INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. **12**: 187–213 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/isaf.242

A MODEL FOR VALUATION OF THE BRANCH OFFICES OF A SAVINGS BANK BASED ON ROUGH SETS

C. PÉREZ-LLERA,[a]* C. FERNÁNDEZ-BAIZÁN,[b] J. L. FANJUL[c] AND J. FEITO[c]

[a] *Dpt. Informática, Universidad de Oviedo, Spain*
[b] *Dpt. Lenguajes, Sistemas Informáticos e Ingeniería del Software, Universidad Politécnica de Madrid, Spain*
[c] *Dpt. Dirección y Economía de la Empresa, Universidad de León, Spain*

SUMMARY

This work proposes a model for the valuation of branch offices of banks based on the rough set theory, which could be used as the basis for a decision-making system for dimensioning strategies of a financial entity. It compares the rough set approach with the competitive discriminant analysis methodology using a common set of data from 421 branches. We pay special attention to data reduction and the creation of decision rules that will allow future branches to be classified. These rules could constitute the basis for the evaluation of the viability of dimensioning strategies for a financial entity. In order to evaluate the predictive capabilities of the decision rules, we present the results of cross-validation tests to evaluate the ability of the model to classify new branches. It appears that the rough sets approach provides a favourable tool for the valuation of branch offices.

1. INTRODUCTION

The environment of Spanish savings banks has passed from a simple and stable state, to the current dynamic and complex situation. The flexibility that has thus far characterized these institutions in the process of adaptation is insufficient to confront the competitive situation in the banking sector today. For this reason, savings banks must develop a series of strategies to assure their viability in the mid to long-term, by means of the strengthening of their internal structure. The structure of these companies must allow for the demands of the market, which makes dimensioning strategies necessary.

The preliminary work carried out in this field shows the absence of any relationship between the global parameters, that is the number of offices and the number of employees per office, and the profitability of the entity. It does seem, however, that there is a direct relationship between the resources of the entity and the number of new openings foreseen in the short to mid-term. In studying a concrete dimensioning strategy, which contemplates openings, closings, relocations, expansions and structural or operative modifications, our goal is to determine whether it is an optimal strategy given the expectations and the capacity for development of the entity.

In general, there are two initial aspects to consider in the processes of rationalization of the commercial networks of Spanish savings banks: *consolidation* (in the market in which they operate, as compared to their competitors) and *expansion* (both inside and outside the borders of their historical/geographical environment, either national or international).

* Correspondence to: C. Pérez-Llera, Dpt. Informática, Universidad de Oviedo, Gijón, Spain.
E-mail: cpllera@uniovi.es

Projects of consolidation and expansion can be individual or collective: limited to a direct positioning of the entity in a certain geographical space, or be dimensioning strategies common to several entities (coalitions, acquisitions, etc.). These become processes of a certain complexity, affecting all levels of the institutions involved.

There are many theories concerning the valuation and analysis of the productivity of financial entities; some contradictary.

Our contribution consists of valuing the profitability of the entity in question, based on the following premise: '*The profitability of the whole (the savings bank) is determined by the profitability of the parts (the branches that integrate the commercial network) +/− the effect of the boundary conditions (positive or negative administration of the central services and their capacity to adapt in the face of the constant changes these institutions undergo).*' We will consider the branch as a unit to measure the return of the entity.

A *strategy of optimal dimensioning* for a savings bank is one that allows maximum profits to be obtained while using minimum resources (economic, human, technical, etc.). Everything is imputed individually to each one of the branches that make up the commercial network as a function of the branches' *nature* (rural or urban), *typology* (size, human resources endowment, volume of business, etc.) and *degree of specialization* (companies, trades, etc.).

In addition to the *qualitative variables* previously referred to (nature, typology and degree of specialization), the analysis and valuation of the branches involves several *quantitative variables* that condition the individual and collective profitability of these units of business:

- Volume of business (resources and investments);
- Operating costs (general expenses, personnel expenses);
- Cost of liabilities;
- Return of assets;
- Ordinary income generated by the intermediation of financial and parafinancial products;
- Reinvestment rate of assets;
- Standard return (defined according to nature, typology and degree of specialization), etc.

The following *positioning variables* must be considered in conjunction with the variables listed above:

- Ratio of banking concentration;
- Market share;
- Coverage of the dominant entity;
- Target segments;
- Socio-economic level of the geographical unit (municipality, area of influence, etc.);
- Financial consumption pattern;
- Expected response.

These positioning variables also depend on the geographical location of the branch. Together, these variables generate the *valuation matrix*, which determines the existence of the *potential market gap* and the short and mid-term viability of the offices that make up the present and future network.

The adaptation of these variables to the technique of discriminant analysis, and to the rough sets model referred to in the following sections, necessitates the development of processes of normalization and discretization. These will in no case denaturalize the information initially present.

Section 2 reviews rough sets theory. The relationship between rough sets theory and other classification methods is presented in Section 3. The variables of the information system are discussed in Section 4. In Section 5 a discriminant analysis of the information system formed by 421 branches is analysed. In Section 6 a rough sets analysis is obtained through three steps: variable reduction, rule generation and reclassification, and cross-validation tests.

2. ROUGH SETS THEORY

2.1. Introductory Remarks

The theory of rough sets (RS) proposed by Pawlak (1991) offers a model to deal with imprecise or incomplete information. From the start, the theory has generated great interest among researchers working in the field of automatic learning and knowledge discovery from databases (data mining). In the financial field in particular, subjects such as business failure (McKee, 2000; Slowinski *et al.*, 1999; Zopounidis *et al.*, 1999) and stock market analysis (Bazan *et al.*, 1994; Grzymala-Busse, 1997; Golan, 1995) have been studied.

The RS theory is used as a method to analyse data in an information system (IS). We applied this theory to the IS formed by all the branches of a Spanish savings bank. In this analysis the main problems are those related to:

- The removal of superfluous variables, in order to obtain minimum subsets of variables (reducts) that assure a satisfactory approximation to the predetermined classification of branches by a decision variable.
- Creation of branch models for each value of the decision variable, which are most representative. These models are described by a set of decision rules (IF . . . THEN . . .) that represents the knowledge obtained from the set of all branches.

Once this analysis has been completed other problems arise, such as the application of results to the classification of new branches.

The information that makes up the IS is:

- The approximation of each branch class;
- The computation of the precision and quality of the classification of the branches;
- The search for reducts of variables;
- The removal of non-significant variables;
- Derivation of a set of rules from the reduced information system.

In practical applications, the IS often contains a mixture of discrete and continuous data. Rough sets analysis gives satisfactory results when the domains of variables are finite groups of low cardinality. This requirement is often fulfilled naturally when the variables have a qualitative (discrete) nature. However, variables that are quantitative can be handled after values have been translated into qualitative terms, for example into low, medium or high. This translation involves a division of the original domain into subintervals, and the assignment of qualitative codes to these subintervals. The definition of the *border values* of these subintervals should take into account the experience, knowledge, habits and conventions used by a financial expert. This is the approach to handling quantitative variables used in this work. It must be kept in mind, however, that the definition of the border values of the subintervals can influence the quality of the classification, so the problem of verifying the sensibility of the results to changes in the definition of border values is very important.

The results obtained using the rough sets theory, in particular, the reduction of variables and the set of rules, are of great importance. The reduction of variables can mean the elimination of superfluous variables in the valuation of branches, diminishing the cost and time of diagnosis of a branch. The set of rules shows all the important relationships between the variables, using a minimum number of decision rules and/or a minimum number of variables. So, the set of rules is easier for the user to assimilate than the original IS.

Moreover, these results represent the knowledge gathered from all the cases registered in the IS. This knowledge is extremely useful to support decisions concerning the classification of new branches. New branches are those which have not been previously dealt with and that are described only by values of condition variables (all the values or a reduced group of them). Initially, the class of each new branch is unknown. The interest of the expert is in predicting this assignment based on the knowledge that arises from past experience. The idea is based on looking for a rule whose conditions satisfy the description of a new branch. When this rule is found, it is used to support the classification. In some cases, when no rule exists, it is necessary to find the *nearest rules*. These rules are those that are close to the description of the branches classified according to a measure of distance.

2.2. Basic Concepts of the Rough Sets Theory

Information System, Indiscernibility Relation and Approximation of Sets

Let OB be a non-empty set called the universe, and let R be an equivalence relation defined in the universe OB, called an indiscernibility relation which represents a classification of the universe into categories of objects which are indiscernible or identical in terms of the knowledge provided by the given attributes. The main notion in rough sets theory is the Boolean approximation space, BoolAS, which is formally defined as:

$$A = (OB, R)$$

Equivalence classes of the relation R are also called R-elementary sets and $[o]_R$ denotes the R-elementary set containing the object $o \in OB$. Any finite union of elementary sets is called a definable set. Let us take $X \subseteq OB$, which represents a concept. X cannot always be exactly defined as the union of various elementary sets. For this reason two new sets are defined: $\underline{R}(X) = \{o \in OB \mid [o]_R \subseteq X\}$ called lower approximation and $\bar{R}(X) = \{o \in OB \mid [o]_R \cap X \neq \varnothing\}$ called upper approximation. Any set defined in terms of its lower and upper approximations is called a rough set.

