

# Improving Group Decision Support Systems using Rough Set

Mohamed Eisa

Computer Science Department, Port Said University  
42526 Port Said, Egypt.

## ABSTRACT

In this paper, a proposed Group Decision Support Systems model based on Rough Set is presented. The model improves decision making process by using rough set as a tool for knowledge discovery on decision support system, where the same feature may evaluate by one decision maker as good and by another one as medium, in this case inconsistent will appear in decision problem. To cope with this problem, the model will be used to reduce inconsistent after computing lower and upper approximations. Moreover, the classification accuracy of the rough set with a single classifier and multiple classifiers was compared. These results indicate that, the model improve the classification accuracy for data sets, rather than using single and multiple classifiers.

## Keywords

Group Decision Support Systems, Rough set, Classification accuracy.

## 1. INTRODUCTION

Group Decision Support Systems (GDSSs) [1] technology supports project collaboration through the enhancement of digital communication with various tools and resources. These types of programs are used to support customized projects requiring group work, input to a group and various types of meeting protocols. Making the right decision in business is usually based on the quality of data and the ability to sift through and analyze the data to find trends in which you can create solutions and strategies. Extracting useful information and ironing out inconsistencies from raw data is a challenging task and draws the research interests of computer scientists globally [2,3], where decision quality depends on data analysis tools and techniques that used for extracting useful information from vast amounts of data. Often, traditional data analysis tools and techniques cannot be used because of the massive size of a data set. Thus, the data sets cannot be processed using existing data analysis techniques, and new methods need to be developed. Therefore, an efficient and effective information processing and decision-evaluation system becomes necessary. The past GDSSs studies concerned only with graphical representation and data analysis [4,5], in which it only focused on the graphical problem representations and a few immediate output dimensions.

Rough set theory proposed by Z.Pawlak in 1982 has been applied to many fields [6], and it is a valid mathematical method to deal with imprecise, uncertain, and vague information. The basic ideas of rough set have already been explored in many fields, such as soft computing, knowledge discovery, machine learning, GDSSs, and web intelligence. Y. Peng et al. [7] and R. Słowiński et al. [8] have done a lot of work in rough set and its applications including rough set in data mining and decision support systems. Also, rough set theory is a fundamental to many types of rule induction algorithms such as Learning from Examples using Rough set System (LERS) [9] , and algorithms developed by J. Błaszczyński et al. [10], all based on the theory proposed by Pawlak [6]. Examples of other algorithms and developments in applications and data mining can be found in the edited volumes by Lin and Cecerone [11]. Also, Lim et al. [12] presented an extensive survey of learning algorithms.

Recently, advances in the development of algorithms and increased computing power have led to some of decision support systems projects such as stock market predication, patient symptom diagnosis, telecommunication churner predication and financial bank customer attrition analysis to solve challenging business problems [13].

In this paper, a proposed Group Decision Support Systems based on Rough Set (GDSS-RS) model for combining decisions using multiple classifiers is presented. The input of the model is the decision table with multiple decision attributes and the output is the combined lower and upper approximations. Also, we implement an application begins with sampling original data into training datasets and assign each set to a classifier that calculate lower and upper approximations from it. These lower and upper calculated from all base classifiers will be combined in a new decision table with low inconsistent. The proposed model has been tested on datasets taken from the Machine Learning Database Repository at the University of California [14]. The main aim of this paper is to improve knowledge discovery on GDSSs by using rough set as a classifier. Further more, we experimentally verify, on several data sets, whether the model can achieve higher classification accuracy than other classifiers. Finally, we generate certain decision rules using MLEM2 algorithm [15].

The remainder of this paper is organized as follows: Section 2 describes the preliminaries and notations. Model construction is then presented in Sections 3, 4, followed by a description of

the experimental design and the results are presented and discussed in Section 5. The paper concludes in Section 6 with a summary and future work.

