An Application of the Rough Set Approach to Credit Rating

JaeKyeong Kim*, SungSik Cho**

Abstract

The credit rating represents an assessment of the relative level of risk associated with the timely payments required by the debt obligation. In this paper, we present a new approach to credit rating of customers based on the rough set theory. The concept of a rough set appeared to be an effective tool for the analysis of customer information systems representing knowledge gained by experience. The customer information system describes a set of customers by a set of multi-valued attributes, called condition attributes. The customers are classified into groups of risk subject to an expert's opinion, called decision attribute. A natural problem of knowledge analysis consists then in discovering relationships, in terms of decision rules, between description of customers by condition attributes and particular decisions. The rough set approach enables one to discover minimal subsets of condition attributes ensuring an acceptable quality of classification of the customers analyzed and to derive decision rules from the customer information system which can be used to support decisions about rating new customers.

Using the rough set approach one analyses only facts hidden in data, it does not need any additional information about data and does not correct inconsistencies manifested in data; instead, rules produced are categorized into certain and possible. A real problem of the evaluation of credit rating by a department store is studied using the rough set approach.

Key words: Credit Rating, Rough Set

1. Introduction

Credit rating of a credit card customers has been, for a long time, a major preoccupation of university researchers and practitioners. The first approach to rating the credit of credit card customers started with the use of demographic features gathered at new customer entrance. These features were long considered as objective indicators of bad customers.

Recently, new methods of credit rating have been developed using the usage information along with demographic features. There were developed statistical tools based on multivariate statistical methods (e.g. discriminant analysis, cluster analysis) which classify credit customers into groups of bad or good, or calculate a score representing the degree of arrearage risk using those demographic features which considered significant. The most common methods are those of 'credit scoring', which establish a discriminant function using some of a customers demographic and

* Kyonggi University, Management Information System
** Graduate School of Management, KAIST

usage features, and classify them into high risk or low risk groups (Capon, 1982). Later, tools were developed which were based on multi-criteria decision aid methodology. They also classify customers into groups of risk, but circumvent many of the problems that exist when using discriminant analysis. Finally, tools based on artificial intelligence were developed.

This paper presents a new method called the rough set approach for the analysis and evaluation of the credit rating. The concept of a rough set introduced by Pawlak (1982) proved to be an effective tool for the analysis of information systems (customer information system) describing a set of objects (customers) by a set of multi-valued attributes. In particular, in the case where the set of objects is classified subject to an expert's opinion, this approach enables one to deal with two basic problems of information systems:

- How to reduce the set of attributes to a subset ensuring as good approximation of the expert's classification as the whole set of attributes.
- How to derive decision rules from the information systems in view of explaining a decision policy of the expert.

The decision rules derived with the rough set approach are expressed in terms of significant attributes without any redundancy typical for original data. They are based on facts because each decision rule is supported by a set of real examples. In comparison with other methods of decision analysis, the rough set approach does not correct vagueness manifested in the representation of a decision situation. Moreover, it does not need any additional information like probability in statistics or grade of membership in fuzzy set theory. A thorough comparison of the rough set theory with discriminant analysis, fuzzy set theory and the evidence theory has been made by Krusinska et al. (1992), Dubois and Prade (1992) and Skowron and Gryzmala-Busse (1993).

In the next section some basic concepts of the rough set theory is presented. The third section describes a real problem of credit rating of credit card customers of a department store and the application of rough set theory. In the concluding remarks, the merits of the proposed approach are discussed and possible new directions in the field of credit rating of credit card customers are proposed.

2. Rough sets and neural network

2.1 Rough sets

Pawlak (1982) first introduced rough set theory. The philosophy of the method is based on the assumption that with every object some information (data, knowledge) can be associated. Objects characterized by the same information are indiscernible in view of the available information. The indiscerniblness relation generated in this way is the mathematical basis for the rough set theory. Slowinski & Zopounidis (1995) employed rough set approach in business failure prediction. They used 12 financial ratios and compared rough set approach with statistical approaches. Slowinski et al. (1997) applied the rough set approach in a real problem considered by a Greek bank which finances industrial and commercial firms in Greece presenting a great activity. The bank was interested in investing its funds in the best and dynamic firms.

