A DECISION-THEORETIC ROUGH SET MODEL

Y.Y. Yao, S.K.M. Wong, and Pawan Lingras

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada, S4S 0A2

Abstract In this paper we propose a generalized rough set model using decision-theoretic approach. Within this model, the lower and upper approximations of a given concept are derived from a Bayesian decision procedure such that the expected risk is minimum. It is explicitly shown that both the algebraic and probabilistic rough set models are special cases of the proposed model. This new formulation not only contributes to a deeper understanding of the rough set theory but also broadens the framework of rough sets for further development of intelligent information systems.

1. Introduction

Representation and management of uncertain knowledge are some of the more difficult tasks in the design and implementation of intelligent information systems. The uncertainty may stem from imprecise and incomplete information about the reality. The notion of rough sets (Pawlak, 1984, 1982) was introduced to provide a systematic framework for the study of intelligent systems characterized by insufficient information. The successful applications of the rough set model in a variety of problems (Kowalczyk, & Szymanski, 1989; Srinivasan, 1989; Pettorossi, Ras, & Zemankova, 1987; Wong, Ziarko, & Ye, 1987; Pawlak, Slowinski, & Slowinski, 1986) have amply demonstrated its usefulness and versatility.

The theory of rough sets is an extension of the ordinary set theory, in which a *concept* is described by a pair of ordinary sets called the lower and upper approximations. That is, the lower and upper approximations specify the range within which the given concept lies. However, there are inherent drawbacks (Pawlak, Wong, & Ziarko, 1988; Wong, Ziarko, & Ye, 1987) in this *algebraic* approach for the characterization of a concept. The algebraic approach seems to ignore the available statistical information which may be crucial to many non-deterministic classification problems. For this reason, a *probabilistic* rough set model (Pawlak, Wong, & Ziarko, 1988; Wong, & Ziarko, 1987, 1986) was proposed to take advantage of the available statistical information.

We show in this paper that a generalized rough set model can be developed directly from the Bayesian decision-theoretic approach. The main advantage of this new formulation is that it not only contributes to a deeper understanding of the rough set theory but also broadens the framework of rough sets for further development of intelligent information systems.

We first summarize some of the key notions about rough sets in Section 2. Then a generalized model is developed in Section 3. We show that the existing algebraic and probabilistic rough set models are indeed special cases of the proposed model.

Methodologies for Intelligent Systems, 5
Z.W. Ras, M. Zemankova, and M.L. Emrich, Editors

2. The Algebraic Rough Set Model

For completeness, we briefly review here the *algebraic* rough set model (Pawlak, 1984, 1982).

Let U denote the universe (a finite ordinary set), and let $R \subseteq U \times U$ be an equivalence (indiscernibility) relation on U. The pair $A = (U, R)$ is called an approximation space. The equivalence relation R partitions the set U into disjoint subsets. Such a partition of the universe is denoted by $U/R = \{E_1, E_2, \ldots, E_n\}$, where E_i is an equivalence class of R. If two elements e_1, e_2 in U belong to the same equivalence class $E \in U/R$, we say that e_1 and e_2 are indistinguishable. The equivalence classes of R and the empty set $\varnothing$ are called the elementary or atomic sets in the approximation space $A = (U, R)$. The union of one or more elementary sets is called a composed set in A. $Com(A)$ denotes the family of all composed sets.

Since it is not possible to differentiate the elements within the same equivalence class, one may not be able to obtain a precise representation for an arbitrary set $X \subseteq U$ in terms of elementary sets in A. Instead, any X may be represented by its lower and upper approximations defined as follows:

$$\underline{A}(X) = \bigcup_{E_i \subseteq X} E_i \ , \qquad \overline{A}(X) = \bigcup_{E_i \cap X \neq \varnothing} E_i \ . \tag{2.1}$$

That is, the lower approximation $\underline{A}(X)$ is the union of all the elementary sets which are subsets of X, and the upper approximation $\overline{A}(X)$ is the union of all the elementary sets which have a non-empty intersection with X. The pair $(\underline{A}(X), \overline{A}(X))$ is the representation of an ordinary set X in the approximation space $A = (U, R)$, or simply the rough set of X.

