Computational Economics **19:** 287–301, 2002.

Rough Sets and Multivariate Statistical Classification: A Simulation Study

MICHAEL DOUMPOS and CONSTANTIN ZOPOUNIDIS
Technical University of Crete, Department of Production Engineering and Management, Financial Engineering Laboratory, University Campus, 73100 Chania, Greece

Abstract. The classification of a set of objects into predefined homogenous groups is a problem with major practical interest in many fields. Over the past two decades several non-parametric approaches have been developed to address the classification problem, originating from several scientific fields. This paper is focused on the rough sets approach and the investigation of its performance as opposed to traditional multivariate statistical classification procedures, namely the linear discriminant analysis, the quadratic discriminant analysis and the logit analysis. For this purpose an extensive Monte Carlo simulation is conducted to examine the performance of these methods under different data conditions.

Key words: classification, Rough sets theory, multivariate statistics, Monte Carlo simulation

1. Introduction

The classification problem involves the assignment of a set of observations (objects, alternatives) described over some attributes or criteria into predefined homogeneous classes. Such problems are often encountered in many fields, including finance, marketing, agricultural management, human recourses management, environmental management, medicine, pattern recognition, etc. This major practical interest of the classification problem has motivated researchers in developing an arsenal of methods in order to develop quantitative models for classification purposes. The Linear Discriminant Analysis (LDA) was the first method developed to address the classification problem from a multidimensional perspective. LDA has been used for decades as the main classification technique and it is still being used at least as a reference point for comparing the performance of new techniques that are developed. Other widely used parametric classification techniques developed to overcome LDA's restrictive assumptions (multivariate normality, equality of dispersion matrices between groups) include Quadratic Discriminant Analysis (QDA), logit and probit analysis and the linear probability model (for two-group classification problems).

During the last two decades several alternative non-parametric classification techniques have been developed, including mathematical programming techniques (Freed and Glover, 1981), multicriteria decision aid methods (Doumpos et al., 2000), neural networks (Patuwo et al., 1993), and machine learning approaches (Quinlan, 1986). Among these techniques, the rough sets theory developed following the concepts of machine learning, has several distinguishing and attractive features, including data reduction, handling of uncertainty, ease of interpretation of the developed classification model, etc. The purpose of this paper is to explore whether these attractive features also lead to higher efficiency in terms of classification accuracy, as opposed to traditional statistical classification procedures. For this purpose an extensive Monte Carlo simulation is conducted. Although there has been some research regarding the comparison of rough sets with statistical classification procedures (Teghem and Benjelloun, 1992), this is the first study to perform a thorough comparison on different data conditions.

The rest of the paper is organized as follows. Section 2 presents a brief outline of the basic concepts of the rough sets approach. Section 3 provides details on the design of the Monte Carlo simulation and the factors considered in the conducted experiments. Section 4 discusses the obtained results for the rough sets approach and compares them with the results of LDA, QDA and logit analysis.

2. Brief Outline of the Rough Sets Approach

Pawlak (1982) introduced the rough sets theory as a tool to describe dependencies between attributes, to evaluate the significance of attributes and to deal with inconsistent data. As an approach to handle imperfect data (uncertainty and vagueness), it complements other theories that deal with data uncertainty, such as probability theory, evidence theory, fuzzy set theory, etc. Generally, the rough set approach is a very useful tool in the study of classification problems, regarding the assignment of a set of objects into pre-specified classes. Recently however, there have been several advances in this field to allow the application of the rough set theory to choice and ranking problems as well (Greco et al., 1997).

The rough set philosophy is founded on the assumption that with every object some information (data, knowledge) is associated. This information involves two types of attributes; condition and decision attributes. Condition attributes are those used to describe the characteristics of the objects. For instance the set of condition attributes describing a firm can be its size, its financial characteristics (profitability, solvency, liquidity ratios), its organization, its market position, etc. The decision attributes define a partition of the objects into classes according to the condition attributes.