The main computational effort in the processing of data in rough sets theory is associated with the determination of attribute relationships in information systems. An information system is a quadruple: $S = (OB, AT, V, f)$ where:

- OB is a finite set of objects;
- AT is a finite set of attributes;
- $V = \cup V_{a \in AT}$, V_a being the values of attribute a;
- f: $OB \times AT \rightarrow V$ is a total function such that $f(o, a) \in V_a$ for every $a \in AT$, $o \in OB$, called information function.

Let $Y = \{Y_1, Y_2, \ldots, Y_n\}$ be a classification (partition) of OB in S, and $P \subseteq AT$. Subsets $Y_j (j = 1, 2, \ldots, n)$ are classes of Y. Then, $\underline{P}(Y) = \{\underline{P}(Y_1), \underline{P}(Y_2), \ldots, \underline{P}(Y_n)\}$ and $\bar{P}(Y) = \{\bar{P}(Y_1), \bar{P}(Y_2), \ldots, \bar{P}(Y_n)\}$ are called the P-lower and the P-upper approximation of classification Y, respectively. $\underline{P}(Y)$ and $\bar{P}(Y)$ are also called the P-positive region and P-negative region of Y. $BN_P(Y) = \bar{P}(Y) - \underline{P}(Y)$ is called the P-boundary or P-doubtful region of Y.

The accuracy of approximation is defined as:

$$\alpha_P(Y) = \textstyle\sum_j card(\underline{P}(Y_j)) / \sum_j card(\bar{P}(Y_j))$$

and expresses the percentage of possible correct decisions when classifying objects employing the knowledge P. Inexactness of a decision is due to the existence of a doubtful region. The accuracy coefficient expresses how large the boundary region is. The approximation of Y by a set, P, of attributes is a measurement of vagueness and captures how accurately the set Y_j is definable in Y based on knowledge, P. Accuracy measurement is intended to capture the degree of completeness of our knowledge about the classification Y.

The coefficient

$$\gamma_P(Y) = \textstyle\sum_j card(\underline{P}(Y_j))/card(OB)$$

is called the quality of approximation of classification Y by the set of attributes P, hereafter referred to as quality of classification. If the quality measurement is 1, then the information system is deterministic; if it is less than 1, it is roughly deterministic; and if it is 0, it is totally non-deterministic. It expresses the percentage of objects that can be correctly classified into classes of Y by employing knowledge P.

Attribute Reduction and Dependency of Attributes

Core and reducts are two fundamental concepts of rough sets. A reduct is the essential part of an IS which can discern all the objects discernible by the original IS. A core is the common parts of all the reducts. The set of attributes $R \subseteq Q$ depends on the set of attributes $P \subseteq Q$ in S (denotation $P \rightarrow R$) iff $IND(P) \subseteq IND(R)$. Discovering dependencies between attributes is of primary importance in the rough sets approach to data analysis. Another important issue is that of attribute reduction, so that the reduced set of attributes provides the same quality of classification as the original set of attributes. The minimal subset $R \subseteq P \subseteq Q$ such that $\gamma_P(Y) = \gamma_R(Y)$ is called Y-reduct of P (or, simply, reduct if there is no ambiguity in the understanding of Y). Y-reduct of Q is also called minimal set (or subset) in S. An information system may have more than one Y-reduct. $\mathrm{RED}_Y(P)$ is the family of all Y-reducts of P. Intersection of all Y-reducts of P is called Y-core of P, i.e. $\mathrm{CORE}_Y(P) = \cap\mathrm{RED}_Y(P)$. The only Y-core of Q is the set of the most characteristic attributes which cannot be eliminated from S without decreasing the quality of approximation of classification Y.

Decision Rules

An information system can be seen as a decision table assuming that $AT = C \cup D$ and $C \cap D = \varnothing$. Formally, a decision table, S, is a quadruple: $S = (OB, C, D, V, f)$. All the concepts are defined in a similar manner to those of information systems; the only difference is that the set of attributes has been divided into two sets, C and D, which are conditions and decisions respectively. From the decision table, S, a set of decision rules can be derived. The D-elementary sets in S are denoted by

Y_j ($j = 1, 2, \ldots, m$) and called decision classes. Describing decision classes in terms of condition attributes from C, we get lower and upper approximations, $\underline{C}(Y_1)$, $\underline{C}(Y_2)$, ..., $\underline{C}(Y_m)$, respectively, as well as the boundaries $BN_C(Y_j) = \bar{C}(Y_j) - \underline{C}(Y_j)$, $j = 1, 2, \ldots, m$.

A decision rule can be expressed as a logical statement:

IF conjunction of elementary conditions THEN disjunction of elementary decisions

An elementary condition of the subset, $A \subseteq C$, and domain V_{a_i} of attribute $a_i \in A$ is defined as: $f(o, a_i) = v_{a_i}$ where $o \in U$, $v_{a_i} \in V_{a_i}$. We denote con_A as the conjunction of elementary conditions and $[con_A]$ as the set of all objects satisfying conjunction con_A. Obviously, if object $o \in [con_A]$, then $[con_A] = [o]_A$. We denote dec_A as the disjunction of elementary decisions and $[dec_D]$ as the set of all objects belonging either to the lower approximation of one of the decision classes (if the number of decisions is 1), or to the common boundary of s decision classes (if the number of decisions is $s > 1$).

The decision rule 'IF con_A THEN dec_D' is consistent iff $con_A \subseteq dec_D$. If the number of decisions is 1, the decision rule is exact otherwise it is approximate. Approximate rules are consequences of an approximate description of decision classes, that is, with the available knowledge one is unable to decide whether some objects (from the boundary region) belong to a given decision class or not. A class, X, is inconsistent if it contains inconsistent examples, that is, examples which have the same values in all the condition attributes as another example, but which have a different value in the decision attribute.

Procedures for the generation of decision rules from a decision table operate on inductive learning principles. The objects are considered as examples of decisions. Procedures for induction of decision rules from decision tables were presented by Grzymala-Busse (1997, 1998b) and Chan (1991), both of whom use the strategy of determining the smallest set of minimum rules that describe a concept. When a concept or category, X, is inconsistent, that is, it contains inconsistent examples, in a decision table, S, it means that X is not definable for all the condition attributes in S. To deal with inconsistent examples, the basic idea is to substitute an inconsistent concept with its lower and upper approximations generated by the set of all the condition attributes, and to induce two sets of rules, one based on the lower and one on the upper approximation. The rules derived from the lower approximation of X are called certain rules (also called exact, deterministic or discriminant) and the rules derived from the upper approximation of X are called possible rules (also called approximate, non-deterministic).

In order to learn rules from a set (whether from the set itself or from its lower or upper approximations), we have used a system for incremental learning of classification rules using examples (Fernández-Baizán *et al.*, 2000) based on the LEM3 (Chan, 1991) and LERS (Grzymala-Busse, 1992, 1997, 1998b) systems, but dealing with probabilistic approximation space (Wong and Ziarko, 1986) rather than Boolean approximation space. As described in Fernández-Baizán *et al.* (2000) probabilistic boundaries are used to determine the lower and upper approximations of a concept, which generates certain rules with a level of certainty β ($0.5 \leq \beta \leq 1$).

A probabilistic approximation space, ProbAS, is a triplet ProbAS = (OB, R, P), where OB is a universe, R is an equivalence relation on OB, and P is a probability measure on subsets of OB. In this approach, the lower and upper approximations of a subset X in OB are defined by using the concept of probabilistic approximation with a level of certainty β ($0.5 \leq \beta \leq 1$).

The R-lower and R-upper β approximations of X in ProbAS are defined as (Fernández-Baizán *et al.*, 2000; Wong and Ziarko, 1986):

$$\underline{R}X = \{o \in U \mid \mathrm{P}(X \mid [o]) \geq \beta\} \quad \text{and} \quad \bar{R}X = \{o \in U \mid \beta > \mathrm{P}(X \mid [o]) \geq 0.5\}$$

Those examples which can be said, with a level of certainty β, to belong to class X belong to the lower β-approximation of X.

In the BoolAS an example is consistent or inconsistent. In ProbAS, consistency should be verified concerning a class and a certainty level β. An example may be consistent with one class and inconsistent with another, or it may be inconsistent with several classes.