## 2. PRELIMINARIES AND NOTATIONS

### 2.1 Decision Support Systems

Decision Support Systems (DSSs) are defined as interactive computer-based intended systems intended to help decision makers utilize data and models in order to identify and solve problems and make decisions. Some key words associated with DSSs are such as: Decision Theory, Decision Analysis, Operation Research, Management Science and Artificial Intelligence (AI). However, the main problem of DSSs is to find ways of extracting useful information, called decision rules. These rules are something people using for decision making. DSSs [16] also, include a knowledge management component which stores and manages a new class of emerging AI tools such as machine learning and case-based reasoning. Machine learning refers to computational methods/tools of a computer system to learn from experience, data and observations, and consequently alter its behavior, triggered by a modification to the stored knowledge. Therefore, we are using rough set as a mathematical tool for knowledge discovery on decision support system. J. Błaszczyński et al. [10], introduce a new version of the rule induction system. This system can be used to select decision rules at each stage of decision making, where the decision rules can be store on knowledge base under the control of knowledge management system.

### 2.2 Rough Set

Rough set is a formal framework for the automated transformation of data into knowledge, especially when vague concepts and uncertain data are involved in the decision process. Also, rough set deals with the classificatory analysis of data tables and transformation of data into knowledge in the form of decision rules. These decision rules may be induced from one table that measured by one decision maker or a group of decision makers.

#### 2.2.1 Information System

Any decision problem involves a set of objects; e.g. actions, states, processes, competitors, etc. The objects are described by its attribute, and such sets of data objects and its attribute can be represented by a table its rows represent objects and columns represent attribute values.

Let  $I = (U, A)$  be an information system [17], where  $U$  is a non-empty set of finite objects (the universe) and  $A$  is non-empty finite set of attributes such that  $a: U \rightarrow V_a$  for every  $a \in A$  and  $V_a$  is the set of values that attribute  $a$  may take. In words, the information system simply assigns a value in  $V_a$  to each attribute  $a$  of each object in universe  $U$ .

#### • Lower approximation

The Lower approximation or positive region is the union of all equivalence classes in  $[x]_p$  which are contained by the given set. Formally it can be defined as the following:

$$\underline{P}(X) = \{x: [x]_p \subseteq X\}.$$

#### • Upper approximation

The upper approximation is the union of all equivalence classes in  $[x]_p$  which have non-empty intersection with the given set. Formally it can be defined as the following:

$$\bar{P}(X) = \{x: [x]_p \cap X \neq \emptyset\}.$$

#### • The accuracy of Rough set

The accuracy of the rough set representing of the set  $X$  can be given by the following:

$$\alpha P(X) = \frac{|P(X)|}{|\bar{P}(X)|}.$$

Where,  $|P(X)|$  is the number of objects contained in the lower approximation of the set  $X$ , and  $|\bar{P}(X)|$  is the number of objects contained in the upper approximation of the set  $X$ .

## 3. GDSS-RS MODEL

Determine the best decision maker views at enterprise system is a critical problem, where decision is making by a group of individuals having the necessity and the opportunity to choose between different options. Furthermore, traditional group meetings can take a long time and the resulting decisions may be mediocre. Also, getting group experts together in one place and at one time can be difficult and expensive. Therefore, Group Decision Support Systems based on Rough Set (GDSS-RS) model are introduced as interactive computer-based systems that facilitate the solution of structured and unstructured problems by a set of decision makers working together as a group. They aid groups, especially groups of managers, in analyzing problem, situations and in performing group decision making tasks. Many major decisions in organizations are made by groups as depicted in Fig 1.

The proposed model consists of sets of learning machines whose decisions are combined to improve the performance of the overall system. They are also hoping that if some learner failed, the overall system can recover the error. The basic idea to construct multiple classifiers from the original data and then aggregate their predications when classifying unknown examples as shown in Fig 2, Where  $\{D_1, D_2, \dots, D_n\}$  refer to data sets and  $\{C_1, C_2, \dots, C_n\}$  refer to multiple classifiers and  $C^*$  is the combiner of all these classifiers outputs. A voting scheme is typically employed to combine the predications, where the class that receives the highest number of votes is assigned to the test instance.