On the basis of these two types of attributes an information table is formed. The rows of the information table represent the objects and its columns represent the condition and decision attributes. Objects characterized by the same information are considered to be indiscernible. This indiscernibility relation constitutes the main mathematical basis of the rough set theory. Any set of all indiscernible objects is called an elementary set and forms a basic granule of knowledge about the universe. Any set of objects being a union of some elementary sets is referred to as crisp (precise) otherwise as rough (imprecise, vague). Consequently, each rough set has a boundaryline consisting of cases (objects) which cannot be classified with certainty as members of the set or of its complement. Therefore, a pair of crisp sets, called the lower and the upper approximation can represent a rough set. The lower approximation consists of all objects that certainly belong to the set and the upper approximation contains objects that possibly belong to the set. The difference between the upper and the lower approximation defines the doubtful region, which includes all objects that cannot be certainly classified into the set. On the basis of the lower and upper approximations of a rough set, the accuracy of approximating the rough set can be calculated as the ratio of the cardinality of its lower approximation to the cardinality of its upper approximation.

On the basis of these approximations, the first major capability that the rough set theory provides is to reduce the available information, so as to retain only what is absolutely necessary for the description and classification of the objects being studied. This is achieved by discovering subsets of the attributes' set that describes the considered objects, which can provide the same quality of classification as the whole attributes' set. Such subsets of attributes are called reducts. Generally, the reducts are more than one. In such a case the intersection of all reducts is called the core. The core is the collection of the most relevant attributes, which cannot be excluded from the analysis without reducing the quality of the obtained description (classification). The decision maker can examine all obtained reducts and proceed to the further analysis of the considered problem according to the reduct that best describes reality. Heuristic procedures can also be used to identify an appropriate reduct (Slowinski and Zopounidis, 1995).

The subsequent steps of the analysis involve the development of a set of rules for the classification of the objects into the classes where they actually belong. The rules developed through the rough set approach have the following form:

IF *conjunction of elementary conditions*
THEN *disjunction of elementary decisions* .

The developed rules can be consistent if they include only one decision in their conclusion part, or approximate if their conclusion involves a disjunction of elementary decisions. Approximate rules are consequences of an approximate description of decision classes in terms of blocks of objects (granules) indiscernible by condition attributes. Such a situation indicates that using the available knowledge, one is unable to decide whether some objects belong to a given decision class or not. Each rule is associated with a strength, that represents the number of objects satisfying the condition part of the rule and belonging to the class that the conclusion of the rule suggests (for approximate rules, the strength is calculated for each class separately). The higher the strength of the rules, the more general the rule is considered to be (its condition part is shorter and less specialized).

The rule induction approach employed in this study is based on the MODLEM algorithm of Grzymala-Busse and Stefanowski (2000). This algorithm is an extension of the LEM2 algorithm (Grzymala-Busse, 1992), which is among the most popular rule induction approaches within the context of the rough sets theory. The major feature of the MODLEM algorithm is that it does not require the realization of a discretization of the attributes values prior to its application. Therefore, it is well suited to problems involving continuous attributes such as the ones studied through the Monte Carlo simulation that is employed in this study. Furthermore, the algorithm leads to the development of a compact set of rules with high strength (Grzymala-Busse and Stefanowski, 2000). The algorithm is applied in combination with the LERS classification system (Grzymala-Busse, 1992), which enables the handling of cases where an object's description matches more than one conflicting rules, or cases where an object does not match any rule.

More details on the rough sets theory can be found in Pawlak and Slowinski (1994a, b).

3. Experimental Design

In order to perform a thorough examination of the classification performance of the rough sets theory as opposed to traditional classification procedures, an extensive experimental study is conducted using several different data conditions. Except rough sets, three other statistical and econometric procedures are considered: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Logit Analysis (LA). The factors considered in this experimental design are illustrated in Table I.

The first of the factors involving the properties of the data investigated in the conducted experimental design involve their distributional form (F_2). While many studies conducting similar experiments have been concentrated on univariate distributions to consider multivariate non-normality, in this study multivariate distributions are considered. This specification enables the investigation of additional factors in the experiment, such as the correlation of the attributes and the homogeneity of the group dispersion matrices. Actually, using a univariate distribution implies that the attributes are independent, a case that is hardly the situation encountered in real-world problems. The first two of the multivariate distributions that are considered (normal and uniform) are symmetric, while the exponential and log-normal distributions are asymmetric, thus leading to a significant violation of multivariate normality. The generation of the multivariate non-normal data is based on the methodology presented by Vale and Maurelli (1983).

