A Rough Set Interpretation of User's Web Behavior: A Comparison with Information Theoretic Measure

George V. Meghabghab

Roane State
Dept of Computer Science Technology
Oak Ridge, TN, 37830
gmeghab@hotmail.com

Abstract

Searching for relevant information on the World Wide Web is often a laborious and frustrating task for casual and experienced users. To help improve searching on the Web based on a better understanding of user characteristics, we address the following research questions: What kind of information would rough set theory shed on user's web behavior? What kind of rules can we extract from a decision table that summarizes the behavior of users from a set of attributes with multiple values in such a case? What kind of decision rules can be extracted from a decision table using an information theoretic measure? (Yao 2003) compared the results of granularity of decision making systems based on rough sets and information theoretic granulation methods. We concur with Yao, that although the rules extracted from Rough Set(RS) and Information Theoretic(IT) might be equal, yet the interpretation of the decision is richer in the case of RS than in the case of IT.

General Introduction to Rough Set Theory and Decision Analysis

The rough set approach to data analysis and modeling (Pawlak 1997, 2002) has the following advantages: a- is based on the original data and does not need any external information (probability or grade of membership); b- It is a suitable tool for analyzing quantitative and qualitative attributes; c-It provides efficient algorithms for finding hidden patterns in data; d-It finds minimal sets of data (data reduction); e-It evaluates the significance of data.

We show that the rough set theory is a useful tool for analysis of decision situations, in particular multi-criteria sorting problems. It deals with vagueness in representation of a decision situation, caused by granularity of the representation. The rough set approach produces a set of decision rules involving a minimum number of most important criteria. It does not correct vagueness manifested in the representation; instead, produced rules are categorized into deterministic and non-deterministic. The set of decision rules explains a decision policy and may be used for decision support. Mathematical decision analysis intends to bring to light those elements of a decision situation that are not evident for actors and may influence their attitude towards the situation. More precisely, the elements revealed by the mathematical decision analysis either explain the situation or prescribe, or simply suggest, some behavior in order to increase the coherence between evolution of the decision process on the one hand and the goals and value system of the actors on the other. A formal framework for discovering facts from representation of a decision situation has been given by (Pawlak 1982) and called rough set theory. Rough set theory assumes the representation in a decision table in which there is a special case of an information system. Rows of this table correspond to objects (actions, alternatives, candidates, patients etc.) and columns correspond to attributes. For each pair (object, attribute) there is a known value called a descriptor. Each row of the table contains descriptors representing information about the corresponding object of a given decision situation. In general, the set of attributes is partitioned into two subsets: condition attributes (criteria, tests, symptoms etc.) and decision attributes (decisions, classifications, taxonomies etc.). As in decision problems the concept of criterion is often used instead of condition attribute; it should be noticed that the latter is more general than the former because the domain (scale) of a criterion has to be ordered according to decreasing or increasing preference while the domain of a condition attribute need not be ordered. Similarly, the domain of a decision attribute may be ordered or not. In the case of a multi-criteria sorting problem, which consists in assignment of each object to an appropriate predefined category (for instance, acceptance, rejection or request for additional information), rough set analysis involves evaluation of the importance of particular criteria: a- construction of minimal subsets of independent criteria b- having the same discernment ability as the whole set; c-non-empty intersection of those minimal subsets to give a core of criteria which cannot be eliminated without it; d-disturbing the ability of approximating the decision; e-elimination of redundant criteria from the decision table;6-

the generation of sorting rules (deterministic or not) from the reduced decision table, which explain a decision; f- development of a strategy which may be used for sorting new objects.

Rough Set Modeling of User Web Behavior

The concept of rough set theory is based on the assumption that every object of the universe of discourse is associated with some information. Objects characterized by the same information are indiscernible in view of their available information. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. The concepts of rough set and fuzzy set are different since they refer to various aspects of non-precision. Rough set analysis can be used in a wide variety of disciplines; wherever large amounts of data are being produced, rough sets can be useful. Some important application areas are medical diagnosis, pharmacology, stock market prediction and financial data analysis, banking, market research, information storage and retrieval systems, pattern recognition (including speech and handwriting recognition), control system design, image processing and many others. Next, we show some basic concepts of rough set theory. 20 Users from Roane State were used to study their web characteristics. The results of the fact based query "Limbic functions of the brain" is summarized in Table 1 (S=1,2;M=3,4;L=5,7;VL=8, 9,10). The notion of a User Modeling System presented here is borrowed from (Pawlak 1991). The formal definition of a User Modeling System (UMS) is represented by S=(U, Ω, V, f) where: U is a non-empty, finite set of users called the universe;Ω is a non-empty, finite set of attributes: CUD, in which C is a finite set of condition attributes and D is a finite set of decision attributes; $V = \cup V_q$ is a non empty set of values of attributes, and V_q is the domain of q (for each qε Ω); f is a User Modeling Function :

