SE 63
/?? 78
p. 15

N93-22359

CERTAIN & POSSIBLE RULES FOR DECISION MAKING USING ROUGH SET THEORY EXTENDED TO FUZZY SETS

André de Korvin

Department of Applied Mathematical Sciences
University of Houston-Downtown

Margaret F. Shipley

Department of Business Management
University of Houston-Downtown
One Main Street
Houston, Texas 77002

(713) 221- 8571
FAX (713) 221- 8632

1. Introduction

Uncertainty may be caused by the ambiguity in the terms used to describe a specific situation. It may also be caused by skepticism of rules used to describe a course of action or by missing and/or erroneous data. [For a small sample of work done in the area, the reader is referred to (Arciszewski & Ziarko 1986), (Bobrow, et.al. 1986), (Wiederhold, et. al. 1986), (Yager 1984), and (Zadeh 1983).]

To deal with uncertainty, techniques other than classical logic need to be developed. Although, statistics may be the best tool available for handling likelihood, it is not always adequate for dealing with knowledge acquisition under uncertainty. [We refer the reader to Mamdani, et. al. (1985) for a study of the limitations of traditional statistical methods.]

Inadequacies caused by estimating probabilities in statistical processes can be alleviated through use of the Dempster-Shafer theory of evidence. [For a sample of works using the Dempster-Shafer theory see (Shafer 1976), (de Korvin, et. al. 1990), (Kleyle & de Korvin 1989), (Strat 1990), and (Yager).] Fuzzy set theory is another tool used to deal with uncertainty where ambiguous terms are present. [Articles in (Zadeh 1979, 1981 & 1983) illustrate the numerous works carried out in fuzzy sets.] Other methods include rough sets, the theory of endorsements and nonmonotonic logic. [The work on rough sets is illustrated in (Fibak, et. al. 1986), (Grzymala-Busse 1988), and (Mrozek 1985 & 1987). Also, see (Mrozek 1985) and (Pawlak 1982) for the application of rough sets to medicine and (Arciszewski & Ziarko 1986) and (Pawlak 1981) for applications to industry.]

J. Grzymala-Busse (1988) has defined the concept of lower and upper approximation of a (crisp) set and has used that concept to extract rules from a set of examples. We will define the fuzzy analogs of lower and upper approximations and use these to obtain certain and possible rules from a set of examples where the data is fuzzy. Central to these concepts will be the idea of the degree to which a fuzzy set A is contained in another fuzzy set B, and the degree of intersection of set A with set B. These concepts will also give meaning to the statement; A implies B. The two meanings will be: 1) if x is certainly in A then it is certainly in B, and 2) if x is possibly in A then it is possibly in B. Next, classification will be looked at and it will be shown that if a classification is well externally definable then it is well internally definable, and if it is poorly externally definable then it is poorly internally definable, thus generalizing a result of Grzymala-Busse (1988). Finally, some ideas of how to define consensus and group opinions to form clusters of rules will be given.

2. Results

We now recall some basic definitions such as lower and upper approximations and the concept of an information system.

Let U be the universe. Let R be an equivalence relation on U. Let X be any subset of U. If [x] denotes the equivalence class of x relative to R, then we define

$\underline{R}(X) = \{x \in U/[x] \subset X\}$ and

$\overline{R}(X) = \{x \in U/[x] \cap X \neq \emptyset\}$.

$\underline{R}(X)$ is called the lower approximation of X and $\overline{R}(X)$ is called an upper approximation of X. Then $\underline{R}(X) \subset X \subset \overline{R}(X)$. If $\underline{R}(X) = X = \overline{R}(X)$, then X is called definable.

An information system is a quadruple (U,Q,V,τ) where U is the universe and Q is a subset of $C \cup D$ where $C \cap D = \emptyset$. The set C is called the set of conditions; D is called the set of decisions. We assume here that Q = C. The set V stands for value and τ is a function from UxQ into V where $\tau(u,q)$ denotes the value of attribute q for element u. The set C induces naturally an equivalence on U by partitioning U into sets over which all attributes are constant. The set X is called roughly C-definable if

$\underline{R}(X) \neq \emptyset$ and $\overline{R}(X) \neq U$.