Factor F_3 defines the number of groups into which the classification of the objects is made. In this experimental design the two-group and the three-group classification problems are considered. This specification enables the derivation of useful conclusions on the performance of the methods investigated, in a wide range of situations that are often met in practice (many real-world classification problems

Table I. Factor investigated in the experimental design.

Factors		Levels	
F_1	Classification procedures	1.	Linear Discriminant Analysis (LDA)
		2.	Quadratic Discriminant Analysis (QDA)
		3.	Logit Analysis (LA)
		4.	Rough sets (ROUGH SETS)
F_2	Statistical distribution of the data	1.	Multivariate normal
		2.	Multivariate uniform
		3.	Multivariate exponential
		4.	Multivariate log-normal.
F_3	Number of groups	1.	Two
		2.	Three
F_4	Size of the training sample	1.	36 objects, 5 attributes
		2.	72 objects, 5 attributes
		3.	108 objects, 5 attributes
F_5	Correlation coefficient	1.	Low correlation: $r \in [0, 0{,}1]$
		2.	Higher correlation: $r \in [0{,}2, 0{,}5]$
F_6	Homogeneity of the group dispersion matrices	1.	Equal
		2.	Unequal
F_7	Group overlap	1.	Low overlap
		2.	High overlap

involve three groups). The fourth factor is used to define the size of the training sample, and in particular the number of objects that it includes (henceforth this number is denoted as m). The factor has three levels corresponding to 36, 72 and 108 objects, distributed equally to the groups defined by factor F_3. In all three cases the objects are described along five attributes. Generally, small training samples contain limited information about the classification problem being examined, but the corresponding complexity of the problem is also limited. On the other hand, larger sample provides richer information, but the complexity is also increased. Thus, the examination of the three levels of the factor enables the investigation of the performance of the classification procedures under all these cases.

The specified correlation coefficients for every pair of attributes define the off-diagonal elements of the dispersion matrices of the groups. The elements in the diagonal of the dispersion matrices, representing the variance of the attributes are specified by the sixth factor, which is considered in two levels. In the first level, the variances of the attributes are equal for all groups, whereas in the second level the variances differ. Denoting the variance of attribute x_i for group j as σ_{ij}^2, the realization of these two situations regarding the homogeneity of the group dispersion matrices is performed as follows:

– For the multivariate normal, uniform and exponential distributions:

$$\text{Level 1: } \sigma_{i1}^2 = \sigma_{i2}^2 = \sigma_{i3}^2 = 1, \quad \forall i = 1, 2, \ldots, 5\,.$$

$$\text{Level 2: } \sigma_{i1}^2 = 1, \sigma_{i2}^2 = 4, \sigma_{i3}^2 = 16, \quad \forall i = 1, 2, \ldots, 5\,.$$

– For the multivariate log-normal distribution, the variances are specified so as to assure that the kurtosis of the data ranges within reasonable levels,[1] as follows:

a) In the case of two groups:

$$\text{Level 1: } \sigma_{i1}^2 = \sigma_{i2}^2 = \begin{cases} 12, & \text{if } m = 36 \\ 14, & \text{if } m = 72 \\ 16, & \text{if } m = 108 \end{cases}, \quad \forall i = 1, 2, \ldots, 5\,.$$

$$\text{Level 2: } \sigma_{i1}^2 = \begin{cases} 12, & \text{if } m = 36 \\ 14, & \text{if } m = 72 \\ 16, & \text{if } m = 108 \end{cases}, \quad \begin{array}{l} \text{and } \sigma_{i2}^2 = 1.5\sigma_{i1}^2, \\ \forall i = 1, 2, \ldots, 5\,. \end{array}$$

b) In the case of three groups:

$$\text{Level 1: } \sigma_{i1}^2 = \sigma_{i2}^2 = \sigma_{i3}^2 = \begin{cases} 4, & \text{if } m = 36 \\ 7, & \text{if } m = 72 \\ 10, & \text{if } m = 108 \end{cases}, \quad \forall i = 1, 2, \ldots, 5\,.$$

$$\text{Level 2: } \sigma_{i1}^2 = \begin{cases} 2, & \text{if } m = 36 \\ 4, & \text{if } m = 72 \\ 6, & \text{if } m = 108 \end{cases}, \quad \begin{array}{l} \text{and } \sigma_{i2}^2 = 1.5\sigma_{i1}^2, \sigma_{i3}^2 = 1.5\sigma_{i2}^2\,, \\ \forall i = 1, 2, \ldots, 5\,. \end{array}$$