$$f: U \times \Omega \rightarrow V$$

such that: $\exists$ f (q , p) ε V_p $\forall$ p ε U and qε Ω

$$f_q: \Omega \rightarrow V$$

such that: $\exists$ f_q (p) =f(q , p) $\forall$ p ε U and qε Ω is the user knowledge of U in S.

Users	W	H	SE	E
1	L	M	M	F
2	S	S	S	SU
3	M	L	L	SU
4	L	L	VL	F
5	L	L	L	F
6	M	M	L	SU
7	S	M	S	F
8	L	L	M	F
9	M	S	L	F
10	M	S	M	SU
11	S	M	M	SU
12	M	S	M	SU
13	M	M	L	SU
14	VL	M	VL	SU
15	S	S	S	SU
16	S	M	S	SU
17	S	S	S	SU
18	S	S	S	SU
19	S	S	S	F
20	S	S	M	F

Table 1. Summary of all users

This modeling system can be represented as a table in which columns represent attributes and rows are users and an entry in the q^{th} row and p^{th} column has the value f(q,p). Each row represents the user's attributes in answering the question. Consider the following example: U={1,2,3,4,5,6,7,8,9,10,11,12,13, 14,15,16,17, 18,19,20}= set of 20 users who searched the query, Ω={Searches, Hyperlinks, Web Pages}={SE,H,W}= set of 3 attributes. Since some users did search more than others, browsed more than others, scrolled down web pages more than others, a simple transformation of table 1 yields a table up to 3 different attributes with a set of values ranging form: small(S), medium(M), large(L), and very large(VL).

Ω={SE,H,W},V_{SE}={M,S,L,VL,L,L,S,M,L,M,M,M,L,VL,S, ,S,S,S,M},V_H={M,S,L,L,L,M,M,L,S,S,M,S,M,M,S,M,S,S,S ,S},V_W={L,S,M,L,L,M,S,L,M,M,S,M,M,VL,S,S,S,S,S,S}.

The modeling system will now be extended by adding a new column E representing the expert's evaluation of the user's knowledge whether the user's succeeded in finding the answer or failed to find the answer. In a new UMS, S is represented by S=(U, Ω, V, f), f_q (p) where q ε Ω and p ε PU={P-E}is the user's knowledge about the query, and f_q(p) where q ε Ω and p = E is the expert's evaluation of the query for a given student. E is the decision attribute. Consider the above example but this time: $\Omega=\Omega_u \cup \Omega_e$={SE,H,W}∪E, where E= SU or F; V_{SE}= {M,S,L,VL,L,L,S,M,L,M,M,M,L,VL,S,S,S,S,S,M},V_H={M ,S,L,L,L,M,M,L,S,S,M,S,M,M,S,M,S,S,S,S},V_W={L,S,M,L ,L,M,S,L,M,M,S,M,M,VL,S,S,S,S,S,S}V_E={F,SU,SU,F,F,S U,F,F,F,SU,SU,SU,SU,SU,SU, SU,SU, SU,F,F}

Lower and upper approximations

In rough set theory the approximations of a set are introduced to deal with indiscernibility. If S= (U,Ω, V, f) is a decision table, and X ⊆ U, then the I_* lower and I^* upper approximations of X are defined, respectively, as follows:

$$I_*(X) = \{x \in U, I(x) \subseteq X\} \quad (1)$$

$$I^*(X) = \{x \varepsilon U, I(x) \cap X \# \emptyset\} \quad (2)$$

where I(x) denotes the set of all objects indiscernible with x, i.e., equivalence class determined by x. The boundary region of X is the set $BN_I(X) = I^*(X) – I_*(X)$. If the boundary region of X is the empty set, i.e., $BN_I(X) = \emptyset$, then the set X will be called crisp with respect to I; in the opposite case, i.e., if $BN_I(X)\# \emptyset$, the set X will be referred to as rough with respect to I. Vagueness can be characterized numerically by defining the following coefficient:

$$\alpha_I(X) = | I_*(X) | / | I^*(X) | \quad (3)$$

where |X| denotes the cardinality of the set X.