It is said that an example, o_i, is consistent with a class, X, if the conditional probability $\mathrm{P}(X \mid [o_i]) \geq \beta$. The conditional probability $\mathrm{P}(X \mid [o_i])$ is defined as $\mathrm{P}(X \cap [o_i])/\mathrm{P}([o_i])$. In the particular case of $\beta = 1$, an example, o_i, is consistent with class X if $\mathrm{P}(X \mid [o_i]) = 1$; that is to say if $[o_i] \subseteq X$. This is the condition of consistency in BoolAS.

In the upper β-approximation definition we use the parameter α instead of the constant 0.5:

$$\bar{R}X = \{o \in U \mid \mathrm{P}(X \mid [o]) \geq \alpha\}$$

and we establish $\alpha = (1 - \beta)$ to maintain the coherence with the meaning of certainty level β.

In Chan (1991) a superior limit, β, in the precision of the upper β-approximation is also considered:

$$\bar{R}X = \{o \in U \mid \beta > \mathrm{P}(X \mid [o]) \geq 0.5\}$$

We work without this superior limit (in the line of the *variable precision rough sets model* of Ziarko, 1993) to conserve the coherence with the definition of upper approximation according to the rough sets theory. The region that defines the previous expression is, in fact, the boundary region (Pawlak, 1991). The generation of possible rules from the examples of the boundary region instead of the examples of the upper approximation has a disadvantage: rules that are obtained only classify the examples of the boundary region and generalization is lost. However, it has an advantage: among the rules generated from the upper β-approximation (considered as in Pawlak, 1991) there will be no rules that have already been generated from the inferior approximation, so we avoid the later process of simplification of rules.

A value $\alpha > 0$ causes the removal of the inconsistent examples of certain upper approximations. This implies the elimination of rules with the smallest conditional probability. A value $\alpha = 0.5$ signifies that the inconsistent examples only are included in the upper approximation of the decision class that has half of the examples at least.

The inconsistent examples whose consistency with a class is greater than β are considered in the lower approximation of that class. This makes the rules generated from the lower approximation more general because they include more examples. When parameter β diminishes, the rules are compacted, and generally the number of rules diminishes.

In consequence, the stochastic approximation space implies an elimination of inconsistencies (stronger rules are generated) and a compacting of the rules set.

In our approach (Fernández-Baizán *et al.*, 2000) we have implemented the subsystem LEM2 of the system LERS (Grzymala-Busse, 1992, 1997, 1998b) as the method of rules generation. This subsystem is based on learning a minimal discriminant description for each concept. The method computes local coverings of attribute–value pairs. Local coverings are constructed from minimal complexes. The minimal complex contains attribute–value pairs, selected on the basis of their relevancy to the concept. In the case of a tie, the next criterion is the maximum of conditional probability of the concept given the attribute–value pair.

Let B be a non-empty lower or upper approximation of a concept represented by a decision–value pair (d, v). Set B depends on a set T of attribute–value pairs t if and only if:

$$\{\varnothing \neq [T] = \cap_{t \in T}[t] \subseteq B\}$$

Set T is a minimal complex of B if and only if B depends on T, and no proper subset T' of T exists, such that B depends on T'. Let $\boldsymbol{T}$ be a non-empty collection of non-empty sets of attribute–value pairs. Then $\boldsymbol{T}$ is a local covering of B if and only if the following conditions are satisfied:

1. Each member T of $\boldsymbol{T}$ is a minimal complex of B;
2. $\cup_{T \in \boldsymbol{T}}[T] = B$;
3. $\boldsymbol{T}$ is minimal, i.e., $\boldsymbol{T}$ has the smallest possible number of members.

To calculate the classification rules from a given set B (be this the set of examples relative to the given set, or its lower or upper approximation), the process is as follows:

- We begin with the attribute–value pairs relevant to the given set B. That is, the a–v pairs which only appear in the training examples of B. Next, we calculate the minimal complex, T, of B.
- The generation of rules finishes when the rules classify all the examples of set B, $\cup_{T \in \boldsymbol{T}}[T] = B$.
- Finally we minimize the total set of rules obtained, $\boldsymbol{T}$, through the elimination of superfluous rules.

In order to evaluate discovered rules, several measurements have been used.

Probability: The conditional probability that the rule defines the concept (class) that appears in the right-hand side of the rule given the conditions in the left-hand side. This probability is measured on the training set.
Specificity: The total number of attribute–value pairs on the left-hand side of the rule. The matching rules with a larger number of attribute–value pairs are more specific.
Strength: The total number of examples correctly classified by the rule during training.
TrMatch: The total number of training examples that match the left-hand side of the rule.

The problem of inducing decision rules has also been extensively investigated in the rough sets field. Nowadays, the most representative approaches and software systems (see, e.g., Orlowska, 1998; Polkowski and Skowron, 1998a,b; Slowinski, 1992; Ziarko and Yao, 2000; Bazan *et al.*, 1998) which are used in the majority of applications are:

- System LERS (Learning from Examples based on Rough Sets) introduced by Grzymala (see, e.g., Grzymala-Busse, 1992, 1997, 1998a,b) which itself has four different options of rule induction; the most popular of them seems to be the LEM2 algorithm.
- Approaches based on a discernibility matrix and boolean reasoning techniques. These concepts have been studied by Skowron and Rauszer (1992) and extended by several additional strategies connected with, e.g., the approximation of reducts, looking for dynamic reducts, boundary region thinning, data filtration and tolerance relation. Their implementations have been developed by Skowron and his collaborators and form a computational kernel of the system *Rosetta* (Ohrn *et al.*, 1998).

- RoughDAS and RoughFamily software systems developed by Slowinski and Stefanowski (1992) and collaborators which offer several rule induction options as, for example, an approach inducing the set of decision rules satisfying the given user's requirements.
- Systems Dataquest and Datalogic Szladow (1993), based on the probabilistic extension of the original rough sets model called *variable precision rough sets model* (Ziarko, 1993).
- System KDD-R (Ziarko and Shan, 1994), oriented towards data mining applications from large databases and capable of finding strongly supported rules.

Decision Making Using Decision Rules

Typically, a rule set induced from training data is used for classifications of unseen cases (sometimes called testing cases). In Grzymala-Busse (1997) two different classification schemes are used. The first one is called naive classification. If no rule exists that matches a new, unseen example or if two (or more) rules match the example but these rules indicate different concepts, the example is treated as an error. The second one is called new classification, and it is a modification of the *bucket brigade algorithm* (Gunn and Grzymala-Busse, 1994; Holland and Holyoak, 1986). With the *new* classification, the decision as to which concept an example belongs to is made using three factors: *strength*, *specificity* and *support*.

In this work, both methods of classification have been implemented, but with certain modifications that mainly improve the results in the *naive* classification. The improvement of the *naive* classification is shown in Figure 1.

In the previous section we saw that the induced rules are accompanied by four measurements: *probability*, *specificity*, *strength* and *trMatch*. The first measurement is used in the *naive* classification for the selection of the rule associated with the new example to classify. The other three measurements are used in the *new* classification to calculate a factor that will determine the class of the new example.

In the *new* classification the decision as to which concept an example belongs to is made on the basis of the support factor, which is defined by means of the *strength* and *specificity factors*:

For each example to classify, all the rules that match the example are computed. Two situations can appear:

All the rules that match the example indicate the same concept or class. In this case the example is classified as pertaining to the concept indicated by the rules.

The rules that match the example indicate different concepts. The example is classified as pertaining to the concept of that rule that has the greatest probability. In the case that there are several rules with the same probability—greater than the rest of rules—and that indicate different concepts, the concept is chosen that is associated to the rule with the greater number of attribute-value pairs. If the number of attribute-value pairs of the rules that indicate different concepts with equal probability is the same, the first concept is chosen.

If an example does not match completely with any rule, it is assigned a concept by default. This is equivalent to the application of the rule by default, that is to say, a rule with no antecedents.

Figure 1. Improvement of the *naive* classification

$$Support = \sum_{\text{matching rules } R \text{ describing } C} Strength(R) * Specificity(R)$$

Support is the sum of the votes of those rules that match the example and indicate the same concept.

In the *new* classification a new example is not classified by choosing the best rule that matches the example, but according to the global knowledge that stores the set of rules that match the example.

In the *naive* classification when no rule matches completely with the example the rule by default is applied. In the *new* classification if complete matching is not possible, all the rules that match partially (rules that match at least one a–v pair of the example) are identified.

For each rule, R, that matches partially, the matching factor, *trMatch*(R), is calculated. If *support* is null for all the concepts, then the following expression for each concept is calculated:

$$\sum_{\text{partially matching rules } R \text{ describing } C} TrMatch(R) * Strength(R) * Specificity(R)$$

As in the previous case, the concept C assigned to the new example to classify is the one that has the greatest value for the previous expression.

With this classification system it is not necessary to generate a class by default at the moment of the rules generation.

