Rough Sets Based Fraud Detection in Electrical Energy Consumers

JOSÉ E. CABRAL JUNIOR, JOÃO ONOFRE P. PINTO, EDGAR M. GONTIJO, JOSÉ REIS FILHO
Electrical Engineering Department
Federal University of Mato Grosso do Sul
Cidade Universitária, s/n – CP 549
Campo Grande, MS - 79070-900
BRAZIL

Abstract: - This article describes the theory and application of Rough Sets of fraud detection in electrical energy consumers from databases. The Rough Sets concept of reduct was used to remove conditional attributes and the minimal decision algorithm (MDA) was used to remove insignificant classes of each conditional attribute. The minimized database approach the consumers behavior, allowing a classification rule system to predict fraud consumers profiles. The achieved results are good enough to demonstrate that Rough Sets is a very powerful technique for this type of problem.

Key-Words: - Rough sets, Fraud Detection, Electrical Energy Consumers, Reduct, Minimal Decision Algorithm (MDA), Classification Rules.

1 Introduction

Intelligent Fraud Detection Systems have been intensively addressed in recent past. So far, mostly of the work has been done for credit card and cell phone fraud detection. The most popular soft computing techniques used for the purpose are artificial neural networks [1],[2], and fuzzy logic [3],[4].
However, fraud detection of electrical energy consumers has barely been reported in literature. In general, this problem is solved by in-site inspection. Most of the time, as reported by some electricity companies, the fraud identification rate of this strategy is 5% or below. This because the decision of who has to be inspected is done by a worker, who although is a specialist, cannot efficiently look into all the data available from all company consumers and make a decision. The result is a very expensive process that sometimes does not results in cost reduction for the company. Rough Sets is a soft computing technique that, recently, is being widely applied to Knowledge Data Discovery (KDD) problems. For instance, Rough Sets was shown to be effective in classification rules determination [5]. However, for the best knowledge of the authors, it has never been applied for any type of fraud detection.
Initially, a brief description of Rough Sets theory is made, approaching the main concepts. In the sequence, the solution in the detection of frauds from databases is presented and finally the results of the gotten system are given.

2 Rough Sets Theory

Rough Sets theory was developed by Zdzislaw Pawlak in early 1980's[6]. It deals with the classificatory analysis of data tables (or databases). The main goal of Rough Sets analysis is to synthesize approximation of concepts from the acquired data. Often this concepts are "rough" or "fuzzy", and consequently, some methods or algorithms are necessary to reach them. This justify the applicability of Rough Sets in knowledge discovery in databases. Some concepts of Rough Sets will be presented on next subsections.

2.1 Information and Decision Systems

A data set is represented as a table. The rows represent the objects (examples, cases), and each column an attribute (variable, property). This table is called an information system [6]. Formally, it is a pair $\mathcal{A}$=(U,A), where U is a non-empty finite set of objects, and A is a non-empty finite set of attributes. Often, an information system has one (or more) special attribute representing a decision or an outcome. It is called decision attribute. The information system plus the decision attribute defines a decision system. Formally, $\mathcal{A}$=(U, A$\cup${d}), where d$\notin$A is a decision attribute. The attributes in the set A are called conditional attributes. An information system and its decision system are showed in Table 1.

Cand.	Diploma	Experience	Decision
x1	MSc	Medium	Accept
x2	MBA	Medium	Accept
x3	MCE	Low	Reject
x4	MBA	Medium	Reject
x5	MBA	High	Accept
x6	MSc	Medium	Reject
x7	MSc	High	Accept
x8	MCE	Low	Reject

Table 1- Information system (gray) and decision system (all the table).

2.2 Indiscernibility in Objects

Some objects in the Table 1 are indiscernible. For example, considering the attributes Diploma and Experience, the objects in each subset {x1,x6}, {x2,x4} and {x3,x8} are indiscernible. Considering only the attribute Experience, {x3,x8}, {x1,x2,x4,x6} and {x5,x7} are indiscernible. The indiscernibility relation is an equivalence relation. For more datails, see [6].

