A Competitive Intelligence framework to support decision-making based on Rough Set Theory

Fatima-Zzahra Cheffah[1], Mostafa Hanoune[2]

[1](Ben M'Sik Faculty of Science, Hassan II Casablanca University, Morocco)
[2](Ben M'Sik Faculty of Science, Hassan II Casablanca University, Morocco)

Abstract: *Given the increasing complexity of the economic context, it is important for each company to master information and build a robust strategic planning process. Competitive Intelligence (CI) is important for companies to manage their information. CI identifies opportunities and determinants of success, anticipates threats and prevents risks. CI becomes an imperative for any company wishing to sustain its growth and innovation sustainably. In addition, decision-makers have a key role to play when making decisions, some of which can have a significant impact and therefore justify the effort to reflect and deliberate on possible options before making a decision. Strategic decisions can be defined as important and far-reaching decisions in terms of actions taken, resources committed, number of actors involved and impact on all future operations. Our answer to this challenge is an approach that uses Rough set theory. Our approach is designed to support decision-making and respects the characteristics of strategic decision support in complex, uncertain and evolving situations. Rough set theory can effectively process data and information in complex system. In this article, we propose a CI approach where we used the rough set theory to generate rules in order to help decision makers make a decision in a complex and multi-criteria situation and under a context of uncertainty. We applied our model on the choice of implementation of an Enterprise Resource Planning (ERP) within the company.*
Keywords: *Competitive Intelligence, Decision-making, Strategic Information System, Rough set theory.*

I. Introduction

In a world in constant transformation, decision-making in all socio-economic sectors demands permanent reactivity, especially with the globalization that is taking place and which only increases the problems related to decision-making. Decision-making is closely linked to the control of external and internal information to the organization. At the very heart of the workings of organizations, the flow of information is constantly increasing. The exponential increase in the volume of information poses an unavoidable challenge to the management of organizations' activities and directly impacts the decision-making issue.

The democratization of the Internet, coupled with the great development of information and communication technologies, has contributed enormously to the development of information. These developments, according to Kislin, "have favored the explosion of supply and mainly the demand for information, which is essential for all human activities, both for its adaptation to its environment and for decision-making"[1]. Nevertheless, it is noteworthy that the constant challenge of obtaining relevant information that can meet the information needs of users increases in proportion to the increase in the volume of information. The question today is how to find the relevant information to the resolution of a decision problem under the best conditions of quality, delay and cost given the heterogeneity and disparity of the sources of information. Rough set theory and its method can effectively process data and information in complex system. It has become a new mathematical tool to process the fuzzy and imprecise problems. The obvious advantage of rough set theory compared with fuzzy set, evidence theory and probability theory methods for processing the uncertainty problems is that it needs not the priori information just data itself. In 1982, Z. Pawlak proposed the data analysis and reasoning theory: rough set [2]. Rough set theory attracts wide concern from researchers in artificial intelligence and pattern recognition fields because it has been applied in data mining, decision analysis, machine learning and intelligent control successfully. The main contexts of rough set theory are approximate classification, knowledge reduction (attributes or attributes values reduction), attributes dependency analysis, getting an optimal or suboptimal decision control algorithm and so on. The study of rough set theory focuses on two aspects: one is the theory research, there are a series of literatures about rough set algebra, rough set topology and its properties, rough set logic, approximate reasoning and so on, which have formed system to process incomplete, imprecise, and uncertain problems; the other is application research, to study the rough set theory applied in many areas such as medical, management, image process, decision analysis and so on. Therefore, how to process the fuzzy, imprecise and incomplete information to obtain potential, innovative and useful knowledge? It is in this context, that we propose a CI approach where we used the rough set theory to generate rules in order to help decision makers make a decision in a complex, multi-criteria and uncertain situation. We applied this approach to help decision makers choose an ERP within their company.

This paper is organized as follows. In section 2, we will present a brave explanation of the functioning of the rough set theory. Then in section 3, we will expose our proposed framework based on the rough set theory. Finally, we will present the result of the application of the framework on ERP selection within a company.

II. Related knowledge of rough set theory

1. Introduction

Rough set theory can be considered as an intelligent mathematical tool discovered by Z. Pawlak in 1982, which deals with vagueness and uncertainty and inconsistencies [2]. The general purpose of the theory is to extract and define rules from observing rough and imperfect data. When comparing rough set theory with other prediction methods, we can confirm that rough set is more appropriate and fit better in the context of prediction and selection because of the following reasons:

- Rough set is able to generate rules from inexact and undecidable data based on the concept of approximation and the concept of "possibly belong to" instead of "surely belong to".
- Rough set deals with both quantitative and categorical data.
- Rough set has proved to be very useful in practice [3].
- Each decision rule is supported by a set of real examples [4].
- Rough set generates a set of easily understandable decision rules [4], which did not require interpretation and additional information, like probability [3], [4], [5].