3. RELATIONSHIP BETWEEN THE ROUGH SETS METHOD AND OTHER CLASSIFICATION METHODS

One result of the use of the rough sets method in data analysis is a set of classification rules in two or more classes. The rules form a description of each decision class in terms of an expression that combines, in the case of analysing a financial entity, simple and observable properties of the branch offices. For example, the expression:

PROFITS = 7 **AND** HALF BALANCE OF RESOURCES = 1
AND HALF BALANCE OF INVESTMENTS = 1

could be the description of the branch class TYPOLOGY1996 (G-type). One could classify new branches by making their characteristics conform to the conditions of the rules.

Some objectives of the rough sets approach are to:

- Distinguish between several classes;
- Capture the essential factors that affect the result of the classification, and not consider irrelevant factors;
- Minimize the number of rules and their conditions;
- Receive strong support from the available data;
- Have a low rate of error for new cases as a result of their generalization.

The previous group of objectives has some overlap with the objectives of statistical approaches to the classification problem. However, the formation of characteristic and discriminant descriptions of a class is not an objective in the statistical techniques. The main objective of the statistical approach is the construction of a probabilistic classifier that approaches the theoretically optimal classification of Bayes. Different methods for the construction of such classifiers have been used (Chen, 1973; Duda and Hart, 1973; Lavalle, 1970; Meisel, 1972). These usually involve very strong suppositions, such as the Gaussian distribution of values in parametric techniques (Duda and Hart, 1973), or require a very large number of examples, as in the case of the estimation method of probability density in non-parametric approaches. Such non-parametric techniques could make the supposition of probabilistic independence of the variables that define the branches in order to estimate the distributions of conditional probability highly questionable. The method of the *nearest neighbour* is based on the (not always correct) supposition that if two branches are very near in terms of any measurement of distance, then they belong to the same class.

Experience indicates that the techniques of the nearest neighbour do not offer satisfactory performance and, in addition, essential information is lost in the process of reducing the problem of classification to the calculation of distances between vectors. Linear discriminant functions are not appropriate when categories are not linearly separable (Duda and Hart, 1973; Nilsson, 1990; Weiss and Kulikowski, 1991). Neither non-linear discriminant functions on multi-layer perceptions suffer from this shortcoming. These, however, require tremendous power of calculation and the final result is strongly dependent on the establishment of the group of initial parameters. Also, all of these approaches suffer from problems of dimensionality. They are not able to identify the essential subset of non-redundant variables. While the efficiency of the systems based on RS improves with the extension of additional outstanding characteristics, the efficiency of statistical techniques may degrade when new characteristics are added (Duda and Hart, 1973).

Zadeh (1977) and others have tried the use of fuzzy sets in the classification of approximate patterns. In this approach the branches are represented in terms of variables whose values are of constant linguistic imprecision, such as low and half. These constants are represented as functions of diffuse membership. The use of diffuse membership functions allows the degree of association of a branch to be expressed with an imprecisely defined linguistic notion. The resulting classification, the target category, is also assumed to be a diffuse rather than a precise category. The central problem in the fuzzy approach is the derivation of a formula that ties the membership functions of an unknown objective category with the given membership functions for the variables of a branch. Once such a formula has been obtained, for example, from some cases of *training*, degrees of membership with regard to the target category can be predicted in new cases. The formula could be perceived as a recognition rule corresponding to the logic rules that represent the descriptions of the target category in the RS theory. The main difficulty in the fuzzy sets method is the lack of objective techniques in order to define characteristics, target categories and membership functions.

As a basic method in machine learning, learning from examples has been studied extensively (*Version Space*, AQ systems, ID systems, CN2 algorithm). In most learning systems, the training examples are classified in advance by the tutor into two disjoint sets, the positive examples and the negative examples. The training examples represent low-level, specific information. The learning task is to generalize these low-level concepts to general rules.

Version Space (Mitchell, 1977) assumes we are trying to learn some unknown target concept defined on the instance space. We are given a sequence of positive and negative examples which are called samples of the target concept. The task is to produce a concept that is consistent with the samples. The set of all hypotheses that are consistent with the sample is called the version space of the sample. The version space is empty in the case that no hypothesis is consistent with the sample. The algorithm of this learning task, the *Candidate-Elimination* algorithm, maintains two subsets of the version space: the set, S, of the most specific hypotheses in the version space and the set, G, of the most general hypotheses. These sets are updated with each new example. The learning process terminates when $G = S$.

AQ systems (Michalski *et al.*, 1986) are designed to find the most general rule in the rule space that discriminates training examples in all other classes. To search this rule space, AQ systems use the AQ algorithm, which is close to the repeated application of the *Candidate-Elimination* algorithm.

ID3 (Quinlan, 1983) and CN2 (Clark and Niblett, 1989) are based on information theory (Clark and Niblett, 1989) searches for classification rules directly using a measure of rule goodness. These algorithms use an information-theoretic approach aimed at minimizing the expected number of tests to classify the objects. The attribute selection part is based on the plausible assumption that the complexity of the decision tree is strongly related to the amount of information conveyed by this message. It builds a decision tree by choosing a good test attribute that partitions the instance into smaller sets for which decision subtrees are constructed recursively. To determine which attribute should be the test attribute for a node, the algorithm applies an information-theoretic measure *gain*. An attribute with the maximal gain is selected as the test attribute.

The main advantage of the rough sets theory is that it does not need any preliminary or additional information about the data (unlike the probability theory which requires knowledge of probabilities, or the fuzzy sets theory which requires knowledge of the degree of membership) and another advantage (versus machine learning algorithms) is that it is directly equipped to handle inconsistent or seemingly conflicting examples in the data material. In the rough sets approach inconsistencies are not corrected or aggregated. Instead the lower and upper approximations of all decision concepts are computed. Thus, the task of rule induction from inconsistent data is reduced to rule induction from consistent data, since both lower and upper approximations of the concept make it feasible. As a consequence of using the approximations, induced decision rules are categorized into certain (exact) and approximate (possible) ones depending on the lower and upper approximations used, respectively.

4. THE INFORMATION SYSTEM

4.1. The Data

In a first approach to the problem, an analysis was carried out using the following socio-economic variables provided by the savings bank for each one of its operative branches up to December 31, 1996:

- RANKING 96
- REGIONAL
- AREA
- TYPOLOGY 95
- TYPOLOGY 96
- WORK LOAD
- PROFIT
- AVERAGE BALANCE OF RESOURCES
- AVERAGE BALANCE OF INVESTMENTS

- VOLUME OF BUSINESS
- PROFITABILITY THRESHOLD

RANKING 96—Ordered classification of the different branches that make up the commercial network of the bank, according to the valuation assigned as a function of work load, resources, investments and profit.
REGIONAL—The geographical environment in which the bank operates is divided into five large areas called *regional sectors*. The commercial administration of the bank is carried out jointly between the sectors while allowing for individualized policies to suit the idiosyncrasies of each region.
AREA—AREA involves a subdivision of the regional sectors. The object is to rationally group those branches that, while they belong to the same regional sector, present subtle individual characteristics as a result of their geographic location and their nature (rural or urban).
TYPOLOGY 9X—Stratification of the commercial network of the savings bank according to the valuations assigned to branches up to December 31, 199X. Seven STRATA or TYPOLOGIES (A, B, C, D, E, F and G) have been configured in order to partition the population or universe.
WORK LOAD—Hours imputed to each branch as a function of the operations carried out, according to the measurement model used. External consultants have confirmed this.
PROFIT—Results (positive or negative) obtained by the branch up to December 31, 1996, registered in the account of losses and earnings of the branch, according to accepted accounting processes.
AVERAGE BALANCE OF RESOURCES—Average balance of liabilities and other funds deposited by clients registered up to December 31, 1996.
AVERAGE BALANCE OF INVESTMENTS—Average balance of the asset operations managed by the branch up to December 31, 1996.
VOLUME OF BUSINESS—Sum of the AVERAGE BALANCES of RESOURCES and INVESTMENTS to December 31, 1996.
PROFITABILITY THRESHOLD—Minimum VOLUME OF BUSINESS needed by a branch in order to generate earnings sufficient to cover total costs.

These variables have been constructed independently of any external information regarding the positioning variables of the branches, given that this information is not taken into account in the computations of TYPOLOGY 96 and RANKING carried out by the entity. The positioning variables are, however, included in the second phase of the project (currently being developed), given their relevant nature and importance in the decision-making process.

An expert was asked to discretize the continuous financial variables in the primary analysis of the groups (we cannot provide the description of the discretization of continuous variables made by means of experts' opinions, because this a private and confidential information of the bank). Discretization is the process by which a group of values is contained together in a symbol range. Discretization is performed because the precision of financial variables is rather doubtful and, moreover, it prevents drawing general conclusions from data in terms of reducts and decision rules.