A. Information system

By an information system we understand the 4-tuple $S=<U,Q,V,\rho>$, where U is a finite set of objects, Q is a finite set of attributes, $V=\bigcup_{q\in Q} V_q$ and

V_q is a domain of the attribute q, and $\rho: U\text{x}Q \to V$ is a total function such that $\rho(x,q) \in V_q$ for every $q \in Q, x \in U$, called an information function. Let $S = < U, Q, V, \rho >$ be an information system and let $P \subseteq Q$ and $x, y \in U$. We say that x and y are indiscernible by the set of attributes P in S iff $\rho(x,q) = \rho(y,q)$ for every $q \in P$. Thus every $P \subseteq Q$ generates a binary relation on U which will be called an indiscerniblness relation, denoted by $IND(P)$. Obviously $IND(P)$ is an equivalence relation for any P. Equivalence classes of $IND(P)$ are called P-elementary sets in S. The family of all equivalence classes of relation $IND(P)$ on U is denoted by $U \mid IND(P)$ or, in short, $U \mid P$.

$Des_p(X)$ denotes a description of P-elementary set $X \in U \mid P$ in terms of values of attributes from P, i.e.

$$Des_p(X) = \{(q,v) : \rho(x,q) = v, \forall x \in X, \forall q \in P\}$$

B. Approximation of Sets

Let $P \subseteq Q$ and $Y \subseteq U$. The P-lower approximation of Y, denoted by $\underline{P}Y$, and the P-upper approximation of Y, denoted by $\overline{P}Y$, are defined as:

$$\underline{P}Y = \bigcup\{X \in U \mid P : X \subseteq Y\}$$

$$\overline{P}Y = \bigcup\{X \in U \mid P : X \cap Y \neq \varnothing\}$$

The P-boundary (doubtful region) of set Y is defined as

$$Bn_p(Y) = \overline{P}Y - \underline{P}Y$$

Set $\underline{P}Y$ is the set of all objects from U which can be certainly classified as elements of Y, employing the set of attributes P. Set $\overline{P}Y$ is the set of objects from U which can be possibly classified as elements of Y, using the set of attributes P. The set $Bn_p(Y)$ is the set of objects which cannot be certainly classified to Y using the set of attributes P only.

With every set $Y \subseteq U$, we can associate an accuracy of approximation of set Y and P in S, or in short, accuracy of Y, defined as:

$$a_p(Y) = \frac{card(\underline{P}Y)}{card(\overline{P}Y)}$$

C. Approximation of a Partition of U

Let S be an information system, $P \subseteq Q$, and let $\Psi = \{Y_1, Y_2, ..., Y_n\}$ be a partition of U. The origin of this partition is independent on attributes from P; it can follow from solving a sorting problem by an expert. Subsets $Y_i, i = 1,...,n$, are categories of partition Ψ. By P-lower and P-upper approximation of Ψ in S we mean sets $\underline{P}\Psi = \{\underline{P}Y_1, \underline{P}Y_2, ..., \underline{P}Y_n\}$ and $\overline{P}\Psi = \{\overline{P}Y_1, \overline{P}Y_2, ..., \overline{P}Y_n\}$, respectively. The coefficient

$$\gamma_P(\Psi) = \frac{\sum_{i=1}^{n} card(\underline{P}Y_i)}{card(U)}$$

is called the quality of approximation of partition Ψ by set of attributes P, or in short, quality of sorting. It expresses the ratio of all P-correctly sorted objects to all objects in the system.

D. Reduction of Attributes

We say that the set of attributes $R \subseteq Q$ depends on the set of attributes $P \subseteq Q$ in S (denotation $P \to R$) iff $IND(P) \subseteq IND(R)$. Discovering dependencies between attributes is of primary importance in the rough set approach to knowledge analysis.