The final factor defines the degree of group overlap. The higher the overlapping is between each pair of groups, the higher is the complexity of the classification problem due to the difficulty in discriminating between the objects of each group. The degree of group overlap in this experimental design is considered using the Hotelling T^2 statistic. This is a multivariate measure of difference between the means of two groups, assuming that the attributes are multivariate normal and that the group dispersion matrices are equal. Studies conducted on the first of these assumptions (multivariate normality) have shown that actually the Hotelling T^2 is quite robust to departures from multivariate normality even for small samples (Mardia, 1975). Therefore, using the Hotelling T^2 in the multivariate distributions considered in this experimental design does not pose a significant problem. To overcome the second assumption regarding the homogeneity of the group dispersion matrices, the modified version of the Hotelling T^2 defined by Anderson (1958) is employed in the case where the dispersion matrices are not equal. The use of these measures of group overlap in the conducted experimental design is performed as follows: Initially, the means of the all five attributes for the first group is fixed to a specific value (1 for the case of multivariate normal, uniform and exponential distribution, and 8 in the case of the log-normal distribution). Then, the means of the attributes for the second group are specified so as the Hotelling T^2 (or its modified version) between the means of groups 1 and 2 is significant at the 1% and the 10% significance level, corresponding to low and high degree of group overlap. Similarly, the means of the third group are specified so as the Hotelling T^2 (or its modified version) between the means of groups 2 and 3 is significant at the 1% and the 10% significance level.

For each combination of the factors F_2–F_7 (192 combinations) a training sample and a validation sample are generated, having all the properties that these factors specify. The size of the training sample is defined by the factor F_4, while the size of the validation sample (holdout sample) is fixed at 216 in all cases. For each factor combination 20 replications are performed. Therefore, during this experiment the number of samples considered is 7,680 ($192 \times 20 = 3{,}840$ training samples matched to 3,840 validation samples). Overall, the conducted experiment involves a $4 \times 4 \times 2 \times 3 \times 2 \times 2 \times 2$ full-level factorial design consisting of 768 treatments (factor combinations).

4. Results

The analysis of the results obtained in this experimental design, is focused only in classification errors for the validation samples. The rough sets approach correctly classify all objects in the training sample and consequently there is no point of conducting a comparison with the statistical classification procedures with regard to their performance in the training sample.

The examination of the results obtained from the experimental design is performed through a seven-way analysis of variance, using the transformed misclassification rates of the methods, on the basis of the transformation:

$$2 \arcsin \sqrt{\text{error rate}} \,. \tag{1}$$

This transformation has been proposed by several researchers in order to stabilize the variance of the misclassification rates (Bajgier and Hill, 1982; Joachimsthaler and Stam, 1988). The ANOVA results presented in Table II, indicate that the seven main effects (the considered factors), three two-way interaction effects and one three-way interaction effect explain more than 76% of the total variance measured using the Hays ω^2 statistic. None of the remaining interaction effects explains more than 1% of the total variance, and consequently they are not reported.

In these results the interaction effects are of major interest. All four interaction effects that are found to explain a high proportion of the total variance in the obtained results, involve the interaction of the factor F_1 (classification procedures) with other factors, in particular the homogeneity of the group dispersion matrices (F_6), the distributional form of the data (F_2), and the training sample size (F_4).

Table II. Major explanatory effects regarding the classification performance of the methods (seven-way ANOVA results).

Effects	df	Sum of squares	Mean squares	F	ω^2
F_6	1	198.249	198.249	198.25	17.76
$F_1 \times F_6$	3	132.835	44.278	132.83	11.90
$F_1 \times F_2$	9	122.608	13.623	122.61	10.97
F_1	3	84.637	28.212	84.64	7.58
F_3	1	83.150	83.150	83.15	7.45
F_2	3	71.409	23.803	71.41	6.39
F_4	2	47.897	23.949	47.90	4.29
F_7	1	42.719	42.719	42.72	3.83
$F_1 \times F_4$	6	29.572	4.929	29.57	2.64
F_5	1	24.603	24.603	24.60	2.20
$F_1 \times F_5 \times F_6$	9	12.452	1.384	12.45	1.11

Table III. Tukey's test for significant differences between methods along all experiments.