The family $Com(A)$ of all composed sets uniquely defines a topological space $T_A = (U, Com(A))$. $Com(A)$ consists of all the open and closed sets in T_A. By definition, for any set $X \subseteq U$, $\underline{A}(X)$ is the greatest composed set contained in X and $\overline{A}(X)$ is the least composed set containing X. Thus, the lower and upper approximations of an arbitrary set X defined by eqn. (2.1) can be interpreted as the interior and closure operations in the topological space T_A. For any subsets $X, Y \subseteq U$:

(I1) $\underline{A}(X \cap Y) = \underline{A}(X) \cap \underline{A}(Y)$

(I2) $\underline{A}(X \cup Y) \supseteq \underline{A}(X) \cup \underline{A}(Y)$

(I3) $\underline{A}(\neg X) = \neg \overline{A}(X)$

(I4) $\overline{A}(X \cap Y) \subseteq \overline{A}(X) \cap \overline{A}(Y)$

(I5) $\overline{A}(X \cup Y) = \overline{A}(X) \cup \overline{A}(Y)$

(I6) $\overline{A}(\neg X) = \neg \underline{A}(X)$

where $\neg X$ denotes the complement of X, i.e., $\neg X = U - X$.

Based on the lower and upper approximations of a set $X \subseteq U$, the universe U can be divided into three disjoint regions, the positive region $POS_A(X)$, the doubtful region $DOU_A(X)$, and the negative region $NEG_A(X)$:

$$POS_A(X) = \underline{A}(X) = \bigcup_{E_i \subseteq X} E_i \ ,$$
$$DOU_A(X) = \overline{A}(X) - \underline{A}(X) = \bigcup_{\substack{E_i \not\subseteq X \\ E_i \cap X \neq \varnothing}} E_i \ , \tag{2.2}$$
$$NEG_A(X) = U - \overline{A}(X) = \bigcup_{E_i \cap X = \varnothing} E_i \ .$$

Note that the upper approximation of a set X is the union of the positive and doubtful regions, namely, $\overline{A}(X) = POS_A(X) \cup DOU_A(X)$. One can say with *certainty* that any element $e \in POS_A(X)$ belongs to X, and that any element $e \in NEG_A(X)$ does not belong to X. However, one can not decide with certainty whether or not an element $e \in DOU_A(X)$ belongs to X. From properties (I3) and (I6), it follows for any set $X \subseteq U$:

(I7) $POS_A(X) = NEG_A(\neg X)$,

(I8) $DOU_A(X) = DOU_A(\neg X)$,

(I9) $NEG_A(X) = POS_A(\neg X)$.

Thus the two rough sets $(\underline{A}(X), \overline{A}(X))$ and $(\underline{A}(\neg X), \overline{A}(\neg X))$ complement each other.

3. A Generalized Rough Set Model

Although the algebraic rough set model has been successfully applied in machine learning, expert system design, and knowledge representation (Kowalczyk, & Szymanski, 1989; Pettorossi, Ras, & Zemankova, 1987; Pawlak, Slowinski, & Slowinski, 1986; Wong, & Ziarko, 1986), it may be inadequate to deal with situations in which the statistical information plays an important role. Consider, for example, two equivalence classes E_1 and E_2 in the partition U/R such that each has 100 elements. Suppose only a single element in E_1 belongs to X, and only a single element in E_2 does not belong to X. In the algebraic rough set model, these two equivalence classes are treated in the same way and both will be included in the doubtful region. From a physical point of view, such an identical treatment of E_1 and E_2 does not seem reasonable. Moreover, the observation that only one element in E_1 belongs to X may be just a result of noise. Therefore, the algebraic rough set model can be sensitive to noise often encountered in many real-world applications. The probabilistic rough set model (Pawlak, Wong & Ziarko, 1988; Wong, & Ziarko, 1987, 1986) was introduced to overcome these shortcomings by incorporating the available statistical information. In what follows, we develop a generalized rough set model based on the Bayesian decision theory. We also show that our model indeed provides a unified view of the existing rough set models.