By means of discriminant analysis (Section 5), the validity of the groups previously defined by the entity has been statistically contrasted using the variable TYPOLOGY 96. This has determined the optimal discriminant function.

On the other hand, rough sets analysis (Section 6) provides a method for the reduction of knowledge and the construction of decision rules corresponding to the observed variables. It is configured as a suitable alternative tool in the valuation processes.

4.2. Frequency of the Observed Variables

- **REGIONAL**
  The administration of the commercial network all over Spain is carried out through a pyramidal structure in which the branches, areas and regions are inserted hierarchically. The distribution of regions is presented in Table I.
- **AREA**
  The distribution of the branches observed in the 17 existent areas, up to December 31, 1996, is shown in Table II.
- **TYPOLOGY**
  If we observe the TYPOLOGY 95 and TYPOLOGY 96 variables, a certain improvement in the positioning of the branches within the network can be seen, since the percentage of lowest category branches (TYPOLOGY = G) has fallen from 54.6% in 1995 to 49.2% in 1996. This reduction in the lowest stratum corresponds to the increments in the ones above it. The situation is described in Table III.

Table I. Frequency of REGIONAL variable

Regions	Absolute frequency	Relative frequency	Accumulated frequency (%)
Region I	116	27.55%	27.55
Region II	55	13.06%	40.61
Region III	74	17.58%	58.19
Region IV	109	25.89%	84.08
Region V	67	15.91%	100.00
TOTAL	421	100.00%	

Table II. Frequency of AREA variable

Area code	Absolute frequency	Relative frequency	Accumulated frequency (%)
821	30	7.13%	7.13
823	35	8.31%	15.44
825	38	9.03%	24.47
829	6	1.43%	25.90
834	10	2.38%	28.28
835	10	2.38%	30.66
855	18	4.28%	34.94
856	19	4.51%	39.45
857	20	4.75%	44.20
858	3	0.71%	44.91
957	25	5.94%	50.85
958	30	7.13%	57.98
961	27	6.41%	64.39
962	31	7.36%	71.75
963	51	12.11%	83.86
968	37	8.79%	92.65
969	31	7.36%	100.00
TOTAL	421	100.00%	

Table III. Frequency of TYPOLOGY variable

Stratum	Typology	1995	1996	% VAR
1	A	1	6	6.00
2	B	6	4	0.67
3	C	5	9	1.80
4	D	12	37	3.08
5	E	67	69	1.03
6	F	100	89	0.89
7	G	230	207	0.90
TOTAL		421	421	1.00

It is necessary to keep in mind the lack of uniformity in the financial years 1995 and 1996, since the determination of the TYPOLOGY 95 variable is conditioned only for the AVERAGE BALANCES of RESOURCES and INVESTMENTS, in α and β, respectively. TYPOLOGY 96, on the other hand, includes the combined evaluation of the following X_i variables, in α_i:

(a) X_1 = Average balance of resources (α_1)
(b) X_2 = Average balance of investments (α_2)
(c) X_3 = Work load (α_3)
(d) X_4 = Results account by branch (α_4)

The valuation assigned to each branch comes from:

$$P_j = \sum_{i=1}^{4} \left[\left(\ln X_{ij} * P * \alpha_i \% \right) \div \left(\sum_{j=1}^{421} \ln X_{ij} \right) \right]$$

where:

P_j = Valuation relative to branch j
ln = Function of assignment used by the bank
X_{ij} = Value of the variable X_i corresponding to branch j
P = Total points to distribute (previously established by the entity)
$\alpha_i\%$ = Applied weight to the variable X_i, for all j

The total group of branches integrating the commercial network is divided into STRATA or TYPOLOGIES (A, B, C, D, E, F and G) according to a scale of valuation defined by the entity for each financial year. Therefore, the TYPOLOGY 96 variable observed up to December 31, 1996 is solely conditioned by the valuation obtained according to the previous formulation, independently of the TYPOLOGY assigned in previous financial years.

5. DISCRIMINANT ANALYSIS

Our objective is to determine those factors that determine efficiency and the appropriate dimensioning strategy for the bank, more concretely, to characterize the branches of the entity based on their nature, typology and degree of specialization; that is to say, according to their stratum.

We have considered a technique of multivariate analysis: discriminant analysis. The premises used were as follows:

- **Target population**: Commercial network of the financial entity subject to study.
- **Unit or individual**: Branch (real or potential).
- **Variables**: Those previously mentioned, unless otherwise expressed.

Discriminant analysis is used for the classification of the population into predefined groups. So, a given individual (a branch) is assigned to a predetermined stratum of the universe as a function of its characteristics. The fundamental aims of discriminant analysis consist of determining whether the variables observed allow the branches to be sorted into defined groups, and discovering the optimum discriminatory function on which future predictions will be based.

In the framework of this study, the predefined groups correspond to the values of the TYPOLOGY 96 variable (A, B, C, D, E, F and G).

Having analysed the nature of each variable, the following variables were found inappropriate or unnecessary for discriminant analysis:

1. The RANKING variable. This variable is a finer measurement than TYPOLOGY, but the TYPOLOGY 96 variable summarizes the information contained in RANKING.
2. The REGIONAL and AREA variables have been excluded in the analysis. They will, however, be used in the second phase of the study, when the external information relative to the positioning variables of the branches that configure the commercial network will be introduced.
3. TYPOLOGY 95. The lack of uniformity from one financial year to the next, together with the determination of the TYPOLOGY of the branches, led us to discard the variable TYPOLOGY 95.
4. THRESHOLD PROFITABILITY was excluded from the calculation, because once the nature (rural or urban) of a branch is fixed, the VOLUME OF BUSINESS determines the THRESHOLD PROFITABILITY.
5. VOLUME OF BUSINESS. This variable is no more than a linear combination of the RESOURCES and INVESTMENTS variables.

Therefore, the developed discriminant analysis includes five variables, one that describes predefined groups (TYPOLOGY 1996), and four independent variables or predictors (PROFIT, WORK LOAD, INVESTMENTS, RESOURCES). Table IV shows the results obtained by means of discriminant analysis. In spite of the limitations observed in connection with the parametric requirements (normality and equality in the covariance matrix for the different groups), we feel that the results are interesting for the main purpose of our study, validation of the model.

The percentage of cases correctly grouped according to the discriminant analysis is 78.62%, although we should observe that this global number is not representative of the situation observed at group level. While predetermined groups A and C support the prediction 100%, others such as B, D, E, F and G present irregular behaviour. For this reason, the distribution of the TYPOLOGY 96 variable, modified in accordance with the prediction obtained in the discriminant analysis, differs significantly from the predefined groups as seen in Table V.

The four standardized canonical discriminant functions obtained by the statistical procedure are described below:

Table IV. Results of the discriminant analysis

'Predetermined'		Prediction						
Group	Cases	GROUP A	GROUP B	GROUP C	GROUP D	GROUP E	GROUP F	GROUP G
A	6	6 100.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
B	4	0 0.00%	2 50.00%	2 50.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
C	9	0 0.00%	0 0.00%	9 100.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
D	37	0 0.00%	0 0.00%	0 0.00%	31 83.80%	6 16.20%	0 0.00%	0 0.00%
E	69	0 0.00%	0 0.00%	0 0.00%	1 1.40%	53 76.8%	14 20.30%	1 1.40%
F	89	0 0.00%	0 0.00%	0 0.00%	1 1.10%	7 7.90%	37 41.60%	44 49.40%
G	207	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	14 6.80%	193 93.20%
TOTAL	421	6 1.42%	2 0.48%	11 2.61%	33 7.84%	66 15.68%	65 15.44%	238 56.53%

Table V. Classification accuracy to group level

TYPOLOGY 1996	Cases 'predetermined'	Cases predicted	VAR %
A	6 1.42%	6 1.42%	0 0.00%
B	4 0.95%	2 0.48%	2 0.48%
C	9 2.14%	11 2.61%	2 0.48%
D	37 8.79%	33 7.84%	4 0.95%
E	69 16.39%	66 15.68%	3 0.71%
F	89 21.14%	65 15.44%	24 5.70%
G	207 49.17%	238 56.53%	31 7.36%
TOTAL	421 100.00%	421 100.00%	0 0.00%

$$F_1(X_1, X_2, X_3, X_4) = 0.51054\,X_1 + 0.48050\,X_2 + 0.70189\,X_3 + 0.54129\,X_4$$
$$F_2(X_1, X_2, X_3, X_4) = -0.11338\,X_1 - 0.87207\,X_2 + 0.71744\,X_3 + 0.42539\,X_4$$
$$F_3(X_1, X_2, X_3, X_4) = 0.86822\,X_1 + 0.19123\,X_2 - 0.09603\,X_3 - 0.65024\,X_4$$
$$F_4(X_1, X_2, X_3, X_4) = 0.25277\,X_1 - 0.65948\,X_2 - 0.38055\,X_3 + 0.79954\,X_4$$

where:

X_1 = PROFITS
X_2 = WORK LOAD
X_3 = INVESTMENTS
X_4 = RESOURCES

Function F_1 describes 97.45% of the parametric model (followed by F_2, F_3 and F_4, whose contributions are 2.35%, 0.12% and 0.07%, respectively). This percentage confers on F_1 a significant role in the predictive analysis of the problem. (Table IV shows a lack of precision of 21.38%, this indiscernibility is focused on the subtable: BDEFG × BDEFG.)