	Mean	Tukey's grouping
LDA	1.2000 (32.20%)	C
QDA	1.0671 (26.99%)	B
LOGIT	1.1891 (31.69%)	C
ROUGH SETS	1.0302 (25.65%)	A

Table III summarizes the results of all methods throughout all experiments, while Tables IV–VII provide further details on the comparison of the methods in terms of the aforementioned two and three-way interaction effects that are found significant through the analysis of variance. Each of these tables reports the average transformed error rate, the true error rate (in parentheses) and the grouping obtained through the Tukey's test on the average transformed error rates [cf. Equation (1)].

The results indicate that, overall, the rough sets approach outperforms all statistical procedures, followed by QDA, while the performances of LDA and LA are similar. The further analysis indicates that when the homogeneity of the dispersion matrices is considered, the rough sets approach is the best classifier in the cases where the dispersion matrices are equal, while in the opposite case the QDA outperforms all the other procedures. However, the results of Table IV clearly indicate that QDA is very sensitive to departures from its assumption regarding the heterogeneity of the group dispersion matrices.

Table IV. Tukey's test for significant differences between methods (factor: F_5).

	Homogeneity of dispersion matrices			
	Equal		Unequal	
	Mean	Tukey's grouping	Mean	Tukey's grouping
LDA	1.2391 (33.98%)	C	1.1610 (30.42%)	C
QDA	1.3283 (38.17%)	D	0.8059 (15.80%)	A
LOGIT	1.2183 (33.01%)	B	1.1599 (30.37%)	C
ROUGH SETS	1.1552 (30.89%)	A	0.9051 (20.42%)	B

The distributional form of the data is also a significant factor in explaining the differences in the error rates of the methods. The results of Table V show that the rough sets approach provides significantly lower error rates than all the statistical procedures, when the data come from non-symmetric distributions, such as the exponential and the log-normal distribution. On the contrary, when the underlying distribution of the data is symmetric then QDA is found to be the best classifier, even in the case of multivariate non-normality (uniform distribution). This is not surprising; Clarke et al. (1979) investigated the performance of QDA for departures from non-normality and they concluded that QDA is quite robust to non-normal data, expect for the case where skewness is high. This justifies the results of this experimental design.

The consideration of the interaction between the homogeneity of the group dispersion matrices and the distributional form of the data (Table VII) clarifies further the above results. According to the results of Table VII, the LDA and the LA outperform both rough sets and QDA when the distribution is symmetric and the group dispersion matrices are equal, whereas in the case of unequal group dispersion matrices QDA outperforms all procedures, followed by the rough sets approach. In the cases of the exponential and the log-normal distributions, which are asymmetric, the rough sets approach provides consistently lower error rates than the statistical procedures, except for the case of the log-normal distribution with unequal group dispersion matrices where the QDA provides the best performance. However, once again the high sensitivity of the QDA to departures from its assumption regarding the heterogeneity of the group dispersion matrices, becomes apparent. For all distributional forms with equal group dispersion matrices, the QDA performs worse than the other classification procedures, expect for the multivariate normal distribution where it outperforms rough sets, whereas its performance when the group dispersion matrices are equal is greatly improved.

Finally, the size of the training sample is also found to be a significant factor. The results of Table VI indicate that the rough sets approach provides consistently

Table V. Tukey's test for significant differences between methods (factor: F_1).

	Distribution							
	Normal		Uniform		Exponential		Log-normal	
	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping
LDA	1.1900 (31.69%)	B	1.2312 (33.57%)	C	1.1642 (30.66%)	C	1.2147 (32.86%)	C
QDA	1.0917 (27.79%)	A	1.0634 (27.02%)	A	1.0392 (26.97%)	B	1.0740 (27.16%)	B
LOGIT	1.1827 (31.36%)	B	1.2200 (33.03%)	C	1.1517 (30.11%)	C	1.2020 (32.25%)	C
ROUGH SETS	1.2795 (35.97%)	C	1.1219 (29.30%)	B	0.6769 (11.83%)	A	1.0422 (25.51%)	A

Table VI. Tukey's test for significant differences between methods (factor: F_4).