It will be called internally C-undefinable if

$\underline{R}(X) = \emptyset$ and $\overline{R}(X) \neq U$.

It will be called externally C-undefinable if

$\underline{R}(X) \neq \emptyset$ and $\overline{R}(X) = U$.

Fuzzy sets defined

Next, we define two functions on pairs of fuzzy sets that will be of importance in the present work.

$$I(A \subset B) = \inf_x \mathrm{Max}\ \{1 - A(x),\ B(x)\} \tag{1}$$

$$J(A \# B) = \mathrm{Max}_x\ \mathrm{Min}\ \{A(x),\ B(x)\}. \tag{2}$$

Here A and B denote fuzzy subsets of the same universe. The function $I(A \subset B)$ measures the degree to which A is included in B and $J(A \# B)$ measures the degree to which A intersects B. It is important to note that for the crisp case, $I(A \subset B) = 1$ iff $A \subset B$ and is 0 otherwise. Similarly, $J(A \# B) = 1$ iff $A \cap B \neq \emptyset$.

The goal is to define the fuzzy terms involved in the decision as a function of the terms used in the conditions. This is accomplished as a function of how much the decision follows the conditions. Let $\{B_i\}$ be a finite family of fuzzy sets. Let A be a fuzzy set. By a lower approximation of A through $\{B_i\}$, we mean the fuzzy set

$$\underline{R}\ (A) = \bigcup_i\ I\ (\ B_i \subset A\ )\ B_i \tag{6}$$

The decision making process may be simplified by disregarding all sets B, if $I\ (\ B_i \subset A\ )$ is less than some threshold α. Then,

$$\underline{R}\ (A)_{\alpha} = \bigcup_i I\ (\ B_i \subset A\ )\ B_i \tag{7}$$

over all B_i for which $I\ (\ B_i \subset A\ ) \geq\ \alpha$.

Similarly, we can define the upper approximation of A through $\{B_i\}$ as

$$\overline{R}\ (A)_{\alpha} = \bigcup_i J\ (\ B_i \# A\ )\ B_i \tag{8}$$

over all B_i for which $J\ (\ B_i \# A\ ) \geq\ \alpha$.

The operators I and J will yield two possible sets of rules: the certain rules and the possible rules. It is straightforward to see that if $\{B_i\}$ are crisp equivalency classes we get the lower and upper approximations as defined by Grzymala-Busse (1988).

Determining Fuzzy Rules

We now show how rules can be obtained from the raw data given in Table 1 after converting this data according to the professor's evaluation of the performance of the students, relative to exams high, exams low, project high, project low, and his belief with respect to each student getting an A. (See Table 2 for the converted data.)

Table 1: Production/Operations Management Grades

Student	Exams (2)	Project (Written & Oral)	Course Grade
1	75	85	75.36
2	94	87	89.53
3	88	89.3	89.93
4	79.5	95	78.06
5	85	97	90.85
6	56.5	88.6	60.89
7	65	91.6	76.15
8	49	76.7	59.22
9	63.5	89.1	69.99
10	57	76.9	55.77
11	70	98	80.3
12	93	88	90.1

It can be observed that none of the course grades was a strong predictor of "success". In other words, the course grades of 90 or slightly better than 90 as a "quality" measure of the final product did not allow the professor strong belief in the awarding of an "A" to the student. The professor's belief in these grades being the best in the class and therefore deserving of an "A" grade was approximately .67. The belief in the lower scores is scaled downward from .67 to .41 (the latter representing belief that 55.77 will be the top score in the class.)

