DECISION MAKING VIA DEGREE OF DEPENDENCY

A.M. Kozae [a], A.A. Abo Khadra [b], T. Medhat [b]

Abstract- In some information systems, we can't eliminate any of its attributes. But we need to eliminate one or more of them with some approximation depending on the characteristics of attributes such as its dangerous, cost, long time of measuring, etc. The aim of this work is to suggest a new approach for reduction of information system attributes with an approximation "some error" depending on the degree of dependency of attributes.

Index Terms- Rough set, degree of dependency, reducts and decision making.

I. Introduction

Rough sets theory was first introduced by Pawlak in the 1980s [6] and it has been applied in many applications such as machine learning, knowledge discovery, and expert systems [2,3,4,7]. It deals with the classificatory analysis of data tables. The data can be acquired from measurements or from human experts. The main goal of the rough set analysis is to synthesize approximation of concepts from the acquired data and makes reduction of data to a minimal representation. Many rough sets models have been developed in the rough set community in the last decades [1,6,8,9,10,11], including VPRS [11] and GRS [5]. Some of them have been applied in the industry data mining projects such as stock market prediction, patient symptom diagnosis, telecommunication churner prediction, and financial bank customer attrition analysis to solve challenging business problems [2,4]. These rough set models focus on the extension of the original model proposed by Pawlak [6,7] and attempt to deal with its limitations such as handling statistical distribution or noisy data.

[a] Department of Mathematics, Faculty of Science, Tanta University, 31527, Tanta, Egypt.
[b] Department of Physics and Engineering Mathematics, Faculty of Engineering, Tanta University, 31521, Tanta, Egypt.

E-mail addresses: akozae55@yahoo.com (A.M. Kozae), aaabokhadra@yahoo.com (A.A. Abo Khadra), tmedhatm@yahoo.com (T. Medhat).

The approach we used depends on the degree of dependency between the condition attributes and decision attributes or classified attribute "attribute which makes a classification of objects as equivalence classes with respect to all condition attributes". The degree of dependency measure always lies in the range [0,1], with 0 indicating no dependency and 1 indicating total dependency.

II. Basic Concepts

A. Information System

A data set is represented as a table, where each row represents a case, an event, a patient, or simply an object. Every column represents an attribute (a variable, an observation, a property, etc.) that can be measured for each object; the attribute may be also supplied by a human expert or user. This table is called an information system IS [5] which can be defined as IS=(U,A,ρ,V), where U is a non-empty finite set of objects called a universe and A is a non-empty finite set of attributes.

$$IS=(U,A,\rho,V)$$

Any subset $X \subseteq U$ wil be called a concept or a category in U, Each attribute $a \in A$ can be viewed as a function that maps elements of U into a set Va, where the set Va is called the value of set of the attribute a.

$$\rho: UxA \rightarrow Va$$

B. Mathematical Model in Information System

To extract knowledge from information system, it is necessary to find mathematical models using the given data.

B.1. Indiscernibility Relation

With any $P \subseteq A$ there is an associated equivalence relation IND(P):

$$IND(P) = \{(x,y)\in U^2: \forall\ a\in P,\ a(x) = a(y)\}$$

The partition of U, generated by IND(P) is denoted U/IND(P) (or U/P) and can be calculated as follows:

$$U/IND(P) = \otimes\{a \in P : U/IND(\{a\})\},$$

Where

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq Ø\}$$

If $(x,y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_P$.

B.2. Lower and upper approximations

Let $X \subseteq U$. X can be approximated using only the information contained within P by constructing the P-lower and P-upper approximations of X:

$$\underline{P}X = \{x : [x]_P \subseteq X\}$$
$$\overline{P}X = \{x : [x]_P \cap X \neq Ø\}$$

B.3. Positive, negative, and boundary regions

Let C and D be equivalence relations over U, then the positive, negative and boundary regions can be defined as:

$$POS(C,D) = \cup_{X \in U} \underline{P}X$$
$$NEG(C,D) = U - \cup_{X \in U} \overline{P}X$$
$$BND(C,D) = U - POS(C,D) \cup NEG(C,D)$$

The positive region contains all objects of U that can be classified to classes of U/D using the information in attributes C. The boundary region, BND(C,D), is the set of objects that can possibly, but not certainly, be classified in this way. The negative region, NEG(C,D), is the set of objects that cannot be classified to classes of U/D.

III. Degree of dependency and reduction

An important issue in data analysis is discovering dependencies between attributes.

Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if all attribute values from D are uniquely determined by values of attributes from C. If there exists a functional dependency between values of D and C, then D depends totally on C. In rough set theory, dependency is defined in the following way: For $C, D \subset A$, it is said that D depends on C in a degree K ($0 \leq K \leq 1$), denoted $C \Rightarrow_k D$, if

$$K(C,D) = \|POS(C,D)\| / \|U\|$$

Where $\| . \|$ is the cardinality of a set.

If k=1, D depends totally on C, if 0< k <1, D depends partially (in a degree k) on C, and if k = 0 then D does not depend on C.

By calculating the change in dependency when an attribute is removed from the set of considered conditional attributes, a measure of the significance of the attribute can be obtained. The higher the change in dependency, the more significant the attribute is. If the significance is 0, then the attribute is dispensable.

A. Reduction of attributes without error ratio

Let $C \subseteq A$, $a \in C$, then a is superfluous attributes [10] in B if:

$$U/IND(C)=U/IND(C-\{a\}).$$

The set M is called a minimal reduct of C iff:
(i) U/IND(M)=U/IND(C).
(ii) $U/IND(M) \neq U/IND(M-\{a\}), \forall a \in M$

Attributes can be divided into condition attributes C and decision attributes D. An attribute $a \in C$ is called superfluous with respect to D if K(C-{a},D)/K(C,D)=1, otherwise a is indispensable in C [5].

Eliminating a superfluous C-attribute will not decrease or increase the degree of dependency. This means that this attribute is not necessary for the decision.

A subset M of the condition attributes is called a reduct of C with respect to D if:
(i) K(M,D)/K(C,D) =1.
(ii) $K(M-\{a\},D)/K(M,D) \neq 1 \quad \forall a \in M$

B. Reduction of attributes with error ratio

For an information system which has condition and decision attributes, we can eliminate one of its condition attributes c_i if $K(C-\{c_i\},D)/K(C,D)=\alpha$.

For an information system which has condition attributes only, we make a classification of all objects according to all condition attributes, i.e. each similar objects will be put at the same equivalence class, then name the equivalence classes as E1, E2,… to En, where E={E1,E2,….,En} is the set of all equivalence classes. By eliminating the similar objects except one of them in each equivalence class, then the number of remaining objects will equal to the number of equivalence classes.

Now there is a new definition of degree of dependency as follow:

$$K(C,E) = \|POS(C,E)\|/\|E\|=1,$$

this means that E must be totally depending on C.

If $K(C-\{c_i\},E)=1$, then c_i is a superfluous attribute, else c_i is a core attribute " without error ratio".

If $\alpha \le K(C-\{c_i\},E)<1$, then c_i is a superfluous attribute, else c_i is a core attribute " with error ratio depending on the value of K, α".

We can change the number α according to the limit of error ratio. The following example indicates the above notion.

Example 1.

In the following information system, we give some mechanical properties of some alloys "29 alloys" according to system in weight percentage as follows:

U={$Bi_{55.5}Pb_{44.5}$, $Bi_{50}Pb_{28}Sn_{22}$, $Bi_{50}Pb_{27}Sn_{13}Cd_{10}$, Standard, $Bi_{50}Pb_{25}Sn_{12.5}Cd_{12.5}$, $Bi_{50}Pb_{25}Sn_{25}$, $Bi_{51.6}Pb_{40.2}Cd_{8.2}$, $Pb_{75}Sn_{25}$, $Pb_{60}Sn_{30}Cd_{10}$, $Bi_{30}Pb_{40}Sn_{20}Cd_{10}$, $Bi_{60}Pb_{20}Sn_{10}Cd_{10}$, $Pb_{50}Sn_{50}$, $Bi_{10}Pb_{50}Sn_{40}$, $Bi_{20}Pb_{50}Sn_{30}$, $Bi_{40}Pb_{50}Sn_{10}$, $Bi_{50}Pb_{50}$, $Bi_{55.5}Pb_{44.5}$, $Bi_{55.5}Pb_{42.5}Sn_2$, $Bi_{55.5}Pb_{39.5}Sb_5$, $Bi_{55.5}Pb_{39.5}Cd_5$, Sn70Pb30, Sn40Pb60, $Sn_{96.5}Ag_{3.5}$, $Sn_{95}Ag_5$, $Sn_{95}Sb_5$, $Sn_{91}Zn_9$, $Sn_{90}Zn_9Bi_1$, $Sn_{88}Zn_9Bi_1Cu_2$, $Sn_{89}Sb_{10}Bi_1$} respectively.