	Training sample size					
	36		72		108	
	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping
LDA	1.1092 (26.04%)	B	1.2342 (33.68%)	C	1.3017 (36.87%)	B
QDA	1.0585 (26.26%)	B	1.0686 (27.11%)	B	1.0741 (27.58%)	A
LOGIT	1.0555 (25.69%)	B	1.2225 (33.11%)	C	1.2894 (36.27%)	B
ROUGH SETS	1.0099 (24.51%)	A	1.0205 (25.27%)	A	1.0601 (27.19%)	A

lower error rates both when small and larger training samples are employed. The differences between the rough sets and the statistical classification procedures are significant for small and medium size training samples (36 and 72 objects), while in the case of a larger training sample the difference between the rough sets and the QDA is not significant. The performances of LDA and LA deteriorate significantly as the training sample size increases, whereas the rough sets approach and QDA provide more robust results. In the case of the QDA, its improvement relative to the other procedures when the training sample size increases can be justified on the basis that larger samples provide a better representation of the true form of the group dispersion matrices, thus enabling the QDA to take full advantage of its assumption regarding the heterogeneity of the group dispersion matrices.

5. Conclusions

The aim of this study was to explore the performance of a machine learning approach, namely the rough sets theory in classification problems. Over the past two decades, the rough sets theory has emerged as a significant methodological tool to address complex decision-making problems, where a classification of the considered objects is involved.

A thorough comparison was performed with traditional statistical classification approaches, on the basis of an extensive experimental design involving several factors regarding the properties of the data involved. The results indicate that the rough sets could be considered as a promising classification approach compared to well-established existing procedures, at least in the cases where the data originate from asymmetric distributions. The major 'rival' of the rough sets, the QDA also

Table VII. Tukey's test for significant differences between methods (factors: F_2, F_5).

	Distribution							
	Normal				Uniform			
	Homogeneity of dispersion matrices				Homogeneity of dispersion matrices			
	Equal		Unequal		Equal		Unequal	
	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping
LDA	1.2200 (33.08%)	A	1.1601 (30.30%)	B	1.2603 (34.94%)	A	1.2021 (32.20%)	C
QDA	1.3164 37.60%	B	0.8671 (17.99%	A	1.3460 39.00%)	B	0.7808 (15.04%)	A
LOGIT	1.2033 (32.34%)	A	1.1622 (30.39%)	B	1.2403 (33.98%)	A	1.1996 (32.07%)	C
ROUGH SETS	1.4090 (42.08%)	C	1.1500 (29.86%)	B	1.3386 (38.70%)	B	0.9053 (19.90%)	B

Table VII. (Continued)

	Distribution							
	Exponential				Log-normal			
	Homogeneity of dispersion matrices				Homogeneity of dispersion matrices			
	Equal		Unequal		Equal		Unequal	
	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping	Mean	Tukey's grouping
LDA	1.2429 (34.18%)	C	1.0855 (27.14%)	C	1.2332 (33.70%)	B	1.1961 (32.02%)	C
QDA	1.3255 (38.07%	D	0.7529 13.88%)	B	1.3251 (38.02%)	C	0.8229 (16.30%)	A
LOGIT	1.2107 (32.71%)	B	1.0927 (27.51%)	C	1.2189 (33.02%)	B	1.1851 (31.49%)	C
ROUGH SETS	0.7631 (14.54%)	A	0.5908 (9.12%)	A	1.1100 (28.24%)	A	0.9744 (22.78%)	B

performs well in many cases, but the form of the group dispersion matrices heavily affects its performance.

These results are encouraging with regard to the performance of the rough sets approach in addressing classification problems. Further analysis regarding additional data conditions that are commonly encountered in real-world problems, such as the existence of qualitative data and outliers, could provide a global view of the true performance of the rough sets approach in a wider range of complex data conditions.

However, despite these issues that need further research and the findings of this extensive experimental study, there are two significant features of the rough sets theory that should be emphasized, with regard to providing meaningful decision support. First, they provide a sound mechanism for data reduction, and second they enable the development of rule-based classification models that are easy to understand. This second feature is very significant in terms of decision support, and it is hardly shared by the statistical approaches explored in this study. These features have been the main reasons for the widespread application of rough sets in addressing numerous real-world problems from the fields of finance (Dimitras et al., 1999), engineering (Stefanowski et al., 1992), automatic control (Szladow and Ziarko, 1992), medicine (Tsumoto, 1998), etc.