Another important issue is that of attribute reduction. in such a way that the reduced set of attributes provides the same quality of sorting as the original set of attributes. The minimal subset $R \subseteq P \subseteq Q$ such that $\gamma_P(\Psi) = \gamma_R(\Psi)$ is called Ψ-reduct of P (or, simply, reduct if there is no ambiguity in the understanding of Ψ) and denoted by $RED_{\Psi}(P)$. Note that an information system may have more than one Ψ- reduct. Intersection of all Ψ-reducts is called the Ψ-core of P, i.e. $CORE_{\Psi}(P) = \cap RED_{\Psi}(P)$. The core is a collection of the most significant attributes in the system. It can also be empty.

E. Decision Tables

An information system can be seen as a decision table assuming that $Q = C \cup D$ and $C \cap D = \varnothing$, where C are called condition attributes, and D, decision attributes. Decision table $S =< U, C \cup D, V, \rho >$ is deterministic iff $C \to D$; otherwise it is non-deterministic. The deterministic decision table uniquely describes the decisions to be made when some conditions are satisfied. In the case of a non-deterministic table, decisions are not uniquely determined by the conditions. Instead, a subset of decisions is defined which could be taken under circumstances determined by conditions.

From the decision table a set of decision rules can be derived. Let $U \mid IND(C)$ be a family of all C-elementary sets called condition classes, denoted by X_i $(i = 1,\dots,k,$ where k is the number of $U \mid IND(C))$. Let, moreover, $U \mid IND(D)$ be the family of all D-elementary sets called decision classes, denoted by Y_j $(j = 1,\dots,n,$ where n is the number of $U \mid IND(D))$.

$Des_C(X_i) \Rightarrow Des_D(Y_j)$ is called the (C, D)-decision rule. The rules are logical statements 'if...then...' relating descriptions of condition and decision classes. The set of decision rules for each decision class Y_j $(j = 1,\dots,n)$ is denoted by $\{r_{ij}\}$. More precisely,

$$\{r_{ij}\} = \{Des_C(X_i) \Rightarrow Des_D(Y_j) : X_i \cap Y_j \neq \varnothing, i = 1,\dots,k\}$$

Rule r_{ij} is deterministic iff $X_i \subseteq Y_j$, and r_{ij} is non-deterministic otherwise.

Procedures for derivation of decision rules from decision tables were presented by Boryczka & Slowinski (1988), Slowinski & Stefanowski (1992), Grzymala-Busse (1992), Siegel et al. (1993) and Ziarko et al. (1993).

F. Decision support using decision rules

Decision rules derived from a decision table can be used for recommendations concerning new objects. Specifically, matching its description to one of the decision rules can support the classification of a new object. The matching may lead to one of four situations (Slowinski & Stefanowski, 1994):

(a) the new object matches one deterministic rule,

(b) the new object matches more than one deterministic rules suggesting, however, the same decision class,

(c) the new object matches one non-deterministic rule or several rules suggesting different decision classes,

(d) the new object does not match any of the rules.

3. A Credit Rating Problem

In this section a real problem statement will be presented followed by the application of the rough set approach.

3.1 Problem Statement

The problem has been considered by a department store, D department store, which has a lot of credit card customers. The sales department of the D department store wants to have more and more credit card customers to increase the sales volume. However, for the management department, more credit card customers means more arrearages and additional cost. So, the management department is interested in rating the credit of card customers. The primary issue is to decrease the arrearages by adjusting the spending line management system. The spending line management system is composed of a lot of rules which determine whether to increase or decrease the spending line of a credit card customer. When the spending line is increased, the sales volume can be increased, but the risk of a large arrearages also goes high.

The spending line management system of the D department store has 8 credit grades, S, A, B, C, D, E, Y, and Z. New customers starts at D or E grade, and every six month the grades are adjusted according to their usage history. One grade difference means 0.3 million or 0.4 million won difference of spending line. For example, when a customer who has D grade gets C grade, his/her spending line goes from 2 million won to 2.3 million won. Customers who delayed the payment for 3 months in a row get the Z grade and their spending line decreases radically and customers who delayed the payment more than 6 times get the Y grade.