And 5 mechanical properties A={Young's modulus E (10^9) Pa, Melting Point, Internal friction Q^{-1} (10^{-3}), Resistivity μΩ.cm , hardness} respectively.

Let U={1,2,3,………..,29}, A={a1,a2,a3,a4,a5},

As shown in the following table

U/A	a1	a2	a3	a4	a5
1	17.775	124.7	60	137.56	6.72
2	15.846	97	70.3	85.42	7.76
3	13.244	71.8	99.7	96.14	7.99
4	19.576	72	59.5	104.74	7.41
5	11.98	70	90.57	90.32	8.09
6	13.65	98	88.86	91.875	8.3
7	12.237	91.5	90.65	122.13	5.61
8	26.5	183	81.2	40.5	9.4
9	33.4	210	64	27	9.5
10	35	120	75	47.6	11.4
11	36	90	70	11.1	11.26
12	27	220	190	23.81	6.63
13	26	208	150	33.33	7.5
14	24	191	120	50	9.18
15	14	166	120	76.92	7.6
16	20	145	100	166.67	6.53
17	40.77	124.7	95.03	119.692	9.43
18	21.69	117.73	78.28	117.01	8.9
19	16.64	132.64	30.19	180.8	11.5
20	13.61	92.85	145.87	144.74	10.744
21	16.7	191	138	36.07	7.98
22	17	235	87	34	8.76
23	43.7	228.17	71.1	16.8	13
24	56.8	227.25	67.8	17.8	13.7
25	34.2	241.24	53.3	22.6	11.5
26	37.2	197.5	74.6	17.2	9.7
27	38.3	196.7	65.4	16.2	5.7
28	46.6	194.7	63.5	16	13.2
29	22.6	197	31.6	35	16.2

Table I
"Mechanical properties of some alloys"

By converting each attribute value to a value from 0 to 1 as follows:

$$V_a=(V_{max}-V_a)/(V_{max}-V_{min}),$$

where V_{max} and V_{min} are the maximum and minimum value for each attribute

As shown in the following table:

U/A	a1	a2	A3	a4	a5
1	0.129	0.212	0.186	0.745	0.327
2	0.086	0.105	0.250	0.437	0.394
3	0.028	0.007	0.434	0.501	0.409
4	0.169	0.007	0.183	0.551	0.372
5	0	0	0.377	0.466	0.415
6	0.037	0.108	0.367	0.475	0.429
7	0.005	0.083	0.378	0.654	0.256
8	0.323	0.439	0.319	0.173	0.499
9	0.477	0.544	0.211	0.093	0.506
10	0.513	0.194	0.280	0.215	0.627
11	0.535	0.077	0.249	0	0.618
12	0.335	0.583	1	0.074	0.313
13	0.312	0.536	0.749	0.130	0.368
14	0.268	0.470	0.561	0.229	0.474
15	0.045	0.373	0.561	0.387	0.374
16	0.178	0.291	0.436	0.916	0.307
17	0.642	0.212	0.405	0.639	0.501
18	0.216	0.185	0.300	0.624	0.467
19	0.103	0.243	0	1	0.634
20	0.036	0.088	0.723	0.787	0.585
21	0.105	0.470	0.674	0.147	0.408
22	0.112	0.642	0.355	0.134	0.458
23	0.707	0.615	0.255	0.033	0.730
24	1	0.611	0.235	0.039	0.775
25	0.495	0.666	0.144	0.067	0.634
26	0.562	0.496	0.277	0.035	0.518
27	0.587	0.492	0.220	0.030	0.262
28	0.772	0.485	0.208	0.028	0.743
29	0.236	0.494	0.008	0.140	0.935

Table II

By dividing the interval [0,1] into 4 parts as follows : [0,0.25[=1, [0.25,0.5[=2, [0.5,0.75[=3, [0.75,1]=4

This is shown in the following table

U/A	a1	a2	a3	a4	a5
1	1	1	1	3	2
2	1	1	2	2	2
3	1	1	2	3	2
4	1	1	1	3	2
5	1	1	2	2	2
6	1	1	2	2	2
7	1	1	2	3	2
8	2	2	2	1	2
9	2	3	1	1	3
10	3	1	2	1	3
11	3	1	1	1	3
12	2	3	4	1	2
13	2	3	3	1	2
14	2	2	3	1	2
15	1	2	3	2	2
16	1	2	2	4	2
17	3	1	2	3	3
18	1	1	2	3	2
19	1	1	1	4	3
20	1	1	3	4	3
21	1	2	3	1	2
22	1	3	2	1	2
23	3	3	2	1	3
24	4	3	1	1	4
25	2	3	1	1	3
26	3	2	2	1	3
27	3	2	1	1	2
28	4	2	1	1	3
29	1	2	1	1	4