2. Rules generation

The aim of this step is to generate rules in the form of (if….. then). So we will use the rough set theory which can be considered as an intelligent mathematical tool discovered by Z. Pawlak, dealing with imperfect data. The general purpose of this theory is to extract and define rules from observing rough and inexact data. Rough set theory is based on two crucial concepts:

- The notion of information system
- The concept of approximation space

Information system: Let U denote a finite non-empty set of objects (actions) called the universe. Further, let A denote a finite nonempty set of attributes (type of service, target customer...). With every attribute $a \in A$, there is a function $a: U \rightarrow Va$ where Va is the set of all possible values of the attribute a, to be called the domain of a.

Based on the previous definition, an information system is a pair $S = (U, A)$.

Any sub set B of A noted $B \subseteq A$ determines a binary relation $I(B)$ on U called an indiscernibility relation and defined as follows [6]:
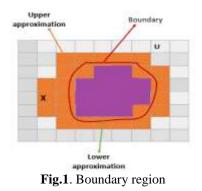
$$(x, y) \in I(B) \; if \, and \; only \; if \; a(x) = a(y) \, for \; every \; a \in A$$

Where $a(x)$ denotes the value of attribute a for element x.

Approximation: Let $S = (U, A)$ an information system, $X \subseteq U$ and $B \subseteq A$. Z. Pawlak defined two operations assigning to every $X \subseteq U$ two sets $B\text{-}(X)$ and $B^*(X)$, called the B-lower and the B-upper approximation of X, respectively, and defined as follows [6]:

$$B_*(X) = \bigcup_{x \in U} \{B(X): B(X) \subseteq X\}, \quad and \quad B^*(X) = \bigcup_{x \in U} \{B(X): B(X) \cap \neq \emptyset\}.$$

Hence, the B-lower approximation of a set is the union of all B-granules that are included in the set, whereas the B-upper approximation of a set is the union of all B-granules that have a nonempty intersection with the set. The two approximations are presented in the following figure:

Fig.1. Boundary region

As showing in the figure, we have a third region called Boundary. It is referred to as the B-boundary region of set X. Logically this region is defined as the intersection between *B*(X)* and *B-(X)*. Mathematically it is defined as follows:

$$BN_B(X) = B^*(X) - B_*(X)$$

3. Rough set theory in practice

Return to the rough set theory context, as saying above, this theory is based on the two concepts of information system and the concept of approximation. To simplify the process, we summarize rough set theory in the following main steps:

- Code the collected data in a decision table
- Extract rules from the decision table
- Extract upper approximation and lower approximation
- Extract boundary set
- Evaluate the generated rules

Knowledge reduction is important in intelligent processing, it is one of the core content in rough set theory. On general, the attributes and equivalence relations in knowledge base are not equally important, even some knowledge is necessary or redundancy. Knowledge reduction means that maintain the ability of classification of the attributes set to delete the unnecessary knowledge.

III. A rough set theory framework for decision-making: Application in ERP selection
1. Enterprise Resource Planning

Enterprise resource planning (ERP) systems are highly complex information systems. The implementation of these systems is a difficult and high cost proposition that places tremendous demands on corporate time and resources. Many ERP implementations have been classified as failures because they did not achieve predetermined corporate goals. Implementing an ERP system is not an inexpensive or risk-free venture. In fact, 65% of executives believe that ERP systems have at least a moderate chance of hurting their businesses because of the potential for implementation problems [7]. This is the reason why, we choose to apply our rough set theory approach on choosing the right ERP to implement within a company depending on critical success criteria. Numerous authors have identified a variety of factors that can be considered to be critical to the success of an ERP implementation [8].