2.3 Set Approximation

Analyzing the decision attributes in a decision system, the class set is found. It is just the set of decision values. For the decision system of Table 1, the class set is {Accept, Reject}.
As it can be observed in Table 1, some objects can represent conflicting information. For example, the objects x1 and x6 possess the same values of conditional attributes, however different values in the decision attribute. To deal with this kind of problem, Rough Sets theory defines yours sets approximation. Either $X \subseteq U$ the set of objects with one determined class, is defined:

- Lower approximation ($\underline{X}$): set of all objects of class X that are not indiscernible with none another object. For the decision system of Table 1, the lower approximation for the class "Accept" is {x5, x7}, and for the class "Reject" is {x3, x8}. Maybe $\underline{X}$ has less elements than X, due to elimination of the indiscernible elements in X to reach $\underline{X}$;
- Upper approximation ($\overline{X}$): set of all objects of class X plus the objects of others classes that are indiscernible with some object of class X. For the decision system of Table 1, the upper approximation for the class "Accept" is {x1, x2, x4, x5, x6, x7} and for the class "Reject" is {x1, x2, x3, x4, x6, x8}. Maybe $\overline{X}$ has more elements than X, due to addiction of some elements to reach $\overline{X}$;
- Boundary region (BNX): set of all objects of class X that are indiscernible. For the decision system of Table 1, the boundary region between the classes "Accept" and "Reject" is {x1, x2, x4, x6}.

The lower and upper approximations, together with boundary region, define the regions for the classes. These regions inform how much an object can be said to be inside a class or not. Fig. 1 illustrates the distribution of the objects inside the regions. The dark blue region delimits all the candidates (objects) that for certain are classified as Accept (a crisp set). The white region delimits all the candidates that positively are classified as Reject (a crisp set). Already the blue region (between dark blue and white) defines the candidates that can be classified as Accept or Reject (a rough set). In other words, the darker the blue, greater is the acceptance certainty.

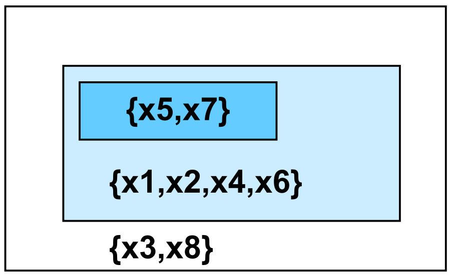

Fig. 1 – Regions for the classes.

It can be interesting to know how much a class is represented or not in a decision system. For such, the accuracy of approximation is defined as:

$$\alpha(X) = |\underline{X}| / |\overline{X}| \qquad (1)$$

If $\alpha(X)=1$, X is crisp (X is precise), and otherwise, if $\alpha(X)<1$, X is "rough" (X is vague).

Candidate	Diploma	Experience	French	Reference	Decision
x1	MBA	Médium	Yes	Excellent	Accept
x2	MBA	Low	Yes	Neutral	Reject
x3	MCE	Low	Yes	Good	Reject
x4	MSc	High	Yes	Neutral	Accept
x5	MSc	Medium	Yes	Neutral	Reject
x6	MSc	High	Yes	Excellent	Accept
x7	MBA	High	No	Good	Accept
x8	MCE	Low	No	Excellent	Reject

Table 2 – Decision System

2.4 Reduct

Given the set of attributes of the decision system defined by Table 2, the reduct of this system can be found. This task consists of eliminating the linear dependent attributes. Or either, to eliminate conditional attributes that do not add any real information to the object. Finding a minimal reduct, a discernibility matrix must be created [6]. This matrix compares each object, identifying in each comparison which attributes possess different values. Later, a discernibility function is applied to the matrix and the linear dependent attributes (and the reduct) are found. To find this minimal reduct from the discernibility matrix is NP-hard. Fortunately, there exist some heuristics that find reducts with a viable computational cost. Although they do not guarantee that the reduct is minimal, the heuristics are more used. The software Rosetta [7] implements some of these heuristics.

2.5 Minimal Decision Algorithm

The minimal decision algorithm (MDA) [5] is used for reductions in decision systems or rule bases. MDA compares the attribute values of an object with the others objects. If it finds attribute values that can be eliminated without the objects becoming indiscernible, the MDA removes this attribute value from the object. Considering object x1 of Table 2. If its first attribute value is eliminated, x1 continues different of all objects. The second attribute value can be eliminated in the same way. Already, the third attribute value of x1 cannot be eliminated because, in case it was eliminated, the object x1 become indiscernible with the object x8.