According to the spending line management system alone, the management department cannot determine who would go bad or not. They want to find another rule that can discriminate bad customers. For this matter, 450 samples of credit card customers in 1998 was chosen. From 450 samples 12 samples were removed for missing data. The bad customers means the customers who delayed the payment more than 6 times. The selected customers were classified into three categories of risk. The result of classification is represented by decision attribute d : d = 0 means 'acceptable' (low-risk group); d = 1 means 'uncertain'; d = 2 means 'unacceptable' (high-risk group). The distribution of customers in the categories is the following: 147 acceptable, 142 uncertain, and 149 unacceptable. The customers were evaluated according to a total of 8 attributes, four of them are quantitative and the other four being qualitative. A part of this multi-attribute information is presented in Table 1. Attributes A_1 – A_5 show some general information and attributes A_6 – A_7 show card usage information. Attributes A_1 – A_5 are gathered at the point of member

Table 1. Original information system

Customers	Attributes								Decision
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	
1	37	2	1	0	210	76000	2	2000000	0
2	31	2	2	2	240	4300780	26	6200000	0
3	34	2	1	0	210	585000	1	2000000	0
4	36	2	1	2	110	907500	4	2000000	0
5	34	2	1	1	210	333360	4	2000000	0
6	36	2	2	0	210	383570	14	3500000	1
7	41	1	1	1	220	6953410	49	3500000	1
8	40	1	1	1	210	8821143	218	4100000	1
9	51	1	1	1	260	416940	5	3500000	1
10	34	2	1	2	230	948003	14	1400000	1
11	41	1	2	0	220	2056220	11	2000000	2
12	44	1	1	2	220	53540	1	2000000	2
13	55	1	2	2	210	1508347	24	2000000	2
14	30	2	1	1	240	137700	2	1700000	2
15	37	1	1	1	110	896050	9	2000000	2

A_1 : Age A_2 : Marriage A_3 : Sex A_4 : Residence style
A_5 : Entrance Branch A_6 : Amount of usage A_7 : Number of usage A_8 : Spending line

Table 2. The definition of norms for quantitative attributes

Attributes	Codes 0	1	2	3	4	5	6
A_1	[0,20)	[20,30)	[30,40)	[40,50)	[50,*)		
A_6	[0,1.2M)	[1.2M,2.7M)	[2.7M,6.4M)	[6.4M,13M)	[13M,24M)	[24M,*)	
A_7	[0,10)	[10,30)	[30,50)	[50,100)	[100, *)		
A_8	[0,1.4M)	[1.4M,2.9M)	[2.9M,4.2M)	[4.2M,6.2M)	[6.2M,8M)	[8M,10.5M)	[10.5M,*)

* : infinite M : Million

entrance. A_5 is the code of the branches of D department store, where a customer entered.

Table 3. The codes for the terms of qualitative attributes

Attributes	Terms	Codes
A_2	Married	1
	Not married	2
A_3	Male	1
	Female	2
A_4	Own house	1
	Rent	2
	etc.	3
A_5	Branch 1	110
	Branch 2	210
	Branch 3	220
	Branch 4	230
	Branch 5	240
	Branch 6	250
	Branch 7	260
	Branch 8	291

3.2 Application of the Rough Set Approach

The application starts with an appropriate conditioning of the information system representing a past experience of the customers with the set of 438 customers (objects). The conditioning consists of translating the values of quantitative attributes A_1, A_6 – A_8, into qualitative terms. The qualitative terms of all attributes are then coded using natural numbers. The translation of the quantitative attributes into qualitative terms is done according to some norms following from the credit card customer's experience and some standards of spending line management system. Table 2 and Table 3 present the definition of norms for the quantitative attributes and the codes of all the qualitative terms respectively.