Consider a problem of deciding whether an object $e \in U$ with description E belongs to a set $X \subseteq U$, where X represents a given concept. Generally one would like to answer this question with a simple *yes* or *no*. However, in many situations the cost of making a wrong decision may be high. In these situations, one may provide the answer *doubtful.* The answer *doubtful* indicates that the available information is insufficient to make any definite conclusions. We can state this decision problem more formally as follows. Let R be an equivalence relation on a universe of discourse U, and P is a probability measure defined on a σ-algebra of subsets of U. Let $U/R = \{E_1, E_2, \ldots, E_n\}$ denote the family of equivalence classes induced by the relation R. Each equivalence class $E \in U/R$ consists of all the objects with a certain description. Consider a set $X \subseteq U$ representing a concept of interest. Suppose all the conditional probabilities $P(X \mid E)$ are known. Let POS_{A_P} be the set of objects corresponding to the answer *yes*, DOU_{A_P} to the answer *doubtful*, and NEG_{A_P} to the answer *no*, where the triple $A_P = (U, R, P)$ represents the available information. Now, the decision problem reduces to dividing the universe U into three disjoint regions POS_{A_P}, DOU_{A_P}, and NEG_{A_P}. There are many ways to construct these three regions. The approach we adopt in this paper is based on the Bayesian decision theory (Duda, & Hart, 1973; Winkler, 1972; DeGroot, 1970).

An element $e \in E$ can be classified into one of the three regions, POS_{A_P}, DOU_{A_P}, and NEG_{A_P}. In other words, corresponding to these regions one can take one of the following actions:

$$a_1\text{: decide } POS_{A_P}; \quad a_2\text{: decide } DOU_{A_P}; \quad a_3\text{: decide } NEG_{A_P}.$$

Let $\lambda(a_i \mid X)$ denote the *loss* incurred for taking action a_i when an element $e \in E$ in fact belongs to X, and $\lambda(a_i \mid \neg X)$ the loss when the element belongs to $\neg X$. $P(X \mid E)$ and $P(\neg X \mid E)$ are the probabilities that an element $e \in E$ belongs to X and $\neg X$, respectively. The expected loss $R(a_i \mid E)$ associated with taking action a_i for all the elements in E can be expressed as:

$$\begin{aligned} R(a_1 \mid E) &= \lambda_{11} P(X \mid E) + \lambda_{12} P(\neg X \mid E)\ , \\ R(a_2 \mid E) &= \lambda_{21} P(X \mid E) + \lambda_{22} P(\neg X \mid E)\ , \\ R(a_3 \mid E) &= \lambda_{31} P(X \mid E) + \lambda_{32} P(\neg X \mid E)\ , \end{aligned} \tag{3.1}$$

where $\lambda_{i1} = \lambda(a_i \mid X)$ and $\lambda_{i2} = \lambda(a_i \mid \neg X)$. In decision-theoretic terminology, an expected loss is called a risk, and $R(a_i \mid E)$ is known as the conditional risk. Whenever a particular element $e \in E$ is encountered, the optimal Bayesian decision procedure suggests that we can minimize our expected loss by selecting the action minimizing the conditional risk. The selection of the action minimizing the conditional risk for every equivalence class E will minimize the overall risk (Duda, & Hart, 1973; Winkler, 1972; DeGroot, 1970).

Stated formally, we can decide a region to which any element $e \in E$ belongs by using the following set of minimum-risk decision rules:

(P) Decide POS_{A_P} if $R(a_1 \mid E) \le R(a_2 \mid E)$ and $R(a_1 \mid E) \le R(a_3 \mid E)$;

(D) Decide DOU_{A_P} if $R(a_2 \mid E) \le R(a_1 \mid E)$ and $R(a_2 \mid E) \le R(a_3 \mid E)$;

(N) Decide NEG_{A_P} if $R(a_3 \mid E) \le R(a_2 \mid E)$ and $R(a_3 \mid E) \le R(a_1 \mid E)$.

The actions that produce the same minimum risk are regarded equivalent in the decision-making process. In other words, the Bayesian decision theory can not differentiate actions producing the same risk. Depending on the application, one may use any convenient tie-breaking criteria to resolve this problem. In the decision rules (P), (D), and (N), we have not explicitly specified the tie-breaking criteria. Since $P(X \mid E) + P(\neg X \mid E) = 1$, the above decision rules may be simplified so that only the probabilities $P(X \mid E)$ are involved. Thus, we can decide the region for any element $e \in E$ based on the probability $P(X \mid E)$ and the loss function λ_{ij} ($i = 1, 2, 3, j = 1, 2$).