3 Problem Solution

3.1 Aplication

Rough Sets theory addresses the analysis of tables (database) aiming to approximate concepts and information from these repositories. Often, this information is imprecise and/or has uncertainty, and it needs algorithm or special methodology to determine it.

At first, to solve the fraud detection of electrical energy consumers problem, costumers data was divided into training data and testing data. This is a standard procedure for supervised learning. In the sequence, the repeated registers were eliminated, and for the training data, only the distinct registers remained. Then, rough sets concepts were used. The lower approximation for the concepts (normal and fraud) was found and the registers that did not belong to this subset were eliminated. After this step, a valid reduct was determined, and the linear dependent attributes were eliminated. The elimination of some attributes reduced the dimension of the training dada. Again, since after some attributes elimination some registers became repeated. The repeated ones were eliminated. Then, the Minimal Decision Algorithm (MDA) was applied to the reduced training data. This algorithm was able to significantly reduce the training data. As before, after the application of the MDA, some registers became repeated, and they were eliminated. Finally, for each remaining register a classification rule was derived. The whole set of rules is called classification rules system. The classification rules system can then be tested using the testing data.

3.2 Results

The database had about 100,000 registers, considering only the inspected units. After filtering inconsistent and irrelevant data, the number of registers fell down to 40,000, which 90% was classified as normal and 10% was classified as fraud. The database was equally divided in training data and testing data. Following the steps explained in aplication subsection, the training data was reduced

from 20000 to 1980 registers. The remaining registers resulted in sparse rules, i.e., not all attributes were used to all rules. This makes the classification rules system not so computation intensive. The application of the classification rules system to the testing data resulted in 20% right classification. This is a very promising result, since the ultimate goal is to reach 30% right classification.

4 Conclusion

- Rough Sets is a powerful tool for fraud detection, mainly when does not exist any previous knowledge of the system, but the database;

- Although it is a computation intensive tool, the rough sets algorithms are easy to understand and to implement;

- An hybrid system involving Rough and Fuzzy sets seems to be a good approach for fraud detection problems;

- The obtained system reaches 20% true classification, but further work is being done in order to reach 30%;

- The main problem to reach the percentage of true classification, as in most data mining cases, was the quality of the data, which in many register did not correspond to reality;

- The upper approximation for the concepts (normal and fraud) has been studied to reach a better goal or the ultimate goal of 30% right classification.

References:

[1]R.Brause, T. Langsdorf, M.Hepp, Neural data mining for card fraud detection, *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference,*1999, pp.103-106.

[2]Garavaglia, S, An application of a counter-propagation neural network: simulating the Standard and Poor's Corporate Bond Rating system, *Artificial Intelligence on Wall Street, Proceedings, First International Conference*, 1991, pp. 278-287.

[3] Deshmukh, A.; Talluru, A rule based fuzzy reasoning system for assessing the risk of management fraud, *Systems, Man, and Cybernetics, 1997, Computational Cybernetics and Simulation, 1997 IEEE International Conference*, Vol. 1, 1997, pp. 669-673.

[4] Syeda, M., Yan-Qing Zhang, Yi Pan, Parallel granular neural networks for fast credit card fraud detection, *Proceedings of the 2002 IEEE International Conference*, Vol. 1, 2002, pp. 572-577.

[5] H. Furuta, M. Hirokane, Y. Mikumo, Extraction Method Based on Rough Sets Theory of Rule-Type Knowledge from Diagnostic Cases of Slope-Failure Danger Levels, *Rough Sets in Knowledge Discovery 2 - Application, Case Studies and Software Systems*, Part 1, Chapter 10, pp. 178-192.

[6] Pawlak, Z., Rough Sets, *International Journal of Computer and Information Sciences*, 1982, pp. 341-356.

[7] Ohrn, A. Rosetta: Technical reference manual. Technical report, Knowledge System Group, Norwergian University on Science and Technology, NO. http://rosetta.lcb.uu.se/general/