Obviously $0 < \alpha_I(X) <= 1$. If $\alpha_I(X) = 1$ the set X is crisp with respect to I; otherwise if $\alpha_I(X) < 1$, the set X is rough with respect to I. Thus the coefficient $\alpha_I(X)$ can be understood as the accuracy of the concept X.

Rough Membership

A vague concept has boundary-line cases, i.e., elements of the universe which cannot be – with certainty- classified as elements of the concept. Here uncertainty is related to the question of membership of elements to a set. Therefore in order to discuss the problem of uncertainty from the rough set perspective we have to define the membership function related to the rough set concept (the rough membership function). The rough membership function can be defined employing the indiscernibility relation I as:

$$\mu^I_X (x) = | X \cap I(x)| / |I(x)| \quad (4)$$

Obviously, $0 < \alpha_I(X) <=1$. The rough membership function can be used to define the approximations and the boundary regions of a set, as shown below:

$$I_*(X) = \{x \varepsilon U: \mu^I_X (x) = 1\} \quad (5)$$

$$I^*(X) = \{x \varepsilon U: \mu^I_X (x) > 0\} \quad (6)$$

$$BN_I(X)= \{x \varepsilon U: 0< \mu^I_X (x) < 1\} \quad (7)$$

Once can see from the above definitions that there exists a strict connection between vagueness and uncertainty in the rough set theory. As we mentioned above, vagueness is related to sets, while uncertainty is related to elements of sets. Thus approximations are necessary when speaking about vague concepts, whereas rough membership is needed when uncertain data are considered.

Decision Algorithm

Usually we need many classification patterns of objects. For example users can be classified according to web pages, number of searches, etc… Hence we can assume that we have not one, but a family of indiscernibility relations I $=\{I_1,I_2,I_3, \ldots I_n\}$ over the universe U. Set theoretical intersection of equivalence relations $\{I_1,I_2,I_3, \ldots I_n\}$ is denoted by:

$$\cap I = \bigcap_{i=1}^{n} I_i \quad (8)$$

is also an equivalence relation. In this case, elementary sets are equivalence classes of the equivalence relation $\cap$ I.. Because elementary sets uniquely determine our knowledge about the universe, the question arises whether some classification patterns can be removed without changing the family of elementary sets- or in other words, preserving the indiscernibility. Minimal subset I' of I such that will be called a reduct of I. Of course I can have many reducts. Finding reducts is not a very simple task and there are methods to solve this problem. The algorithm we use has been proposed by (Slowinski and Stefanowski 1992), and it is summarized by the following procedure that we name **SSP**: a- Transform continuous values in ranges; b-Eliminate identical attributes; c-Eliminate identical examples; d-Eliminate dispensable attributes; e-Calculate the core of the decision table; f-Determine the reduct set; g- Extract the final set of rules.

Application of Rough set theory to the query: Limbic Functions of the Brain

In table1, users {2,15,161,17,18,7,18} are indiscernible according to the attribute SE=S, users {1,4,5} are indiscernible for the attribute W=L. For example the attribute W generates 4 sets of users: $\{2,11,15,16,17,18\}_S$, $\{3,6,10,12,13,9\}_M$, $\{1,4,5,8\}_L$, and $\{14\}_{VL}$. Because users {2,15,17,18} were SU and user {19} failed, and are indiscernible to attributes W=S, H=S, and SE=S, then the decision variable for SU or F cannot be characterized by W=S, H=S, and SE=S. Hence users {2,15,17,18} and {19} are boundary-line cases. Because user {16} was SU and user {7} has failed, and they are indiscernible to attributes W=S, H=M, and SE=S, then the decision variable for SU or F cannot be characterized by W=S, H=M, and SE=S. Hence users {16} and {7} are boundary-line cases. The remaining users: {3,6,10, 11,12,13,14} have characteristics that enable us to classify them as being SU, while users {1,4,5,8,9,20} display characteristics that enable us to classify them as F, and users {2,7,15,16,17,18,19} cannot be excluded from being SU or F. Thus the lower approximation of the set of being SU is: {3,6,10,11, 12,13,14} and the upper approximation of being SU is:{2,7,15,16,17,18,19,3,6,10,11,12,13,14}. Similarly in the concept of F, its lower approximation is: {1,4,5,8,9,20} and its upper approximation is: {1,4, 5,8,9,20,3,6,10,11,12,13,14}. The boundary region of the set SU or F is still:{2,7,15,16,17,18,19}.The accuracy coefficient of "SU" is (by applying (3)) :