The professor recognized the high exam scores of 94 and 93, with belief of .99/EH and .98/EH, respectively (EH: Exams High). The low exam score of .49 was designated .92/EL (EL: Exams Low) by the professor. Since all project grades were relatively close and relatively high, the professor saw little differentiation between the "top" score and the other scores. The "top" project score is .54 high and .46 low. (.54/PH and .46/PL, respectively) This contrasts with the worse project score being .43/EH and .59/EL, where .59 is the highest belief that a project grade is a "low" score. This approach was considered to be consistent since although exam grades varied from 49 to 94, no project grade was below a 76.7. It was felt that keeping the project grades from being too strongly biased toward "high" would prevent the decision rules from being overly biased toward high project grades. Enough differentiation was considered to allow the rough set formulation to consider both attributes in the decision rules for awarding a "top" score of "A" to a student. Each student's scores were translated into belief with respect to EH, EL, PH, PL and "A".

For our example of twelve POM students, x_1, $x_2, \ldots, x_{12}$, we let EH:exams high PH:project high

EL:exams low PL:project low "A": Top Grade

Thus, for the first student, x_1, the belief that the exams were high is .79/EH, and that the exams were low is .60/EL; that the project grade was high is .47/PH and that it was low is .53/PL. The strength of belief for an A is .56/"A". In addition, EH may be viewed as a fuzzy set of students, such that EH = $.79/x_1 + .99/x_2 + \ldots + .98/x_{12}$, where x_2 is an excellent example of EH (.99) while x_8 is not such a good example (.52). (See Table 2 below for all the professor's evaluative scores.)

Table 2: Professor's Evaluative Scores

Student	EH	EL	PH	PL	"A"
1	.79	.60	.47	.53	.56
2	.99	.48	.48	.52	.66
3	.93	.51	.50	.50	.67
4	.84	.57	.53	.47	.58
5	.89	.53	.54	.46	.67
6	.58	.81	.49	.51	.45
7	.68	.69	.51	.49	.56
8	.52	.92	.43	.58	.44
9	.67	.71	.50	.51	.52
10	.60	.79	.43	.59	.41
11	.74	.64	.54	.46	.59
12	.98	.48	.49	.51	.67

Using our rough set theory formulas as they have been developed for fuzzy systems of attributes and decisions, we compute:

$I(EH \subset "A") = .41$ $\quad$ $I(EH \cap PH \subset "A") = .51$

$I(EL \subset "A") = .41$ $\quad$ $I(EH \cap PL \subset "A") = .42$

$I(PH \subset "A") = .51$ $\quad$ $I(EL \cap PH \subset "A") = .51$

$I(PL \subset "A") = .42$ $\quad$ $I(EL \cap PL \subset "A") = .42$

with a lower approximation for $\alpha = .50$ defined by:

$\underline{R} = .51\ PH \cup .51\ (EH \cap PH)$.

The extracted rules would imply that high project scores and high exam scores both impact a high course grade with certainty .51.

Possibility rules can be determined by computing:

J(EH #"A") = .67   J(EH ∩ PH # "A") = .54

J(EL # A") = .59   J(EH ∩ PL # "A") = .53

J(PH #"A") = .54   J(EL ∩ PH # "A") = .54

J(PL #"A") = .53   J(EL ∩ PL # "A") = .53

with an upper approximation at $\alpha = .60$ defined as:

$\overline{R} = .67\ EH$.

Thus, we can see that the factors dictating the "best" in the class are:

1) If project grades are high, an "A" score will be attained. (Certainty = .51)
2) If project grades and exam grades are high, an "A" score will be attained. (Certainty = .51)
3) If exam grades are high, an "A" score will be attained. (Possibility = .67)

Indeed, these rules reflect the fact that exam grades are more heavily weighted than the project grade toward determining the final course grade. Additionally, these two grades comprise the majority of the weighted scores from which the course grade is calculated.

Belief & Possibility

We can use the functions I and J to determine two meanings of A implies B. The belief that if x is certainly in A then it is certainly in B is given by:

$$I[\ \underline{R}\ (A) \subset \ \underline{R}\ (B)] \tag{9}$$

and the belief that if x is possibly in A then it is possibly in B can be defined by:

$$J[\ \overline{R}\ (A)\ \#\ \ \overline{R}\ (B)\ ] \tag{10}$$

This interpretation follows from the fact that $\underline{R}(A)$ are objects certainly in A and $\overline{R}(A)$ are objects possibly in A. We now turn to the study of classifications.