6. DATA ANALYSIS BASED ON ROUGH SETS

6.1. Basic Characteristics

We have carried out the main steps of the rough sets approach:

- Approximation of categories or concepts or classes;
- Computation of the approximation accuracy;
- Computation of the classification quality;
- Search for reducts of variables;
- Reduction of the information system;
- Derivation of a set of rules from the reduced IS.

The system accepts data from the IS in table form, the lines refer to branches and the columns to the condition and decision variables of the branches.

The system may contain both qualitative and quantitative entry data, each of which is coded using different norms. For this reason two types of information systems are considered: original information system (OIS) and coded information system (CIS). The OIS contains qualitative and quantitative variables that take the values of their original domains. The CIS contains variables with coded values.

Variable values are coded in two different ways depending on whether they are qualitative or quantitative. For the former, the values are simply coded with natural numbers. For the latter, the original values are substituted by codes that coincide with the number of the subinterval containing the value, defined by the expert.

In this section a rough sets analysis is obtained through three steps: variables reduction (Section 6.2), rules generation based on an extension (Fernández-Baizán *et al.*, 2000) of algorithm LEM3 (Chan, 1991) and LEM2 (Grzymala-Busse, 1992, 1997, 1998b) (Section 6.3), and reclassification/cross-validation tests (Section 6.4).

6.2. Reduction of Variables

We have considered classification according to the variable TYPOLOGY 96 in seven classes (A, B, C, D, E, F and G).

Using the complete group of condition variables in order to characterize the classes of TYPOLOGY 96, and $\alpha = 0$, $\beta = 1$, we obtain a quality of classification of 92.6% which means that 7.4% of the branches cannot be correctly classified into a class of TYPOLOGY 96 by employing all the available knowledge. We concluded that 92.6% of the branches are classified correctly in the seven groups of TYPOLOGY 96 according to the values taken by the complete group of condition variables.

Using the complete group of condition variables in order to characterize the classes of TYPOLOGY 96 and $\alpha = 0.5$, $\beta = 0.5$, we obtain a quality of classification of 100% which means that the branches can be correctly classified into a class of TYPOLOGY 96 by employing all the available knowledge (the complete group of condition variables).

The next step is to verify if the group of condition variables, C, is dependent, that is, if C contains any dispensable variable which can be eliminated from C without decreasing the quality of classification induce by TYPOLOGY 96. Variables, c, belonging to C were progressively eliminated. In this way, we found that the combined C formed by the eight condition variables is dependent and has the following reduct formed by seven variables:

- REGIONAL
- TYPOLOGY 1995
- WORK LOAD
- PROFIT
- AVERAGE BALANCE OF RESOURCES
- AVERAGE BALANCE OF INVESTMENTS
- VOLUME OF BUSINESS

Next, variables of the reduct were progressively eliminated and any decrease in the precision of category definition and classification quality observed. The goals are:

- To find a small set of condition variables that is both relevant and able to describe the predetermined classification induced by the variable TYPOLOGY 96.
- To try to use the same variables as those used by the discriminant analysis approach.

Some of the results of this process of reduction are shown in Table VI.

The reduced set of variables selected for the rough sets analysis is the fifth reduction from Table VI (that which has a quality of approximation of classification of 99% for $\alpha = \beta = 0.5$). This set contains the same variables as those used for the discriminant analysis. This permits us to compare the results of the two methods of analysis.

The reduced information system is shown in Table VII.

6.3. Rules Generation

The rough sets analysis of the coded information table has been performed using the system presented in Fernández-Baizán *et al.* (2000): an incremental learning system of production rules from examples based on probabilistic approximation space. In this work, *probabilistic limits* in the precision of the upper and lower approximations of a concept are used. These allow the generation of *certain rules* with a certainty level β ($0.5 \le \beta \le 1$). We also use two classification methods: *naive* and *new*.

The set of rules generated from the training set in Figure 2 is presented in the Appendix. The calculus of Figure 2 corresponds to the values: $\alpha = 0.5$ and $\beta = 0.5$, which obtain the best accuracy and quality results.

Table VI. Sensitivity of the quality of classification in the reduction process

Sets of condition variables	Eliminated variables	Accuracy of approximation quality of approximation of classification *induced by condition variables set with respect to that induced by TYPOLOGY 96*, $\alpha = 0$, $\beta = 1$	Accuracy of approximation quality of approximation of classification *induced by condition variables set with respect to that induced by TYPOLOGY 96*, $\alpha = 0.5$, $\beta = 0.5$
REGIONAL TYPOLOGY 95 WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS TRADE VOLUME PROFITABILITY THRESHOLD		86.2%/92.6%	100%/100%
REGIONAL TYPOLOGY 95 WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS TRADE VOLUME	PROFITABILITY THRESHOLD	86.2%/92.6%	100%/100%
REGIONAL WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS TRADE VOLUME	TYPOLOGY 95 PROFITABILITY THRESHOLD	85%/91.9%	100%/100%
WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS TRADE VOLUME	REGIONAL TYPOLOGY 95 PROFITABILITY THRESHOLD	71.8%/83.6%	100%/100%
WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS	REGIONAL TYPOLOGY 95 PROFITABILITY THRESHOLD TRADE VOLUME	63%/77.6%	100%/99%

6.4. Classification of Branches: Reclassification and Cross-validation Tests

Reclassification Test Using Various Sets of Decisions Rules

The reclassification tests (i.e. tests in which the same set of examples is used to generate the set of rules and then to test it) for the reduced information system in Table VII is shown in Table VIII.

Table VII. Reduced information system

Conditions	Decision
WORK LOAD PROFIT AVERAGE BALANCE OF RESOURCES AVERAGE BALANCE OF INVESTMENTS	TYPOLOGY 96

Condition attributes: Work load, profit, average balance of resources, average balance of investments
Decision attribute: Typology 96
Number of training branches: 421
Unknown values: none
Attributes number: 5
Decision classes number: 7
α: 0.5
β: 0.5
Accuracy of approximation: 100%
Quality of approximation of classification: 99%

Figure 2. Information about the training set

Table VIII. Results of reclassification varying factor β

β	Percentage of successes (naive classification)	Percentage of successes (new classification)	Number of rules
1	93.35%	92.16%	115
0.75	93.34%	93.34%	106
0.5	93.34%	93.11%	93

Cross-validation Tests

If the number of cases in the data is greater than or equal to 100 and less than 1000, the *10-fold cross-validation* is recommended (Hamburg, 1983; Holland and Holyoak, 1986). In this technique, all cases are randomly reordered, and then a set of all examples is divided into 10 mutually disjoint subsets of approximately equal size. For each subset, all remaining examples are used for training, i.e. for rule induction, while the subset is used for testing. The error rate is computed as:

$$error\ rate = \frac{total\ number\ of\ misclassifications}{number\ of\ examples}$$

Ten-fold cross-validation is commonly accepted as a standard method of validation for rule sets. However, using this method twice, with different preliminary random reordering of all cases, yields—in general—two different estimates for the error rate. So it appears that cross-validation is not a sufficiently reliable tool to estimate the quality of the rule induction systems. Repeating the method a number of times and finding the average rate of error could overcome this difficulty. This is called *N-fold cross-validation with k repetitions*.

The results for the training set in Figure 2 are presented in Table IX. The results of the 10-fold cross-validation with 40 repetitions for the *naive* and *new* methods respectively are illustrated in Figures 3 and 4.