Note

[1] In contrast to the other distributions considered in this experimental design where the coefficients of skewness and kurtosis are fixed, in the log-normal distribution the skewness and kurtosis are defined by the mean and the variance of the attributes for each group. The procedures for generating multivariate non-normal data are able to replicate satisfactory the prespecified values of the first three moments (mean, standard deviation and skewness) of a statistical distributions. However, the error is higher for the fourth moment (kurtosis). Therefore, in order to reduce this error and consequently to have better control of the generated data, both the mean and the variance of the attributes for each group in the case of the multivariate log-normal distribution are specified so as the coefficient of kurtosis is lower than 40.

References

Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Bajgier, S.M. and Hill, A.V. (1982). An experimental comparison of statistical and linear programming approaches to the discriminant problem. *Decision Sciences*, **13**, 604–618.

Clarke, W.R., Lachenbruch, P.A. and Broffitt, B. (1979). How non-normality affects the quadratic discriminant function. *Communications in Statistics: Theory and Methods*, **8** (13), 1285–1301.

Dimitras, A.I., Slowinski, R., Susmaga, R. and Zopounidis, C. (1999). Business failure prediction using rough sets. *European Journal of Operation Research*, **114**, 49–66.

Doumpos, M., Zopounidis, C. and Pardalos, P.M. (2000). Multicriteria sorting methodology: Application to financial decision problems. *Parallel Algorithms and Applications*, **15** (1–2), 113–129.

Freed, N. and Glover, F. (1981). Simple but powerful goal programming models for discriminant problems. *European Journal of Operational Research*, **7**, 44–60.

Greco, S., Matarazzo, B. and Slowinski, R. (1997). Rough set approach to multi-attribute choice and ranking problems. In G. Fandel and T. Gal (eds.), *Multiple Criteria Decision Making*. Springer-Verlag, Berlin, pp. 318–329.

Grzymala-Busse, J.W. (1992). LERS – A system for learning from examples based on rough sets. In R. Slowinski (ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer Academic Publishers, Dordrecht, pp. 3–18.

Grzymala-Busse, J.W. and Stefanowski, J. (2000). Three discretization methods for rule induction. *International Journal of Intelligent Systems* (to appear).

Joachimsthaler, E.A. and Stam, A. (1988). Four approaches to the classification problem in discriminant analysis: An experimental study, *Decision Sciences*, **19**, 322–333.

Mardia, K.V. (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test. *Applied Statistics*, **24**, 163–171.

Patuwo, E., Hu, M.Y. and Hung, M.S. (1993). Two-group classification using neural networks. *Decision Sciences*, **24**, 825–845.

Pawlak, Z. (1982). Rough sets. *International Journal of Information and Computer Sciences*, **11**, 341–356.

Pawlak, Z. and Slowinski, R. (1994a). Decision analysis using rough sets, *International Transactions in Operational Research*, **1** (1), 107–114.

Pawlak, Z. and Slowinski, R. (1994b). Rough set approach to multi-attribute decision analysis, *European Journal of Operational Research*, **72**, 443–459.

Quinlan J.R. (1986). Induction of decision trees. *Machine Learning*, **1**, 81–106.

Slowinski, R. and Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management*, **4**, 27–41.

Stefanowski, J., Slowinski, R. and Nowicki, R. (1992). The rough sets approach to knowledge analysis for classification support in technical diagnostics of mechanical objects. In F. Belli and F.J. Radermacher (eds.), *Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, Lecture Notes in Economics and Mathematical Systems 604*, Springer-Verlag, Berlin, pp. 324–334.

Szladow, A.J. and Ziarko, W.P. (1992). Knowledge-based process control using rough sets. In R. Slowinski (ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, pp. 49–60.

Teghem, J. and Benjelloun, M. (1992). Some experiments to compare rough sets theory and ordinal statistical methods. In R. Slowinski (ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, pp. 267–286.

Tsumoto, S. (1998). Automated extraction of medical expert system rules from clinical databases based on rough set theory. *Information Sciences*, **112**, 67–84.

Vale, D.C. and Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, **48** (3), 465–471.

Ziarko, W., Golan, D. and Edwards, D. (1993). An application of DATALOGIC/R knowledge discovery tool to identify strong predictive rules in stock market data. In *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, 89–101. Washington D.C.