Classifications

The study of classifications is of great interest because in learning from examples, the rules are derived from classifications generated by simple decisions. In this section, we turn our attention to classifications. Of course, the traditional meaning is to partition. In our setting, we have ill-defined boundaries, so we need to relax the concept of partitions by requiring that the sets not overlap too much.

As earlier, consider a finite family of fuzzy sets, $\{B_i\}$. Let π denote a finite family of fuzzy sets

$$\pi = \{A_1,\ A_2,\ \ldots,\ A_n\}$$

We define

$$\underline{P}\pi_\alpha = \{\ \underline{R}(A_1)_\alpha,\ \ldots,\ \underline{R}(A_n)_\alpha\},$$

$$\overline{P}\pi_\alpha = \{\ \overline{R}(A_1)_\alpha,\ \ldots,\ \overline{R}(A_n)_\alpha\}$$

where the lower and upper α-approximations are generated by the finite sequence $\{B_i\}$.

We can develop the following relationship:

$d^\circ[A = B] = \text{Min}\ \{\ I\ (A \subset B),\ I(\ B \subset A)\}$

using the following definitions:

$$d^\circ[\underline{P}\pi_\alpha = \pi] = \underset{k}{\text{Min}}\ \{d^\circ[\underline{R}(A_k)_\alpha = A_k]\}$$

$$d^\circ[\overline{P}\pi_\alpha = \pi] = \underset{k}{\text{Min}}(d^\circ[\overline{R}(A_k)_\alpha = A_k])$$

π will be called $\{B_i\}$ definable to the degree β with threshold α if

$$\text{Min}\ \{\ d^\circ[\underline{P}\pi_\alpha = \pi],\ d^\circ[\overline{P}\pi_\alpha = \pi]\} \geq \beta.$$

If we define

$$d^\circ[\underline{P}\pi_\alpha = \overline{P}\pi_\alpha] = \underset{k}{\text{Min}}\ \{\ d^\circ[\underline{R}(A_k)_\alpha = \overline{R}(A_k)_\alpha]\},$$

it can be shown that if $\beta \geq \frac{1}{2}$, then

$d^\circ[\underline{P}\pi_\alpha = \pi] \geq \beta$ and $d^\circ[\overline{P}\pi_\alpha = \pi] \geq \beta$ imply that

$d^\circ[\underline{P}\pi_\alpha = \overline{P}\pi_\alpha] \geq \beta$.

Recall that the following result is shown in information systems. For classifications, if $\overline{P}A_k$ is the universal set for each k, then $\underline{P}A_k$ is empty for each k. Also, if $\underline{P}A_k$ is nonempty for each k, the $\overline{P}A_k$ is not the universal set for any value of k. We would like to get the analog of this by showing if $\underline{R}(A_k)_\alpha$ "has some substance" for some k, then $\overline{R}(A_j)_\alpha$ for $j \neq k$ is "not too large", and if $\overline{R}(A_k)_\alpha$ is "fairly substantial", $\underline{R}(A_j)_\alpha$ for $j \neq k$ cannot be "too large". In this sense, the results of Grzymala-Busse (1988) will be generalized.

We would like $\{A_k\}$ and $\{B_i\}$ to somewhat approximate a partition. We define the following two conditions:

(*) For every $0 < \epsilon < 1$, there exists $0 < \delta < 1$ such that if $B_i(x_0) > \epsilon$, then $B_\ell(x_0) < 1 - \delta$ for $\ell \neq i$.

(**) For every pair j,k with $j \neq k$ and all x, $A_k(x) + A_j(x) \leq 1$.

Conditions (*) and (**) both express that the overlap is not too large and obviously hold for partitions. We note that if (**) holds for $\{B_i\}$ then it implies (*). Indeed, in this case we pick $\delta = \epsilon$. Thus, the results that follow may be shown assuming condition (**) for $\{B_i\}$ and $\{A_k\}$.