Consider a special loss function with $\lambda_{11} \le \lambda_{21} < \lambda_{31}$ and $\lambda_{32} \le \lambda_{22} < \lambda_{12}$. That is, the loss of classifying an element $e \in X$ into the POS_{A_P} region is less than or equal to the loss of classifying e into the DOU_{A_P} region, and both of these losses are strictly less than the loss of classifying e into the NEG_{A_P} region. The reverse order of losses are associated with classifying an element that does not belong to X. With this loss function, the minimum-risk decision rules (P), (D), and (N) can be simplified to:

(P) Decide POS_{A_P} if $P(X \mid E) \geq \beta$ and $P(X \mid E) \geq \gamma$,

(D) Decide DOU_{A_P} if $\delta \leq P(X \mid E) \leq \beta$,

(N) Decide NEG_{A_P} if $P(X \mid E) \leq \gamma$ and $P(X \mid E) \leq \delta$,

where

$$\gamma = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})} ,$$

$$\beta = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})} , \qquad \delta = \frac{\lambda_{22} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{21} - \lambda_{22})} . \tag{3.2}$$

From the assumptions, $\lambda_{11} \leq \lambda_{21} < \lambda_{31}$ and $\lambda_{32} \leq \lambda_{22} < \lambda_{12}$, it is not difficult to verify that $\beta \in (0, 1]$, $\gamma \in (0, 1)$, and $\delta \in [0, 1)$. One can also see that the loss function λ_{ij} should satisfy the condition $\delta \leq \beta$; otherwise, the classification problem would reduce to the one with only two disjoint regions instead of three as originally assumed. The decision rules (P), (D), and (N) provide a systematic way of partitioning the universe U into three disjoint regions using the parameters β, γ, and δ. It should perhaps be emphasized that these parameters are determined by a loss function which can be easily obtained from the user.

Now let us introduce the tie-breaking criteria for two separate cases: (i) $\delta < \beta$, and (ii) $\delta = \beta$.

Case (i) $\delta < \beta$

In this case, we have $\delta < \gamma < \beta$. When the risk of choosing POS_{A_P} or DOU_{A_P} is the same, we will always decide POS_{A_P}. Similarly, if the risk of choosing NEG_{A_P} or DOU_{A_P} is the same, we will always decide NEG_{A_P}. With these tie-breaking criteria, we arrive at a much simpler set of decision rules:

(P1) Decide POS_{A_P} if $P(X \mid E) \geq \beta$,

(D1) Decide DOU_{A_P} if $\delta < P(X \mid E) < \beta$,

(N1) Decide NEG_{A_P} if $P(X \mid E) \leq \delta$.

Using these decision rules, the universe U is partitioned into the positive, doubtful, and negative regions, which can be expressed explicitly in terms of the pair of parameters δ and β as follows:

$$\begin{aligned} POS_{A_P}(X, \beta, \delta) &= \bigcup_{P(X \mid E_i) \geq \beta} E_i , \\ DOU_{A_P}(X, \beta, \delta) &= \bigcup_{\delta < P(X \mid E_i) < \beta} E_i , \\ NEG_{A_P}(X, \beta, \delta) &= \bigcup_{P(X \mid E_i) \leq \delta} E_i . \end{aligned} \tag{3.3}$$

Based on these three regions, we may define the lower and upper approximations of X as $\underline{A_P}(X, \beta, \delta)$ and $\overline{A_P}(X, \beta, \delta)$:

$$\underline{A}_P(X, \beta, \delta) = POS_{A_P}(X, \beta, \delta) = \bigcup_{P(X \mid E_i) \geq \beta} E_i \ , \tag{3.4}$$

$$\begin{aligned} \overline{A}_P(X, \beta, \delta) &= POS_{A_P}(X, \beta, \delta) \cup DOU_{A_P}(X, \beta, \delta) \\ &= U - NEG_{A_P}(X, \beta, \delta) = \bigcup_{P(X \mid E_i) > \delta} E_i \ . \end{aligned} \tag{3.5}$$