α(SU)=|{3,6,10,11,12,13,14}|/({2,7,15,16,17,18,19,3,6,10, 11,12,13,14}|=7/14=0.5

The accuracy coefficient of "F" is (by applying (3)):

α(F)=|{1,4,5,8,9,20}}|/({2,7,15,16,17,18,19,1,4,5,8,9,20}=6 /13=0.45

We also compute the membership value of each user to the concept of "SU" or "F". By applying (4) we have:

μ_{SU}(1)=|{3,6,10,11,12,13,14,15,16,17,18}∩{1}|=|{1}| =0
μ_{SU}(2)=|{3,6,10,11,12,13,14,15,16,17,18}∩{2,7,15,16,17,18,19}|=|{2,7,15,16,17,18,19}|=4/7
μ_{SU}(3)=|{3,6,10,11,12,13,14,15,16,17,18} ∩ {3}|=|{3} |=1
μ_{SU}(4)=|{3,6,10,11,12,13,14,15,16,17,18} ∩ {4}|=|{4} |=0
μ_{SU}(5)=|{3,6,10,11,12,13,14,15,16,17,18} ∩ {5}|=|{5} |=0
μ_{SU}(6)=|{3,6,10,11,12,13,14,15,16,17,18} ∩ {6}|=|{6} |=1
μ_{SU}(7)=|{3,6,10,11,12,13,14,15,16,17,18}∩{2,7,15,16,17, 18,19}|=|{2,7,15,16,17,18,19} |=4/7
μ_{SU}(8)=|{3,6,10,11,12,13,14,15,16,17,18}∩{1}|=|{8} |=0
μ_{SU}(9)=|{3,6,10,11,12, 13,14,15,16,17,18}∩ {1}|=|{9} |=0
μ_{SU}(10)=|{3,6,10,11,12,13,14,15,16,17,18}∩{10}|=|{10}|=1
μ_{SU}(11)=|{3,6,10,11,12,13,14,15,16,17,18}∩{11}|=|{11}|=1
μ_{SU}(12)=|{3,6,10,11,12,13,14,15,16,17,18}∩{12}|=|{12}|=1
μ_{SU}(13)=|{3,6,10,11,12,13,14,15,16,17,18}∩{13}|=|{13}|=1
μ_{SU}(14)=|{3,6,10,11,12,13,14,15,16,17,18}∩{14}|=|{14}|=1
μ_{SU}(15)=|{3,6,10,11,12,13,14,15,16,17,18}∩{15}|=|{15}|=1
μ_{SU}(16)=|{3,6,10,11,12,13,14,15,16,17,18}∩{16}|=|{16}|=1
μ_{SU}(17)=|{3,6,10,11,12,13,14,15,16,17,18}∩{17}|=|{17}|=1
μ_{SU}(18)=|{3,6,10,11,12,13,14,15,16,17,18}∩{18}|=|{18}|=1
μ_{SU}(19)=|{3,6,10,11,12,13,14,15,16,17,18}∩{2,7,15,16,17, 18,19 }| /|{2,7,15,16,17,18,19}=4/7
μ_{SU}(20)=|{3,6,10,11,12,13,14,15,16,17,18}∩{20}|=|{20}|=0

Applying the **SSP** procedure from steps a-d result in:

Users	Users	W	H	SE	E
19	**19'**	S	S	S	F
7	**7'**	S	M	S	F
20	**20'**	S	S	M	F
1	**1'**	L	M	M	F
8	**8'**	L	L	M	F
9	**9'**	M	S	L	F
5	**5'**	L	L	L	F
4	**4'**	L	L	VL	F
2,15,17,18	**2'**	S	S	S	SU
16	**16'**	S	M	S	SU
10,12	**10'**	M	S	M	SU
11	**11'**	S	M	M	SU
6,13	**6'**	M	M	L	SU
3	**3'**	M	L	L	SU
14	**14'**	VL	M	VL	SU

Table2. Result of applying steps **a-d of SSP.**

The number of users is reduced from 20 to 15 users because of steps **a-d of SSP**. The result of applying of steps e through f is displayed in tables 3 and 4. (X stands for any value)