Table IX. Results of cross-validation

	Percentage of successes (naive classification)	Percentage of successes (new classification)
10-fold	76.96%	79.34%
10-fold with 40 repetitions	77.28%	79.86%

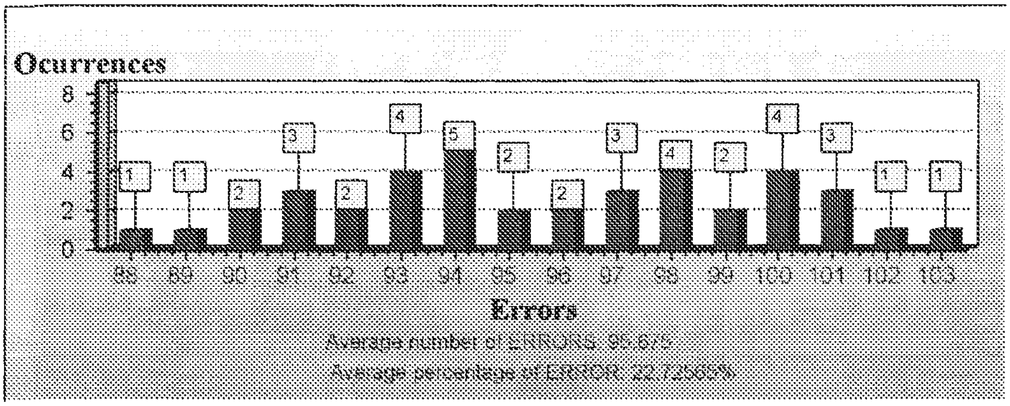

Figure 3. Ten-fold cross-validation with 40 repetitions (naive classification)

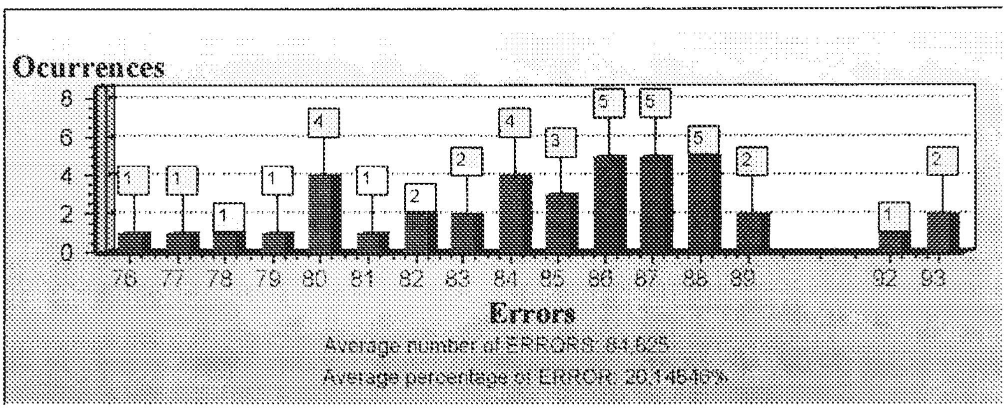

Figure 4. Ten-fold cross-validation with 40 repetitions (new classification)

7. LIMITATIONS OF THIS RESEARCH

Discriminant analysis is a well-known statistical classification method which allows the researcher to study differences between groups of objects with respect to several variables simultaneously. Discriminant analysis requires two important assumptions: (a) that each group is drawn from a population which has a multivariate normal distribution, (b) that the population covariance matrices are equal for each group. If these assumptions are violated, could this account for the difference in classification accuracy? Further studies may look into relaxing both assumptions.

8. CONCLUSIONS AND FUTURE WORK

The ability to differentiate categories of the variable TYPOLOGY 96 provided by the rough sets method has been compared with that of the discriminant analysis method. The comparison has been carried out using the reclassificatory power of a group of variables previously reduced by the rough sets theory to the set:

- WORK LOAD
- PROFIT
- AVERAGE BALANCE OF RESOURCES
- AVERAGE BALANCE OF INVESTMENTS

The results of the comparison are summarized in Table X.

The favourable results obtained in this work encourage us to propose the method of rough sets to model the valuation of branch offices. It can be observed (Figure 4) that the cross-validation of a rough sets approach with the *new* method (average percentage of successes 79.86%) even improves the reclassification results of discriminant analysis (78.6%).

It is clear that the amount and quality of the data (as well as discretization methods) are crucial in the generation of classification rules using the rough sets method. Discretization, we recall, is the

Table X. Comparison of results

Condition variables:
WORK LOAD
PROFIT
AVERAGE BALANCE OF RESOURCES
AVERAGE BALANCE OF INVESTMENTS
Decision variable:
TYPOLOGY 96

Method	Reclassification test
DISCRIMINANT ANALYSIS	78.6%
ROUGH SETS	93.34% *(Using the training set in Figure 2, the set of rules in the Appendix and the naive classification)*

process by which a group of values is contained in a range. Part of our future work will be generating the ranges automatically and contrasting them with those facilitated by an expert. The rough sets community has been committed to constructing efficient discretization algorithms. Rough sets methods combined with boolean reasoning (Brown, 1990) lead to several successful approaches (Grzymala-Busse and Stefanowski, 2001; Komorowski *et al.*, 1999; Nguyen and Skowron, 1995; Slowinski and Vanderpooten, 2000).

Better methods of discretization are needed especially the indicators that have to be discretized very finely. An index of ranges for some variables that could lead to more interesting results is:

+2	Jump up
+1	Tendency up
0	Without change
−1	Tendency down
−2	Jump down

In conclusion, these initial results need more work before the research that we have begun can be applied to processes of decision making in commercial applications. These first steps are promising.

APPENDIX

The set of certain rules obtained from the training set defined in Figure 2 is the following:

ACCURACY OF THE APPROXIMATION: 100%
QUALITY OF THE APPROXIMATION OF THE CLASSIFICATION: 99%
$\alpha = 0.5$, $\beta = 0.5$

WORK LOAD	PROFIT	RESOURCES	INVESTMENTS	TYPOLOGY 96	*Probability*	*Specificity*	*Strength*	*TrMatch*
11	–	–	–	A	1	1	6	6
–	11	–	10	B	1	2	2	2
–	11	9	–	B	1	2	2	2
10	10	–	–	C	1	2	3	3
–	–	7	8	C	1	2	1	1
9	–	–	7	C	1	2	1	1
–	–	7	3	C	1	2	2	2
–	–	7	4	C	1	2	2	2
–	–	–	6	D	1	1	1	1
–	–	–	5	D	1	1	1	1
7	10	–	–	D	1	2	5	5
–	10	5	–	D	1	2	4	4
–	–	7	1	D	1	2	4	4
5	10	–	–	D	1	2	1	1
–	10	–	2	D	0.86	2	6	7
9	–	–	2	D	0.86	2	6	7
9	–	–	1	D	1	2	5	5
–	–	5	2	D	1	2	6	6
8	9	5	–	D	0.6	3	3	5
8	9	6	–	D	1	3	1	1
7	9	6	–	D	1	3	1	1
8	–	4	2	D	1	3	4	4

APPENDIX (*Continued*)

WORK LOAD	PROFIT	RESOURCES	INVESTMENTS	TYPOLOGY 96	*Probability*	*Specificity*	*Strength*	*TrMatch*
8	–	–	3	D	1	2	2	2
–	12	–	2	D	1	2	1	1
7	–	4	1	E	1	3	11	11
8	–	4	1	E	1	3	4	4
7	9	4	–	E	1	3	6	6
6	–	4	–	E	1	2	1	1
–	7	4	–	E	1	2	3	3
5	9	4	–	E	0.75	3	3	4
7	9	3	–	E	1	3	6	6
7	9	5	–	E	1	3	2	2
7	9	2	1	E	1	4	1	1
–	8	6	–	E	1	2	2	2
7	7	3	–	E	1	3	1	1
7	8	2	–	E	1	3	1	1
7	8	3	–	E	0.78	3	7	9
6	9	–	–	E	1	2	6	6
–	9	–	4	E	1	2	2	2
–	9	2	3	E	1	3	1	1
–	9	2	2	E	1	3	1	1
5	9	3	–	E	0.6	3	3	5
–	7	3	2	E	1	3	1	1
8	8	–	–	E	1	2	4	4
6	8	–	2	E	1	3	2	2
5	–	5	–	E	1	2	1	1
–	8	2	2	E	1	3	4	4
–	11	–	1	E	1	2	1	1
5	8	2	–	F	1	3	15	15
4	8	1	–	F	1	3	4	4
5	8	1	–	F	1	3	2	2
5	8	4	–	F	1	3	1	1
6	8	2	1	F	1	4	2	2
4	8	–	1	F	0.90	3	319	21
3	–	3	–	F	0.75	2	3	4
6	8	3	1	F	0.6	4	3	5
–	1	–	7	F	1	2	1	1
5	6	–	2	F	1	3	1	1
5	9	2	1	F	1	4	1	1
5	7	–	–	F	0.63	2	5	8
4	9	–	–	F	1	2	2	2
4	7	–	2	F	1	3	2	2
–	1	3	–	F	1	2	2	2
6	7	2	–	F	1	3	2	2
–	5	–	3	F	1	2	1	1
6	6	–	1	F	1	3	1	1
7	7	2	–	F	1	3	1	1
–	6	3	–	F	1	2	2	2
6	6	1	–	F	1	3	1	1
3	10	–	–	F	1	2	1	1
–	7	5	–	F	1	2	1	1
8	6	–	–	F	0.50	2	1	2
–	–	6	2	F	1	2	1	1
–	3	–	–	G	1	1	17	17
–	2	–	–	G	1	1	2	2
2	–	–	–	G	1	1	43	43
–	5	–	1	G	1	2	33	33
–	5	–	2	G	1	2	3	3

APPENDIX (*Continued*)

WORK LOAD	PROFIT	RESOURCES	INVESTMENTS	TYPOLOGY 96	*Probability*	*Specificity*	*Strength*	*TrMatch*
–	4	–	1	G	1	2	25	25
5	4	–	–	G	1	2	8	8
7	4	–	–	G	1	2	1	1
4	4	–	–	G	1	2	10	10
3	7	–	–	G	1	2	30	30
3	–	2	1	G	0.94	3	29	31
3	8	1	–	G	0.50	3	2	4
–	6	1	1	G	1	3	22	22
4	7	–	1	G	0.G3	3	7	11
4	6	–	–	G	1	2	G	G
G	6	2	2	G	1	4	1	1
7	6	2	–	G	1	3	1	1
5	6	–	1	G	0.83	3	5	G
–	7	3	1	G	1	3	1	1
4	1	–	–	G	1	2	4	4

RULE BY DEFAULT: G
TOTAL NUMBER OF RULES (without including the rule by default): 93

REFERENCES

Bazan JG, Skowron A, Synak P. 1994. Market data analysis: a rough set approach. ICS Report C194, Warsaw University of Technology, Warsaw.