The algebraic approximations of X can be derived from the approximations $\underline{A}_P(X, \beta, \delta)$ and $\overline{A}_P(X, \beta, \delta)$. Consider the following loss function:

$$\lambda_{31} = \lambda_{12} = 1 \ , \qquad \lambda_{11} = \lambda_{21} = \lambda_{32} = \lambda_{22} = 0 \ .$$

This means that there is a unit cost if the system classifies an element belonging to X into the negative region of X or if an element not belonging to X into the positive region; otherwise there is no cost. For such a loss function, eqn. (3.2) yields $\beta = 1$ and $\delta = 0$. Hence, according to eqns. (3.4) and (3.5) we have:

$$\underline{A}_P(X, 1, 0) = \bigcup_{P(X \mid E_i) = 1} E_i \ , \qquad \overline{A}_P(X, 1, 0) = \bigcup_{P(X \mid E_i) > 0} E_i \ . \tag{3.6}$$

Suppose the probabilities $P(X \mid E_i)$ can be estimated from the cardinalities of $X \cap E_i$ and E_i, namely, $P(X \mid E_i) = |X \cap E_i| / |E_i|$. In this case, $\underline{A}_P(X, 1, 0)$ and $\overline{A}_P(X, 1, 0)$ can be rewritten as:

$$\underline{A}_P(X, 1, 0) = \bigcup_{E_i \subseteq X} E_i \ , \qquad \overline{A}_P(X, 1, 0) = \bigcup_{E_i \cap X \neq \varnothing} E_i \ . \tag{3.7}$$

These are exactly the lower and upper approximations of X as defined by eqn. (2.1) in the algebraic theory of rough sets. The results we obtain here clearly indicate that the algebraic rough set model is indeed a special case of our generalized model.

Case (ii) $\delta = \beta$

We have $\delta = \gamma = \beta$ in this case. Whenever the risk of classifying an equivalence class into POS_{A_P} or DOU_{A_P} is the same, we will always decide DOU_{A_P}. Similarly, if the risk of classifying an equivalence class into NEG_{A_P} or DOU_{A_P} is the same, we will always choose DOU_{A_P}. From these tie-breaking criteria, we immediately obtain the following decision rules:

(P2) Decide POS_{A_P} if $P(X \mid E) > \gamma$,

(D2) Decide DOU_{A_P} if $P(X \mid E) = \gamma$,

(N2) Decide NEG_{A_P} if $P(X \mid E) < \gamma$.

The above decision rules partition the universe U into three disjoint regions based on the parameter γ:

$$\begin{aligned} POS_{A_P}(X, \gamma, \gamma) &= \bigcup_{P(X \mid E_i) > \gamma} E_i \ , \\ DOU_{A_P}(X, \gamma, \gamma) &= \bigcup_{P(X \mid E_i) = \gamma} E_i \ , \\ NEG_{A_P}(X, \gamma, \gamma) &= \bigcup_{P(X \mid E_i) < \gamma} E_i \ . \end{aligned} \tag{3.8}$$

Similar to case (i), we can define the lower and upper approximations of X as follows:

$$\underline{A}_P(X, \gamma, \gamma) = POS_{A_P}(X, \gamma, \gamma) = \bigcup_{P(X \mid E_i) > \gamma} E_i \ , \tag{3.9}$$

$$\begin{aligned}\overline{A}_P(X, \gamma, \gamma) &= POS_{A_P}(X, \gamma, \gamma) \cup DOU_{A_P}(X, \gamma, \gamma) \\ &= U - NEG_{A_P}(X, \gamma, \gamma) = \bigcup_{P(X \mid E_i) \geq \gamma} E_i \ .\end{aligned} \tag{3.10}$$

Consider the following loss function:

$$\lambda_{11} = \lambda_{32} = 0 \,, \quad \lambda_{21} = \lambda_{22} = 1/2 \,, \quad \lambda_{31} = \lambda_{12} = 1 \,.$$