Users	W	H	SE	E
19'	S	S	S	F
7'	S	M	S	F
20'	S	S	M	F
9'	X	X	L	F
8'	L	X	X	F
1'	L	X	X	F
5'	L	X	X	F
4'	L	X	X	F
2'	S	S	S	SU
16'	S	M	S	SU
11'	S	M	M	SU
10'	X	X	M	SU
6'	M	M	X	SU
3'	M	L	X	SU
14'	VL	X	X	SU

Table 3: Core of the set of final data (**Step e- of SSP**)

Users	W	H	SE	E
19'	S	S	S	F
7'	S	M	S	F
20'	S	S	M	F
9'	X	X	L	F
8'	L	X	X	F
1'	L	X	X	F
5'	L	X	X	F
4'	L	X	X	F
2'	S	S	S	SU
16'	S	M	S	SU
11'	S	M	M	SU
10'	M	S	M	SU
6'	M	M	X	SU
3'	M	L	X	SU
14'	VL	X	X	SU

Table 4: Set of reduct set (Step f- of **SSP**)

Rules extracted:

Contradictory rules:

If (W=S), (H=S), and (SE=S) then User= SU or F.
If (W=S), (H=M), and (SE=S) then User= SU or F.

Rules on Success:

If (W=S), (H=M), and (SE=M) then User= SU
If (W=M), (H=S), and (SE=M) then User= SU

If (W=M), ((H=M) or (H=L)) then User= SU
If (W=VL) then User= SU
Rules on Failure:
If (W=S) and (H=S) and (SE=M) then User= F
If (W=M) and (H=S) and (SE=L) then User= F
If (W=L) then User= F

Contradictory rules also called inconsistent or possible or non-deterministic rules have the same conditions but different decisions, so the proper decision cannot be made by applying this kind of rules. Possible decision rules determine a set of possible decision, which can be made on the basis of given conditions. With every possible decision rule, we will associate a credibility factor of each possible decision suggested by the rule. We propose to define a membership function. Let $\delta(x)$ denote the decision rule associated with object x. We will say that x supports rules $\delta(x)$. Then $C(\delta(x))$ can be denoted by:

$$C(\delta(x))= 1 \text{ if } \mu^{I}_{X}(x) =0 \text{ or } 1.$$

$$C(\delta(x))= \mu^{I}_{X}(x), \text{ if } 0 < \mu^{I}_{X}(x) < 1 \qquad (9)$$

A consistent rule is given a credibility factor of 1, and an inconsistent rule is given a credibility factor smaller than 1 but not equal to 0. The closer it is to one the more credible the rule is. The credibility factor of both inconsistent rules is 4/7 >.5 which makes more credible than incredible (being equal to 0).

ID3

ID3 uses a tree representation for concepts (Quinlan, 1983). To classify a set of instances, we start at the top of the tree. And answer the questions associated with the nodes in the tree until we reach a leaf node where the classification or decision is stored. ID3 starts by choosing a random subset of the training instances. This subset is called the window. The procedure builds a decision tree that correctly classifies all instances in the window. The tree is then tested on the training instances outside the window. If all the instances are classified correctly, then the procedure halts. Otherwise, it adds some of the instances incorrectly classified to the window and repeats the process. This iterative strategy is empirically more efficient than considering all instances at once. In building a decision tree, ID3 selects the feature which minimizes the entropy function and thus best discriminates among the training instances.

The ID3 Algorithm:

1. Select a random subset W from the training set.
2. Build a decision tree for the current window:
    a. Select the best feature which minimizes the entropy function H:
    $$H = \sum_{i} -pi \text{ lop } pi \qquad (10)$$
    Where pi is the probability associated with the ith class. For a feature the entropy is calculated for each value. The sum of the entropy weighted by the probability of each value is the entropy for that feature.
    b. Categorize training instances into subsets by this feature.
    c. Repeat this process recursively until each subset contains instances of one kind or some statistical criterion is satisfied.
3. Scan the entire training set for exceptions to the decision tree.
4. If exceptions are found, insert some of them into W and repeat from step 2. The insertion may be done either by replacing some of the existing instances in the window or by augmenting it with new exceptions**.**

Rules extracted (See figure 1)**:**
Contradictory rules:
If (W=S), (H=S), and (SE=S) then User= SU or F.
If (W=S), (H=M), and (SE=S) then User= SU or F.
Rules on Success:
If (W=VL) then User= SU
If (W=M), (H=S), and (SE=M) then User= SU
If (W=M) and ((H=L) or (H=L)) then User= SU
If (W=S) and (SE=M) and (H=M) then User= SU
Rules on Failure:
If (W=L) then User= F
If (W=M) and (H=S) and (SE=L) then User= F
If (W=S) and (H=S) and (SE=M) then User= F