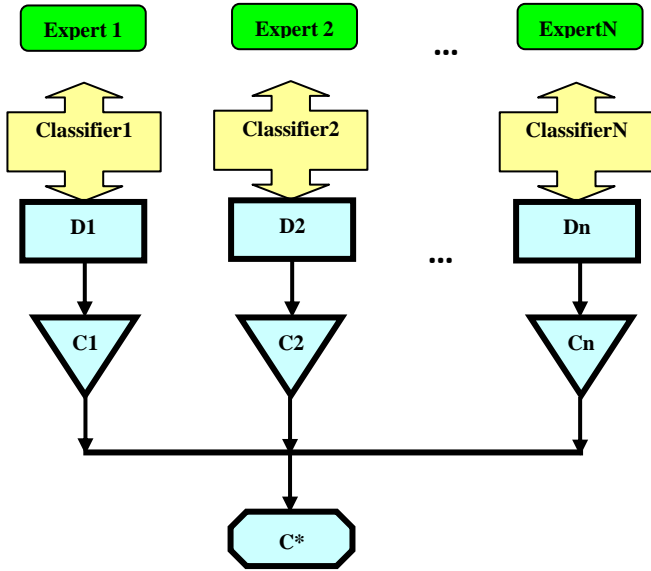


Fig 1: GDSS-RS architecture.

Decision Tables [18] are a special case of information system. They are characterized by two types of attributes namely, decision and condition attributes. Decision attribute represents the classification or partition of a data table. If there are  $n$  distinct values for a decision attribute, it implies that the data table is classified into  $n$  different groups or classes. The other set of attributes form the condition attributes whose values as a whole determine the value of the decision attribute. Formally, it can be written as  $Dt = \{U, A \cup \{d\}, V, f\}$ , where  $Dt$  is a decision table,  $U$  is the non-empty set of samples  $\{x_1, x_2, \dots, x_n\}$ , called a universe,  $A$  is a finite set of attributes called conditional attributes  $\{a_1, a_2, \dots, a_n\}$ , and  $d$  is a decision attribute, where  $d \notin A$ .  $V_a$  is the value domain of attribute  $a$ ; and  $f$  is an information function,  $f : U \times A \rightarrow V$ .

More formally, If the decision table are evaluated by group decision makers  $\{d_1, d_2, \dots, d_n\}$  called decision attributes,  $Dt = \{U, A \cup \{d_i\}, V, f\}$ ,  $i = 1, 2, \dots, n$ . Thus, we have  $n$ -lower approximations and  $n$ -upper approximations for a subset  $A$  of  $U$  respectively as the following:

$$n\text{-lower} = E_n(A) = \bigcup_{i=1}^n p_i(A),$$

$$n\text{-upper} = \bar{E}^n(A) = \bigcap_{i=1}^n \bar{p}_i(A).$$

The accuracy measure of a set ( $A$ ) under  $n$ -decision attributes can be characterized by the following coefficient

$$\alpha^{(n)}(A) = \frac{|E_n(A)|}{|\bar{E}^n(A)|}.$$

Where  $|E_n(A)|$  is the number of objects that belong to lower approximations of a set ( $A$ ) and,  $|\bar{E}^n(A)|$  is the number of objects that belong to upper approximations of a set ( $A$ ).

- 1- Let  $D$  be the training set  $\{d_1, d_2, \dots, d_n\}$ , and  $T$  be the test data.
- 2- For  $i=1$  to  $K$  do
- 3- Build a base classifier  $C_i$  from  $D_i$
- 4- End for
- 5- For each test record  $x \in T$  do
- 6-  $C^*(x) = \text{vote}(C_1(x), C_2(x), \dots, C_K(x))$
- 7- End for

Fig 2: A logical view of GDSS-RS model.

#### 4. GDSS-RS APPLICATION

GDSS-RS application creates classification models by examining already data (expertise) and inductively finding a predictive pattern. This data come from an historical database, such as people who have already undergone a particular medical treatment. They may come from an experiment in which a sample of the entire database is tested in the real world and the results used to create classifier. For example, a sample of a mailing list would be sent an offer, and the results of the mailing used to develop a classification model to be applied to the entire database. Sometimes an expert classifies a sample of the database, and this classification is then used to create the model which will be applied to the entire database.