We first show that under conditions (*) and (**), whenever $\underline{R}(A_k)_\alpha$ is bounded away from 0, then $\overline{R}(A_j)_\alpha$ for $j \neq k$ is bounded away from 1. Suppose $\underline{R}(A_k)_\alpha(x_0) > \epsilon$, then for some i, $I(B_i \subset A_k) > \epsilon$ and $B_i(x_0) > \epsilon$, so for $\ell \neq i$ from condition (*), we have $B_\ell(x_0) < 1 - \delta$. For any $\ell \neq i$ we have $J(B_\ell \# A_j) B_\ell(x_0) < 1 - \delta$. Now

$$J(B_i \# A_j) = 1 - I(B_i \subset \neg A_j);$$

$$I(B_i \subset A_k) = \operatorname*{Min}_x \operatorname{Max} \{1-B_i(x), A_k(x)\};$$

$$I(B_i \subset \neg A_j) = \operatorname*{Min}_x \operatorname{Max} \{1-B_i(x), 1-A_j(x)\}.$$

Condition (**) implies $I(B_i \subset A_k) \leq I(B_i \subset \neg A_j)$ for all $j \neq k$. From the above it follows that $J(B_i \# A_j) < 1 - \epsilon$. Thus,

$$\overline{R}(A_j)_\alpha(x_0) < \operatorname{Max} \{ 1-\epsilon, 1-\delta\}.$$

We now show a rough converse to the above. If $\overline{R}(A_k)$ is bounded away from 0, then for $j \neq k$, $\underline{R}(A_j)_\alpha$ is bounded away from 1. Suppose $\overline{R}(A_k)_\alpha(x_0) > 1 - \epsilon$ for some k, then

$$J(B_{i_0} \# A_k) B_{i_0}(x_0) > 1 - \epsilon \text{ for some } i_0.$$

Pick $j \neq k$. Then

$$I(B_{i_0} \subset A_j) = 1 - J(B_{i_0} \# \neg A_j).$$

Now, $J(B_{i_0} \# \neg A_j) = \operatorname*{Max}_x \operatorname{Min} \{B_{i_0}(x), 1 - A_j(x)\};$

$$J(B_{i_0} \# A_k) = \operatorname*{Max}_x \operatorname{Min} \{B_{i_0}(x), A_k(x)\}.$$

By (**) it follows that $J(B_{i_0} \# \neg A_j) \geq J(B_{i_0} \# A_k)$.

From above, $I(B_{i_0} \subset A_j) \leq 1 - J(B_{i_0} \# A_k) \leq \epsilon$.

Since $B_{i_0}(x_0) > 1 - \epsilon$, by (*), $B_i(x_0) < \theta$ for $i \neq i_0$ where $0<\theta<1$.

Therefore, $\underline{R}(A_j)_\alpha(x_0) \leq \text{Max} \{ \epsilon, \theta \}$.

Consensus

We can define consensus between two rows of a table by

Consensus $[Row_i, Row_j] = \text{Min} \{ I[Row_i \subset Row_j], I[Row_j \subset Row_i] \}$

Here, Row_i and Row_j are considered to be fuzzy subsets of the set of all attributes and decisions. If γ is some predetermined threshold, we pick some x_1 and then all x_j for which Consensus $[Row_1, Row_j] \geq \gamma$. If any of the x's are left over, we start again with the first x available. We thus get fuzzy sets $S_1, S_2, \ldots, S_\ell$ where $\mu_{S_i}(\ell_i) = 1$ for some ℓ_i (which we might call the leader of S_i) and $\mu_{S_i}(x)$ = Consensus (l_i, x) provided $\mu_{S_i}(x)$ exceeds γ. Within each S_i we then can recompute the symptoms/decisions for x_j taking $\mu_{S_i}(x_j)$ into account If $1 \leq i \leq \ell$, then we have ℓ (aggregated) decisions and using fuzzy cardinality we can compute the "firing strength" of each block of rules. This approach has the advantage of taking consensus of opinions into consideration in the decision. The detailed methodology will be discussed in a later paper.