2. A framework of the rough set theory in ERP selection

In this part, we adopt rough set theory to build an ERP selection model to help mangers make efficient decision choosing the most corresponding ERP to the company. The main idea consists in calculating lower and upper approximations based on specific characteristic of attributes and then producing the rules for services selection. When managers need to choose an ERP, they may encounter a confused situation where two different ERP providers X and Y both provide the same kind of product. According to the users' preference, it is hard to tell out whether the service from provider X is better than that from provider Y. That is to say that one property of service provided by X may be better than that service provided by Y, while Y provides a better quality of service of another property for ERP users. Even though it seems clear that the overall quality of service of X is more suitable for ERP users, it is still difficult for them to decide directly to accept the service from which provider, because higher quality of service usually means higher cost. Instead of making the choice by the managers or users themselves, a decision-making helper can choose the best service provider for the required ERP. The core part of the decision-making helper is the decision support tool. It takes the ERP users' preferences and the properties of services from different providers gathered by decision-making helper as input. It can make the best choice for ERP users, which can help them choose the service effectively and accurately.

As the knowledge is generally not equally important, with unnecessary or redundant items, knowledge reduction concept is used. Knowledge reduction aims to maintain the classification ability of the knowledge base under the certain conditions of removing unnecessary knowledge. The process of reducing information leads to a set of attributes that are independent and no further can be deleted without losing consistency. The process of reducing knowledge information is also known as attributes reduction [9]. Extracting rules from knowledge expression system is one of the main tasks in the field of data mining and knowledge discovery. Normally, four types of rules can be mined from data, such as characteristic, association, discriminator, and classification rules. Here, we focus on extracting the association rules from the information system we constructed. These rules will help users in making efficient selection of ERP. On the basis of the workflow described in the figure 2, we construct the candidate ERP selection sets and their sets of attributes (the subjective and objective evaluation parameters) to generate the information system. By combining the ERP characteristics, many metrics can be quantitatively measured (e.g., availability, elasticity, service response time, and cost per task). We can segment assessment metrics level, such as memory Reading/Writing, throughput, the

speed of CPU and so on. As the company's data security and privacy are crucial, security and privacy could also be the assessment criteria. The attribute values can be extracted from the magnanimity date sets.
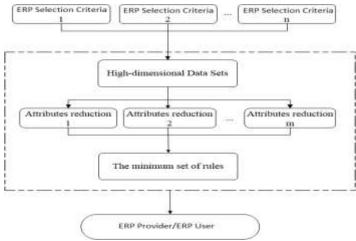

Fig.2. ERP Selection based on rough set theory

The massive amounts of raw data usually make decision process very complicated. Since rough set methods deal only with discrete attributes, a series of pre-processing such as discretization of some continuous attributes is necessary. The information system falls into two types: the complete and the incomplete information system. Incomplete information system is the one with missing values of some attributes. In reality, most of the information systems are incomplete. Recall that one of the biggest advantages of the rough set is that it can deal with imprecise, inconsistent and incomplete information, which motivate this work and the selection of this mining tool. When dealing with incomplete information systems, there are two ways to achieve knowledge reduction: First consists in changing the incomplete information system into a complete one through data remove or complement. Second is to set null as default value for missing data. After pre-processing data, attributes are reduced and the minimum set of rules is deduced. In the following, we will give an example of ERP selection based on rough set theory in which we apply knowledge reduction.

3. Application of rough set theory to sample dataset

For the sake of simplicity, we used many definition of rough set theory [9] and we established a simple instance given in Table 1. Without losing generality, we assume a complete information system, and we choose some keywords as the attributes. Then, all the attribute values are processed with the discretization. Table 1 represents the decision information system. $U = \{X1, X2... X14\}$ is the universe that corresponds to the ERP criteria set. $C = \{a1, a2, a3, a4\}$ is the set of condition attributes, where $a1, a2, a3$ and $a4$ are respectively the *response speed*, the *service feedback*, the *cost* and the *extensibility*.
$D = \{d\}$ is the decision attribute, where *(d)* is the cost effectiveness. It is easy to notice how much it is complicated with such a data set to make an efficient decision on the ERP selection, without the use of any further tool. Also, the amount of available data is pretty much higher than a table in 14 rows and 5 columns.

