21	241	$\Rightarrow$	131	1530	1.00
8	131	150	$\Rightarrow$	21	1530	1.00
9	131	241	$\Rightarrow$	21	1530	0.71
10	150	241	$\Rightarrow$	261	1530	1.00

Test description

Regarding table 3, one can describe the tests that have been occurred as follows;

Line 1 states that if somebody selects the option 1(no evidence or words without backing) to the second question (describe the status of existing software) and s/he selects the option 1 to the third question as well (describe

the status of information systems), this person will probably 71% select the option 1 to the fourteen question (the rate of existing standard codified).
2. The remained lines of this table can be implemented like the above-mentioned commentary by using the mentioned guidelines.
3. With putting together the obtained rules, one can achieve the maps of strategies- strategy planning in order to realize the objective (reducing the gap between current situation and desired situation) –the results of which have been mentioned in tables 4 and 5.

Table 4, the analysis of software gap and mining solutions (strategies)

Unmet (strategies)	Existing component	Desirable component
Definition and form	-	Software Engineering reference
Creation	-	Software developing framework
Creation	-	Software storage
Creation	-	Archives mechanism and software versions records
Review and approval	-	Software standards determination
Review and approval	-	Software quality index
Review and approval	-	Software outsourcing framework
Review and approval	-	Software purchase framework
Creation	-	Software development organization
Creation	-	Software quality control organization
Supply and training	Insufficient	Skilled manpower
Sufficient	Available	Software development tools
Creation	-	Infrastructure software development

Table 5, hardware gap analysis, network, security, and solutions mining(strategies)

Unmet (solution)	Existing components	Desirable components
Review and approval	-	Determination of hardware standards
Review and approval	-	Hardware quality indexes
Review and approval	-	Hardware purchase framework
Review and approval	-	Hardware purchase framework
Creation	-	Hardware quality control organization
Creation	-	Security policy codification
Upgrade	Insufficient	Security infrastructure
Upgrade	Insufficient	Network infrastructure (passive)
Upgrade	Insufficient	Network infrastructure (active)
Upgrade	Insufficient	Network infrastructure services
Creation	-	Monitoring organization and network management
Creation	-	Computational infrastructure
Upgrade	Insufficient	Data centre infrastructure
Supply and training	Insufficient	Skilled manpower

4. Conclusion

Nowadays, the main issue is not lack of information; however, it is drowning in information. Recognizing alternatives to receive useful information and extracting hidden knowledge in information need more attention. This contribution illustrates theeffect of data mining based on association rulesin strategic management. Data Mining requires a set of system appearances to derive information about system structure.

However the field is young and further research is needed to understand the full potential of refinement of decision models by information gained from Data Mining.

From the foregoing, the original focus of data mining was tabular data; an extremely effective set of techniques has been established directed at the mining of tabular data, such as; Text mining, Image Mining, Graph mining[8].

REFERENCES

[1].Agrawal R, Imielinski T, Swami A,“Mining association rules between sets of items in large databases”, In: Proceedings of the ACM SIGMOD Conference on Management of Data, 1993.

[2]. Bart Goethals, "Survey on Frequent Pattern Mining", ACM Symposium On Applied Computing, Nicosia, Cyprus, March 2004.

[3].Shenoy P, Haritsa R, et al, "Turbo-Charging vertical mining of large database", In chain Proceedings of the 2000 ACM SIGMOD International Conference on management of data, Volume 29(2), ACM Perss,2000.

[4]. Zaki M.J, "scalable algorithms for association mining", IEEE Transaction on knowledge and data Engineering, 12(3),May,June 2000.

[5].European Foundation for Quality Management, Assessor Training Models, Brussels representativeoffice Belgium, 14, 2003.

[6]. Kotler, Philip & Armstrong, Garry,"Principle of Marketing",Prentice Hall, 5 edition c1997.

[7]. Tak-chung Fu, "A review on time series data mining", Artificial Intelligence, 24, 2011, 164–181.

[8] France Coenen, "Data mining past, present and future", Cambridge University Press, Vol. 26:1, 25–29.

[9]. Stephan Meisel,"Synergies of Operations Research and Data Mining", European Journal, 206, 2010, 1–10.

[10]. David J. KetchenJr,"The intersection of strategic management and supply chain management", Industrial Marketing Management, 33, 2004, 51– 56.

[11]. J. Han, M. Kamber, "Data Mining: Concepts and Techniques", first ed, Morgan Kaufmann, San Francisco, San Diego, New York, Boston, London, Sydney, Tokyo, 2001.

[12]. I. Witten, E. Frank, "Data Mining", first ed, Hanser, Munich, Vienna, 2001.

[13]. T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning", first ed., Springer-Verlag, Berlin, Heidelberg, New York, Hong Kong, London, Milan, Paris, Tokyo, 2001.