Our application for group decision support systems based on rough set is depicted in Fig 3. It consists of three modules: preprocessing data, extracting decision rules and classification new instances. Dataset presented in tabular form is the input to the system that filtering before entering to the system because imperfect on input data. The goal of our model is to enhance learning, improve performance and generating decision rules from group decision making. Thus, we implement combiner application that combine the outputs of multiple classifiers result in better performance based on rough set theory. our application begins with sampling original data into training datasets and assign each set to a classifier that calculate lower and upper approximations from it. Also, we refer to a sample of our application code in Fig 4, implemented using C# language and SQL server data base management system as tool.

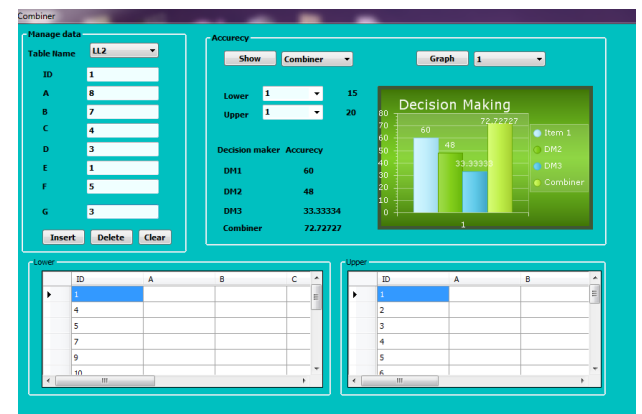


Fig 3: Group Decision Support Systems with GUI.

The main features of the application:

- ◆ Comprehensive set of data pre-processing tools.
- ◆ Graphical User Interfaces (GUI).
- ◆ Visualization and pre-processing of data including methods for discretization and missing value completion.
- ◆ Construction classifiers for both smaller and vast data sets.