Table 1: The decision information system of the ERP selection

Universe	Condition Attribute				Decision Attribute
U	a_1	a_2	a_3	a_4	d
x_1	fast	bad	high	yes	low
x_2	fast	bad	high	no	low
x_3	normal	bad	high	yes	high
x_4	slow	good	high	yes	high
x_5	slow	very good	normal	yes	high
x_6	slow	very good	normal	no	low
x_7	normal	very good	normal	no	high
x_8	fast	good	high	yes	low
x_9	fast	very good	normal	yes	high
x_{10}	slow	good	normal	yes	high
x_{11}	fast	good	normal	no	high
x_{12}	normal	good	high	no	high
x_{13}	normal	bad	normal	yes	high
x_{14}	slow	good	high	no	low

The reduction algorithm of discernibility matrix is described as follows:

> **Input:** The information system of ERP selection criteria;
> **Output:** The attributes Reduction of the ERP selection criteria system: *Red*;
>   1: Input the information table of ERP selection criteria;
>   2: Set *Red=∅, count(ai)=0*, for *i=1, n*;
>   3: Compute the discernibility matrix and weight frequent of attributes *count(ai)*; \\ every new item *C* of *DM, count(ai):=count(ai)+n / | c |, ai ∈| c |*.
>   4: Merge all the same items and order the discernibility matrix according to the length of item and frequent;
>   5: for each *m* of *DM* ;
>   6: if (*m∩Red ==∅*);
>   7: choose the attribute *a* of *m*, *maxi=count(a)*;
>   8: *Red=Red∪{a}*
>   9: end if;
>   10: end for;
>   11: return *Red*.

Step 1(discernibility matrix): By using reduction method, all objects are discernible in the information system. The obtained discernibility matrix from Table1 is:

$$M_{14\times14}(DT) = (c_{ij})_{14\times14}$$

$$= \begin{bmatrix} - & & & \\ - & - & & \\ \vdots & \vdots & \ddots & \\ \{\alpha_1,\alpha_3\} & \{\alpha_1,\alpha_3,\alpha_4\} & \cdots & - \\ - & - & \cdots & \{\alpha_1,\alpha_2,\alpha_3,\alpha_4\} & - \end{bmatrix}_{14\times14}$$

Step 2 (Attributes Reduction): We reduce redundant knowledge which is invalid for making decision in Table 1 as below:
The 45 disjunctive logic expressions which meet "non empty" and "non -" are extracted from the discernibility matrix. We get:

$$\begin{aligned} L_{1,3} &= \alpha_1, \\ L_{2,3} &= \alpha_1 \vee \alpha_4, \\ L_{1,4} &= \alpha_1 \vee \alpha_2, \\ L_{2,4} &= \alpha_1 \vee \alpha_2 \vee \alpha_4, \\ &\vdots \\ L_{13,14} &= \alpha_1 \vee \alpha_2 \vee \alpha_3 \vee \alpha_4 \end{aligned}$$

After performing logical conjunction on those expressions we obtain the following conjunctive logic expression:

$$\begin{aligned} L_\wedge(\vee) &= L_{1,3} \wedge L_{2,3} \wedge L_{1,4} \wedge \cdots \wedge L_{13,14} \\ &= \alpha_1 \wedge (\alpha_1 \vee \alpha_4) \wedge (\alpha_1 \vee \alpha_2) \wedge \\ &\quad (\alpha_1 \vee \alpha_2 \vee \alpha_4) \wedge \cdots \wedge (\alpha_1 \vee \alpha_2 \vee \alpha_3 \vee \alpha_4) \end{aligned}$$

Transforming the conjunctive logic expression give the conjunctive form:

$$L'_\vee(\wedge) = (\alpha_1 \wedge \alpha_2 \wedge \alpha_4) \vee (\alpha_1 \wedge \alpha_3 \wedge \alpha_4)$$

Step 3 (Core of the attributes): the *RED_D(C)* set contains all the relative attributes reduction of the decision information system regarding the decision attribute and is given by:

$$RED_D(C) = \{\{\alpha_1,\alpha_2,\alpha_4\}, \{\alpha_1,\alpha_3,\alpha_4\}\}$$

The condition attribute *α2* or *α3* is unnecessary for decision attribute D. Thus, condition attributes *α1* and *α4* are then the core of the reduction attributes.

$$CORE_D(C) = \{\alpha_1, \alpha_4\}$$

Core is the common attributes which are in reductions sets. In other words, condition attributes *α1* and *α4* are necessary, they can never be reduced from information table. Deleting any of them will affect the classification ability with equivalence relation.