Bazan JG, Nguyen HG, Nguyen TT, Skowon A, Stepaniuk J. 1998. Synthesis of decision rules for object classification. In *Incomplete Information: Rough Set Analysis*, Orlowska E (ed.). Physica-Verlag: Berlin; 23–57.

Brown FM. 1990. *Boolean Reasoning*. Kluwer Academic: Dordrecht.

Chan CC. 1991. Incremental learning of production rules from examples under uncertainty: a rough set-approach. *International Journal of Software Engineering and Knowledge Engineering* **1**(4): 439–461.

Chen C. 1973. *Stanstical Pattern Recognition*. Haiden: New York.

Clark P, Niblett T. 1989. The CN2 induction algorithm. *Machine Learning Journal* **3**: 261–283.

Duda RO, Hart PE. 1973. *Pattern Classification and Scene Recognition*. Wiley: New York.

Fernández-Baizán MC, Pérez-Llera C, Feito-García J. 2000. LEM3 algorithm generalization based on probabilistic approximation space. In *Rough Sets and Current Trends in Computing*, Lecture Notes in Computer Science Vol. 2005, Ziarko W, Yao Y (eds). Springer-Verlag: Berlin.

Golan R, Ziarko W. 1995. A methodology for stock market analysis utilizing rough sets theory. In *Proc. IEEE/IAFE 1995 Conference on Computational Intelligence in Financial Engineering*, New York; 32–40.

Grzymala-Busse JW. 1992. LERS—a system for learning from examples based on rough sets. In *Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory*, Slowinski R (ed.). Kluwer Academic: Dordrecht; 3–18.

Grzymala-Busse JW. 1997. A new version of the rule induction system LERS. *Fundamenta Informaticae* **31**: 27–39.

Grzymala-Busse JW. 1998a. Applications of the rule induction system LERS. In *Rough Sets in Knowledge Discovery: Methodology and Applications*, Polkowski L, Skowron A (eds). Physica-Verlag: Berlin; 366–375.

Grzymala-Busse JW. 1998b. Knowledge acquisition under uncertainty: a rough sets approach. *Journal of Intelligent Robotic Systems* **1**: 3–16.

Grzymala-Busse JW, Stefanowski J. 2001. Three approaches to numerical attribute discretization for rule induction. *International Journal of Intelligent Systems* **16**(1): 29–38.

Gunn JD, Grzymala-Busse JW. 1994. Global temperature stability by rule induction: an interdisciplinary bridge. *Human Ecology* **22**: 59–81.

Hamburg M. 1983. *Statistical Analysis for Decision Making*, 3rd edn. Harcourt Brace Jovanovich: New York.

Holland JH, Holyoak RE. 1986. *Induction, Processes of Inference, Learning, and Discovery*. MIT Press: Cambridge, MA.

Komorowski J, Pawlak Z, Polkowski L. 1999. Rough sets: tutorial. In *Rough Fuzzy Hybridation: A New Trend in Decision-making*, Pal SK, Slowinski R (eds). Springer-Verlag: Berlin; 3–98.

Lavalle IH. 1970. *An Introduction to Probability, Decision and Inference*. Holt, Rinehart, Winston: New York.

McKee TE. 2000. Developing a bankruptcy prediction model via rough sets theory. *International Journal of Intelligent Systems in Accounting, Finance and Management* **9**(3): 159–173.

Meisel WM. 1972. *Computer-Oriented Approaches to Pattern Recognition*. Academic Press: New York/London.

Michalski RS, Mozetic L, Hong J, Lavrac N. 1986. The multipurpose incremental learning system AQ 15 and its testing application to three medical domains. In *Proc. 1986 AAAI Conference*, Philadelphia, PA; 1041–1045.

Mitchell TM. 1977. Version Space: a candidate-elimination approach to rule learning. *Proc. 5th IJCAI Conference*, Cambridge, MA; 305–310.

Nguyen HS, Skowron A. 1995. Quantization of real-valued attributes. In *Proc. Second International Joint Conference on Information Sciences*, Wrightsville Beach, NC; 34–37.

Nilsson MJ. 1990. *The Mathematical Foundations of Learning Machines*. Morgan Kaufmann: San Mateo, CA.

Ohrn A, Komorowski J, Skowron A, Synak P. 1998. The design and implementation of a knowledge discovery toolkit based on rough sets—the ROSETTA system. In *Rough Sets in Knowledge Discovery: Methodology and Applications*, Polkowski L, Skowron A (eds). Physica-Verlag: Berlin; 376–399.

Orlowska E (ed.). 1998. *Incomplete Information: Rough Set Analysis*. Physica-Verlag: Berlin.

Pawlak Z. 1991. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic: Dordrecht.

Polkowski L, Skowron A (eds). 1998a. *Rough Sets in Knowledge Discovery: Methodology and Applications*. Physica-Verlag: Berlin.

Polkowski L, Skowron A (eds). 1998b. *Rough Sets in Knowledge Discovery: Applications, Case Studies and Software Systems*. Physica-Verlag: Berlin.

Quinlan JR. 1983. Learning efficient classification procedures and their applications to chess end-games. In *Machine Learning: An Artificial Intelligence Approach* I. Morgan Kauffman: San Mateo, CA; 463–482.

Skowron A, Rauszer C. 1992. The discernibility matrices and functions in information systems. In *Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory*, Slowinski R (ed.). Kluwer Academic: Dordrecht; 331–362.

Slowinski R (ed.). 1992. *Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory*. Kluwer Academic: Dordrecht.

Slowinski R, Stefanowski J. 1992. RoughDAS and RoughClass software implementations of the rough set-approach. In *Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory*, Slowinski R (ed.). Kluwer Academic: Dordrecht; 445–456.

Slowinski R, Vanderpooten D. 2000. A generalized definition of rough approximations based on similarity. *IEEE Transactions on Data and Knowledge Engineering* **12**: 331–336.

Slowinski R, Zopounidis C, Dimitras AI, Susmaga R. 1999. Rough set predictor of business failure. In *Soft Computing in Financial Engineering*, Almeida Ribetro R, Zimmermann HJ, Yager RR, Kacprzyk J (eds). Physica-Verlag: Berlin.

Szladow A. 1993. Datalogic/R for databases mining and decision support. In *Proc. 2nd International Workshop on Rough Sets and Knowledge Discovery*, Banff, Canada; 511–513.

Weiss SM, Kulikowski CA. 1991. *Computer Systems That Learn*. Morgan Kaufmann: San Mateo, CA.

Wong SKM, Ziarko W. 1986. INFER—an adaptative decision support system based on the probabilistic approximate classification. In *Proc. 6th International Workshop on Expert Systems and their Applications*; 713–726.

Zadeh LA. 1977. Fuzzy sets and their applications to pattern classification and clustering analysis. In *Classification and Clustering*, Van Ryzin J (ed.). Academic Press: New York/London.

Ziarko W. 1993. Variable precision rough sets model. *Journal of Computer and Systems Sciences* **46**(1): 39–59.

Ziarko W, Shan N. 1994. KDD-R: a comprehensive system for knowledge discovery in databases using rough sets. In *Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery*, Lin TY, Wildberger AM (eds). Simulation Councils: San Diego; 93–96.

Ziarko W, Yao Y (eds). 2000. *Rough Sets and Current Trends in Computing*, Lecture Notes in Computer Science Vol. 2005. Springer-Verlag: Berlin.

Zopounidis C, Slowinski R, Doumpos M, Dimitras AI, Susmaga R. 1999. Business failure prediction using rough sets. *Fuzzy Economic Review* **4**(1): 3–33.