Table III

After applying the rough set technique "indiscernibility relation" on Table III, we get to the equivalence classes as shown in Table IV:

Equivalence 0bjects	Equivalence Classes E
{1,4}	E1
{2,5,6}	E2
{3,7,18}	E3
{8}	E4
{9,25}	E5
{10}	E6
{11}	E7
{12}	E8
{13}	E9
{14}	E10
{15}	E11
{16}	E12
{17}	E13
{19}	E14
{20}	E15
{21}	E16
{22}	E17
{23}	E18
{24}	E19
{26}	E20
{27}	E21
{28}	E22
{29}	E23

Table IV

The following table indicates the elimination of similar objects except one of them for each equivalence class.

E/A	a1	a2	a3	a4	a5
E1	1	1	1	3	2
E2	1	1	2	2	2
E3	1	1	2	3	2
E4	2	2	2	1	2
E5	2	3	1	1	3
E6	3	1	2	1	3
E7	3	1	1	1	3
E8	2	3	4	1	2
E9	2	3	3	1	2
E10	2	2	3	1	2
E11	1	2	3	2	2
E12	1	2	2	4	2
E13	3	1	2	3	3
E14	1	1	1	4	3
E15	1	1	3	4	3
E16	1	2	3	1	2
E17	1	3	2	1	2
E18	3	3	2	1	3
E19	4	3	1	1	4
E20	3	2	2	1	3
E21	3	2	1	1	2
E22	4	2	1	1	3
E23	1	2	1	1	4

Table V

The following table gives the degree of dependency for each attributes.

Degree of dependency	Ratio	Percentage
K(A-{a1},A) =	21/23	91.30%
K(A-{a2},A) =	18/23	78.26%
K(A-{a3},A) =	13/23	56.52%
K(A-{a4},A) =	17/23	73.91%
K(A-{a5},A) =	23/23	100.00%

Table VI

From Table VI, we can eliminate the attribute a5 without any error ratio, or eliminate the attribute a1 with error ratio 8.7% and so on.

The selection of attribute which can be eliminated depends on the degree of dependency and the its characteristic "dangerous, cost,…etc"

IV. Conclusion

The use of all information resulting from an experimental measurement could be a time consuming task. Therefore, the process of eliminating superfluous attributes is important for saving time and effort. The use of these attributes in the analysis would not affect the final decision.

By using the degree of dependency and the characteristic of attribute, we can eliminate it according to the limit of error ratio.

References

[1] Hu, X., Cercone N., Han, J., Ziarko, W, "GRS: A Generalized Rough Sets Model", in Data Mining, Data Mining, Rough Sets and Granular Computing, T.Y. Lin, Y.Y.Yao and L. Zadeh (eds), Physica-Verlag, 447- 460, 2002.

[2] Lin T.Y., "From rough sets and neighborhood systems to information granulation and computing in words", Proceedings of European Congress on Intelligent Techniques and Soft Computing, 1602-1607, 1997.

[3] Lin T.Y., "Granular computing on binary relations I: data mining and neighborhood systems, II: rough set representations and belief functions", In: Rough Sets in Knowledge Discovery , Lin T.Y., Polkowski L., Skowron A., (Eds.). Physica-Verlag, Heidelberg ,107-140, 1998.

[4] Lin T.Y., Yao Y.Y., Zadeh L.A., (Eds.) " Rough Sets, Granular Computing and Data Mining", Physica-Verlag, Heidelberg, 2002.

[5] Medhat T., "Topological applications on information analysis by rough sets", Master thesis, Egypt, Tanta University, Faculty of Engineering, 2004.

[6] Pawlak Z., "Rough Sets", International Journal of Information and computer Science, 11(5):341-356, 1982.

[7] Pawlak Z., "Rough Sets - Theoretical Aspects of Reasoning about data.", Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[8] Polkowski, L., Skowron, A., "Rough mereology", Proc. ISMIS, Charlotte, NC, 85-94, 1994.

[9] Polkowski, L., Skowron, A., "Rough mereology: A new paradigm for approximate reasoning", J. of Approximate Reasoning, 15(4), 333-365, 1996.

[10] Skowron, A., Stepaniuk, J., "Tolerance approximation spaces", Fundamenta Informaticae, 27(2-3), 245-253, 1996.

[11] Ziarko, W., "Variable Precision Rough Set Model", Journal of Computer and System Sciences, 46(1), 39-59, 1993.