The use of norms translating the quantitative attributes into qualitative terms is not imposed by the rough set approach but by a practical interpretation of the quantitative attributes. Even if an attribute represents a continuous measure, such as financial ratio,

Table 4. Coded information system

Customers	Coded Attributes A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	Decision d
1	2	2	1	0	210	0	0	1	0
2	2	2	2	2	240	2	1	4	0
3	2	2	1	0	210	0	0	1	0
4	2	2	1	2	110	0	0	1	0
5	2	2	1	1	210	0	0	1	0
6	2	2	2	0	210	0	1	2	1
7	3	1	1	1	220	3	2	2	1
8	3	1	1	1	210	3	4	2	1
9	4	1	1	1	260	0	0	2	1
10	2	2	1	2	230	0	1	1	1
11	3	1	2	0	220	1	1	1	2
12	3	1	1	2	220	0	0	1	2
13	4	1	2	2	210	1	1	1	2
14	2	2	1	1	240	0	0	1	2
15	2	1	1	1	110	0	0	1	2

the expert usually interprets the values of this attribute in qualitative terms, i.e. low, medium or high. The norms used for this interpretation come from tradition, habits or convention. As they are consequently used from the beginning of problem setting until final explanation of decision rules, they do not falsify the original image of the decision situation.

The result of using the norms and the codes for translation of the original information system, one obtains the coded information system presented in Table 4.

3.3 Presentation of Rules

Using the method described in the previous section the approximations are made for each particular category of customers by each approximation is calculated. The results are very satisfactory since the accuracy of all approximations is near perfect, i.e. near to one. Therefore the quality of sorting is also equal to one. This result is very significant for the D department store's management department, because in spite of a large information system the firms are very well discriminated among them.

Table 5. Accuracy of approximation of each category by 8 attributes.

		Predicted		
	d	0	1	2
Actual	0	147		
	1		142	
	2	2	1	146
	Hit Ratio	0.987	0.993	1

The next step of the rough set analysis is the construction of minimal subsets of independent attributes ensuring the same quality of sorting as the whole set Q, i.e. reducts of Q in the information system. But, in this experiment only one reduct was found:

$$\{A_2, A_3, A_4, A_5, A_6, A_8\}$$

Usually there can be several reducts that ensure the same quality of sorting. Then the intersection of all reducts is the core of attributes. In our case, only one reduct was found, thus being the core attributes. Our reduct shows that attributes A_1, A_7 are redundant to discriminating decision category. With this reduct sorting rule is generated. Table 6. shows some example of rules generated:

Table 6. Rules generated

	A_2	A_3	A_4	A_5	A_6	A_8	*d*
rule #1	2	1	2	220	1	2	0
rule #2	1	2	2	210	0	0	2
rule #3	2	2	2	210	0	3	1

4. Conclusions

The aim of this paper has been to show that the rough set theory is a useful tool for discovery of a preferential attitude of the decision making in multi-attribute sorting problems, in particular, multi-attribute evaluation of the credit rating of credit card customers of a department store. The concept of a rough set appeared to be an effective tool for the analysis of customer information systems representing knowledge gained by experience. Using the rough set approach one analyses only facts hidden in data, it does not need any additional information about data and does not correct inconsistencies manifested in data; instead, rules produced are categorized into certain and possible.

The set of derived decision rules can be used by the D department store to support determining whether a customer would be a good or bad customer. The decision support is made by matching a new applicant firm to one of decision rules.

References

[1] Dubois, D., and Prade, H., 'Putting rough sets and

fuzzy sets together', in Slowinski, R. (ed), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic, Dordrecht, 1992, pp. 203-232.

[2] Capon, N., (1982). "Credit Scoring Systems: A Critical Analysis," *Journal of Marketing*, Vol.46, pp.83-88.

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

[4] Krusinska, E., Slowinski, R. and Stefanowski, J. "Discriminant versus rough set approach to vague data analysis", *Applied Stochastic Models and Data Analysis*, 8, 1992, pp.1-17.

[5] Pawlak, Z., "Rough sets," *International Journal of Information and Computer Sciences*, Vol. 11, pp. 341-356.

[6] Skowron, A. and Grzymala-Busse, J.W., 'From the rough set theory to the evidence theory', in Fedrizzi, M., Kacpryk, J. and Yager, R.R. (eds), *Advances in the Dempster-Shafer Theory of Evidence*, John Wiley, New York, 1993.