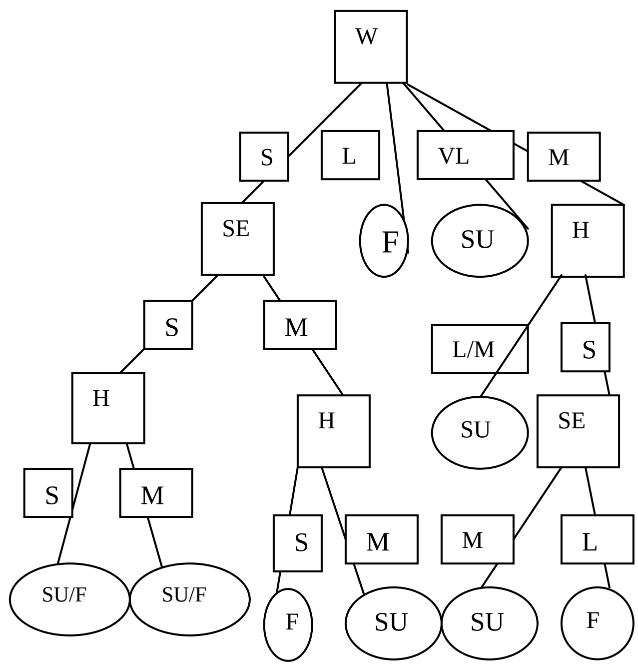

Fig1.Rules extracted by ID3.

It seems that the 9 rules extracted by ID3 are the same extracted by Rough set theory. The 3 parameters were not enough to separate these cases between success and failure.

Conclusion

This application of the rough set methodology shows the suitability of the approach for the analysis of user's web information system's behavior. Rough set theory was never applied on user's behavior and the latter was analyzed very little (Meghabghab, 2003) considering the amount of emphasis on understanding user's web behavior (Lazonder et al., 200). Moreover, we show in this paper how using even a small part of rough set theory can produce interesting results for web behavior situations: a- The proposed rules provide a good classification of user's behavior except in the case of contradictory rules where the 3 attributes are not enough to distinguish between the users; b- The information system was reduced from 20 users at one time to 15 and then 9 rules were extracted that cover all cases of user's behavior. Information theoretic classification measure has been around for a long time and applied in many areas (Quinlan 1983). But it does not mine the relations between attributes, the vagueness that is existent in the attribute set that rough set theory does. It just provides a simple set of accurate classification rules. The rough set theory approach is very rich in interpretation and can help understand complex relations in any decision environment.

References

1. Lazonder, A. W., Biemans, J. A.; and Wopereis, G. J. H. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science*, 51 (6): 576-581.
2. Meghabghab, G. 2003. The Difference between 2 Multidimensional Fuzzy Bags: A New Perspective on Comparing Successful and Unsuccessful User's Web Behavior. *LCNS*: 103-110.
3. Pawlak, Z. 1982. Rough sets. *International Journal of Computer and Information Science* 11: 341–356.
4. Pawlak, Z. 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data.* Dordrecht: Kluwer Academic.
5. Pawlak, Z. 1997. Rough sets approach to knowledge-based decision support. *European Journal of Operational Research* 99: 48–59.
6. Pawlak, Z. 2002. Rough sets and intelligent data analysis. *Information Sciences* 147: 1–12.
7. Pawlak, Z., and Slowinski, R. 1994. Decision analysis using rough sets. *International Transactions in Operational Research* 1(1): 107–114.
8. Quinlan, J.R. 1983. Learning efficient classification procedures and their application to chess and games. In Machine Learning, Michalski, R.S., Carbonell, J.G., and Mitchell T.M., (Eds), Tioga Publishing Company, Palo Alto, CA.
9. Slowinski, R. and Stefanowski, J. 1992. RoughDAS and Rough-class software implementation of the rough sets approach, in Intelligent Decision Support—Handbook of Applications and Advantages of the Rough Sets Theory, R. Slowinski (ed.), 445-456. Dordrecht: Kluwer Academic.
10. Yao Y.Y., 2003. Probabilistic approaches to rough sets. *Expert System*, 20(5): 287-297.