REFERENCES

Arciszewski, T. and Ziarko, W. 1986. "Adaptive expert system for preliminary engineering design," Proceedings 6th International Workshop on Expert Systems and their Applications. Avignon, France, 1, 696-712 .

Bobrow, D.G., Mittal, S. and Stefik, M.J. 1986. "Expert systems perils and promises" Communications of the ACM, 29, 880-894.

Cheeseman, P. 1986. "Induction of models under uncertainty," Proceedings of ACM SIGART International Symposium on

Methodologies for Intelligent Systems, Knoxville, Tennessee, 130-144.

Fibak, J., Slowinski, K. and Slowinski, R. 1986. "The application of rough set theory to the verification of indications for treatment of duodenal ulcers by HSV," Proceedings 6th International Workshop on Expert Systems and their Applications, Avignon, France, 1, 587-594.

Grzymala-Busse, J.W. 1988. "Knowledge acquisition under uncertainty: a rough set approach," Journal of Intelligent and Robotic Systems 1, 3-16.

Kleyle, R. and de Korvin, A. 1989. "A unified model for data acquisition and decision making," The Journal of the American Society for Information Science 15, 149-161.

de Korvin, A., Kleyle, R. and Lea, R. 1990. "An evidential approach to problem solving when a large number of knowledge systems are available," The International Journal of Intelligent Systems, 5, 293-306.

Mamdani, A., Efstathiou, J. and Pang, D. 1985. "Inference under uncertain expert systems 85," Proceedings Fifth Technical Conference British Computer Society, Specialist Group on Expert Systems, 181-194.

Mrozek, A. "Information systems and control algorithms", 1985. Bulletin Polish Academy of Science, Technical Science, 33, 195-204.

Mrozek, A. 1987. "Rough sets and some aspects of expert systems realization," Proceedings 7th International Workshop on Expert Systems and their Applications, Avignon, France, 597-611.

Pawlak, Z. 1981. "Rough sets. Basic Notions," Institute Computer Science, Polish Academy of Science Report No. 431, Warsaw.

Pawlak, Z. 1981. "Classification of objects by means of attributes," Institute Computer Science Polish Academy Science Report No. 429, Warsaw.

Pawlak, Z. 1982. "Rough sets," International Journal of Information Computer Science ,11, 341-356.

Pawlak, Z. 1983. "Rough classifications," International Journal of Man-Machine Studies, 20, 469-483.

Pawlak, Z. 1985. "Rough sets and fuzzy sets," Fuzzy Sets and Systems, 17, 99-102.

Shafer, G. 1976. Mathematical Theory of Evidence, Princeton University Press.

Strat, T.M. 1990. "Decision analysis using belief functions," International Journal of Approximate Reasoning, 4, 391-417.

Wiederhold, G.C., Walker, M., Blum, R., and Downs, S. 1986. "Acquisition of knowledge from data," Proceedings ACM SIGART International Symposium on Methodologies for Intelligent Systems, Knoxville, Tennessee, 78-84.

Yager, R.R. 1984. "Approximate reasoning as a basis for rule based expert systems," IEEE Transactions on Systems, Man and Cybernetics, 14, 636-643.

Yager, R.R. "Decision making under Dempster-Shafer uncertainties," Iona College, Machine Intelligence Institute Tech. Report MII-915.

Zadeh, L.A. 1983. "The rule of fuzzy logic in the management of uncertainty in expert systems," Fuzzy Sets and Systems 11, 119-227.

Zadeh, L.A. 1979. "Fuzzy sets and information granularity," Advances in Fuzzy Set Theory and Applications, 3-18.

Zadeh, L.A. 1981. "Possibility theory and soft data analysis," Mathematical Frontiers of the Social and Policy Sciences. Eds. L. Cobb and R.M. Thrall, 69-129. Westview Press, Boulder, Colorado.