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Data.SqlClient;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
```

Fig 4: Sample code of our application.

## 5. EXPERIMENTAL RESULTS

### 5.1 Data and Materials

We applied our model to data sets in Table 1, which were coming from machine learning repository [14] at the University of California and others from data sets attached with WEKA [19] and RSES [20] software.

### 5.2 The Discussion of the Results

The aim of GDSSs is to improve performance and make certain decisions; therefore, we employ rough set as a tool for knowledge discovery on GDSSs. Rough set should correctly predicate the class labels of new objects. We evaluate the performance of our model by measuring its classification accuracy. Two groups of tests are conducted in our experiments. In the first group of experiments, classification accuracy for multiple (Bagging) and individual classifiers (Naïve-bayes) running on data sets measured. The results are presented in Tables 2, where we use 10-fold cross validation. In the second group we calculate the number of certain decision rules derived from data sets. We run 10-fold cross validation with rough set to illustrate the utility of derived rules. We then present number of certain decision rules derived from our experimental data sets. Our results depicted in Fig 5. On the other hand, we compare our results for classification accuracy using rough set against the use of single and multiple classifiers. These results showed that

our model improves classification accuracy for data set more than the other classifiers.

Table 1. Experimental data sets.

Data sets	No. examples	No. attributes	No. classes
Iris	150	5	3
Labor	57	17	2
Zoo	101	17	7
Weather	14	5	2
Soybean	683	36	19
Australian	690	14	2
Heart	270	14	2
Diabetes	768	9	2
Car	1728	7	4
Cancer	569	32	2

Table 2. The classification accuracy computed by single and multiple classifiers and GDSS-RS.

Data sets	Naïve-bayes	Bagging	GDSS-RS
Iris	84.07	83	86.3
Labor	75.91	75.7	67.1
Weather	42.86	60	60
Soybean	88.3	88.9	90.3
Australian	76.81	86.7	90.4
Heart	92.97	93.4	95.7
Diabetes	95.33	94	96.2
Car	92.5	91.4	92.7
Cancer	88.9	94.2	96.7

The second observation is that rough set classifier only cannot handle continuous attributes and need for any discretization method .Also, results indicate that our approach improve the classification accuracy for data sets with large number of examples (features) , rather than using single and multiple classifiers.

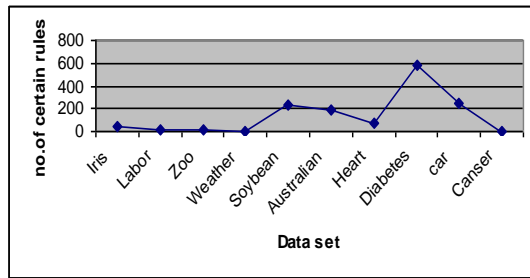


Fig 5: GDSS-RS model improved the quality of decision rules.

## 6. CONCLUSIONS AND FUTURE RESEARCH

We present an application for combining group decisions by aggregating the predications of multiple classifiers based on rough set. All of these approximations sets will be the inputs to our application. Our experimental results indicate that constructing GDSSs based on rough set improve the quality of decision rules which extracting from data. It also, improves the classification accuracy computed by classifiers. In the future work, we will define a new definition for knowledge discovery by using rough set techniques and fuzzy set. So it will be very simple and more powerful than rough set only. Also, we will make algorithm for combining decisions in general.

## 7. REFERENCES

- [1] Gray, P., "Group decision support systems" Dec. Support Syst. 3(1987)233-242.
- [2] Shih H.S., Wang C.H., Lee E.S., "A multi-attribute GDSS for aiding problem solving" Mathematical and Computer Modeling. 39(2004), 1397-1412.
- [3] Chun, K.J. , Park, H.K., "Examining the conflicting results of GDSS research" Information & Management, 33 (1998), 313-325.
- [4] Chih-Fong Tsai, Yu-Chieh Hsiao "Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches" Decision Support Systems, 50 (1) (December 2010), pp. 258–269.
- [5] Fan, SC, Shen, QP, Luo, XC , and Xue, XL "A Comparative Study of Traditional and GDSS-Supported Value Management Workshops" Journal of Construction Engineering and Management (2010).
- [6] Z.Pawlak, and A. Skorwon "Rudiments of rough sets" *Information Sciences*, vol.177, no.1, pp.3-27, 2007.
- [7] Y. Peng, Y. Zhang, Y. Tang, S.M. Li, "An incident information management framework based on data integration, data mining, and multi-criteria decision making" *Decision Support Systems*, 51 (2) (2011), pp. 316–327.
- [8] R. Słowiński, S. Greco, B. Matarazzo "Rough sets in decision making" R.A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York (2009), pp. 7753–7786.
- [9] Grzymala-Busse J.W." A new version of the rule induction system LERS" *Fundamental Informaticae* 1997; 31: 27-39. (mathematical tools to discover patterns hidden in data).
- [10] J. Błaszczyński, R. Słowiński, M., Szelag "Sequential covering rule induction algorithm for variable consistency rough set approaches" *Information Sciences*, 181 (2011), pp. 987–1002.
- [11] T. Y. Lin and N. Cercone "Rough Sets and Data Mining" Norwell, MA: Kluwer, 2000.
- [12] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Mach. Learn.*, vol. 40, pp. 203–228, 2000.
- [13] Lara Khansa, Divakaran Liginlal "Predicting stock market returns from malicious attacks: a comparative analysis of vector autoregression and time-delayed" neural networks *Decision Support Systems*, 51 (4) (2011), pp. 745–759.
- [14] C. Blake and C. J. Merz "UCI repository of machine learning databases" [Machine-readable data repository]. Univ. of California, Dept. Information and Computer Science, Irvine, CA., 2001.
- [15] J. W. Grzymala-Busse "MLEM2: A new algorithm for rule induction from imperfect data" *Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Annecy, France, IPMU 2002, July 1-5, 2002, 243-250.
- [16] Efraim Turban, Jay E. Aronson, Ting-Peng Liang (2008). "Decision Support Systems and Intelligent Systems". p. 574.
- [17] Z.Pawlak, "Rough sets and intelligent data analysis", *Information Sciences*, vol.147, pp.1-12, 2002.
- [18] Z.Pawlak, and R. Slowinski, "Rough set approach to multi-attribute decision analysis" *European Journal of Operational Research*, vol.72, pp.443-449, 1994.
- [19] The WEKA Homepage, <http://www.cs.waikato.ac.nz/~ml>, 2008.
- [20] The RSES Homepage, <http://logic.mimuw.edu.pl/~rses>