Step 4 (Generated rules): According to the two above attributes reduction results, we randomly select one of them to generate the associate rules such as the attribute reduction *α1, α2, α4*. The some decision rules are the following:

$$R1\ (\alpha_1, fast) \wedge (\alpha_2, bad) \wedge (\alpha_4, yes) \rightarrow (d, low)$$
$$R2\ (\alpha_1, fast) \wedge (\alpha_2, bad) \wedge (\alpha_4, no) \rightarrow (d, low)$$
$$R3\ (\alpha_1, general) \wedge (\alpha_2, bad) \wedge (\alpha_4, yes) \rightarrow (d, high)$$
$$R4\ (\alpha_1, general) \wedge (\alpha_2, verygood) \wedge (\alpha_4, no) \rightarrow (d, high)$$
$$R5\ (\alpha_1, general) \wedge (\alpha_2, good) \wedge (\alpha_4, no) \rightarrow (d, high)$$
$$R6\ (\alpha_1, low) \wedge (\alpha_2, good) \wedge (\alpha_4, yes) \rightarrow (d, high)$$
$$R7\ (\alpha_1, low) \wedge (\alpha_2, verygood) \wedge (\alpha_4, yes) \rightarrow (d, high)$$

As decision system contains a lot of information samples, each sample forms a basic decision rule, so there may be a lot of redundant rules. To obtain minimal decision rules to guarantee the ease of use which our main goal, we will reduce the basic set of rules. For decision rules with same decision values, if there are condition attributes with different values, then it is possible to reduce these attribute values to obtain the minimum rule set. For example, in decision rules *R1* and *R2*, the decision attribute *d* with the same value *low*, and the values of the condition attribute *α4* are different, so we can reduce these two rules. Hence, *R1* and *R2* are combined into rule *R'1*. Similarly, *R3, R4* and *R5* are combined into rule *R'2* and so on. In the following are given the minimum set of rules we obtain after reduction:

$$R_1' = (\alpha_1, fast) \wedge (\alpha_2, bad) \rightarrow (d, low)$$
$$R_2' = (\alpha_1, general) \rightarrow (d, high)$$
$$R_3' = (\alpha_1, low) \wedge (\alpha_4, yes) \rightarrow (d, high)$$

Analysis and interpretation of the results decision rules as follows:

Rule R'1: Even if *response speed* of the ERP is fast, but the *user feedback* is bad, this leads to the *cost* is low.
Rule R'2: Only if the *response speed* value of the ERP is general, even if the *cost* is high, when users choose the ERP, the values of the other indexes of the ERP can be ignored.
Rule R'3: ERP has high *cost* when it is valid of the *extensibility*, although *response speed* is low.

These three rules give meaningful information for the decision-makers to choose an ERP for their company and also the ERP providers. Company managers can rely on these rules to make efficient decision. And ERP providers can improve the quality of their ERP product focusing on particular aspects according to these decision rules.

IV. Conclusion

Rough set theory is a useful tool for analyzing big datasets, which can be used for mining the information hidden in datasets. In this article, we proposed a decision model based on the rough set theory to help decision-makers making efficient decision when they have to decide between many complex choices. On a simple example and given some key assessment attributes according to the objective and subjective metrics, we had reduced the redundant knowledge and we deduce the associate rules. Those were also reduced to get the minimal set in order to propose easy and efficient selection system. In future works, we will implement the method using Matlab, and try to evaluate the method for the parameters importance using rough set theory.

References

[1]   Kislin, Philippe. 2007. Modélisation du problème informationnel du veilleur dans la démarche d'intelligence économique. doctoral diss. in information and communication sciences. Nancy 2 University, November 2007, pp 16.
[2]   Z. Pawlak. Rough sets. International journal of computer and information sciences, pp. 341-356, 1982.
[3]   Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough Sets, Magazine Communications, 1995;Vol(38) Issue 11, 88-95
[4]   R. Slowinski, C. Zopounidis, Application of the rough set approach to evaluation of bankruptcy risk. International Journal of Intelligent Systems in Accounting, Finance, and Management, 1995; 4(1), 27–41.
[5]   Rong-Ho Lin, Yao-Tien Wang, Chih-Hung Wu, Kuan-Wei Huang, Chun-Ling Chuang Corrigendum "Developing a business failure prediction model via RST, GRA and CBR" [Experts Systems with Applications 36 (2P1) (2009) 1593–1600].
[6]   Z. Pawlak, Rough set theory and its applications, journal of telecommunications and information technology, 2002; vol.3, pp7-10
[7]   S. Cliffe, ERP implementation, Harvard Business Review77 (1)(1999), pp16–17
[8]   Elisabeth J. Umble a, Ronald R. Haft b, M. Michael Umble a; Enterprise resource planning: Implementation procedures and critical success factors; European Journal of Operational Research; Volume 146, Issue 2, 16 April 2003, pp 241-257
[9]   Y. LIU, Cloud Services Selection Based on Rough Set Theory, Technological University of Troyes, 2016, pp 47,72