That is, there is a unit cost if the system classifies an element belonging to X into the negative region or an element not belonging to X into the positive region. If any element is classified into the doubtful region there is half a unit cost. For other cases there is no cost. Substituting this loss function into eqn. (3.2), we obtain $\delta = \beta = \gamma = 1/2$. It is interesting to note that replacing γ by 1/2 in eqn. (3.8) immediately leads to the same disjoint regions as defined by Pawlak, Wong, and Ziarko (1988). Also, $\underline{A}_P(X, 1/2, 1/2)$ and $\overline{A}_P(X, 1/2, 1/2)$ are identical to their probabilistic lower and upper approximations of X. We have thus demonstrated that our approach based on the Bayesian decision theory is a generalization of the probabilistic rough set model as well.

For the pair of real numbers $\beta \in (0, 1]$ and $\delta \in [0, 1)$ with $\delta \leq \beta$ derived from the class of loss functions satisfying the conditions $\lambda_{11} \leq \lambda_{21} < \lambda_{31}$ and $\lambda_{32} \leq \lambda_{22} < \lambda_{12}$, we have defined the lower and upper approximations $\underline{A}_P(X, \beta, \delta)$ and $\overline{A}_P(X, \beta, \delta)$ for a given set $X \subseteq U$. Although these lower and upper approximations no longer correspond to the interior and closure operations in the topological space $T_A = (U, Com(A))$, they satisfy the following properties:

(II1) $\underline{A}_P(X \cap Y, \beta, \delta) \subseteq \underline{A}_P(X, \beta, \delta) \cap \underline{A}_P(Y, \beta, \delta)$

(II2) $\underline{A}_P(X \cup Y, \beta, \delta) \supseteq \underline{A}_P(X, \beta, \delta) \cup \underline{A}_P(Y, \beta, \delta)$

(II3) $\underline{A}_P(\neg X, \beta, \delta) = \neg \overline{A}_P(X, 1-\delta, 1-\beta)$

(II4) $\overline{A}_P(X \cap Y, \beta, \delta) \subseteq \overline{A}_P(X, \beta, \delta) \cap \overline{A}_P(Y, \beta, \delta)$

(II5) $\overline{A}_P(X \cup Y, \beta, \delta) \supseteq \overline{A}_P(X, \beta, \delta) \cup \overline{A}_P(Y, \beta, \delta)$

(II6) $\overline{A}_P(\neg X, \beta, \delta) = \neg \underline{A}_P(X, 1-\delta, 1-\beta)$

These properties are in fact a generalized version of the properties (I1)-(I6) in the algebraic rough set model. They subsume those given by Pawlak, Wong, and Ziarko (1988) in the probabilistic rough set model. From properties (II3) and (II6), we obtain:

(II7) $POS_{A_P}(X, \beta, \delta) = NEG_{A_P}(\neg X, 1-\delta, 1-\beta)$,

(II8) $DOU_{A_P}(X, \beta, \delta) = DOU_{A_P}(\neg X, 1-\delta, 1-\beta)$,

(II9) $NEG_{A_P}(X, \beta, \delta) = POS_{A_P}(\neg X, 1-\delta, 1-\beta)$.

Thus we say that $(\underline{A}_P(X, \beta, \delta), \overline{A}_P(X, \beta, \delta))$ and $(\underline{A}_P(\neg X, 1-\delta, 1-\beta), \overline{A}_P(\neg X, 1-\delta, 1-\beta))$ complement each other with respect to $A_P = (U, R, P)$.

4. Conclusion

In this paper we have suggested a generalized rough set model based on the Bayesian decision theory. The lower and upper approximations of a given concept are derived from a Bayesian decision procedure, where a set of objects are partitioned into three disjoint regions. These three regions are uniquely determined by a loss function which can be easily obtained from the user. It has been shown that both the algebraic and probabilistic rough set models can be viewed as special cases of the proposed generalized model.

Reference

DeGroot, M.H. (1970). *Optimal Statistical Decisions,* McGraw-Hill, New York.

Duda, R. O., & Hart., P. E. (1973). *Pattern Classification and Scene Analysis,* Wiley, New York.

Kowalczyk, A., & Szymanski, J. (1989). Rough reduction of decision tables, in Janicki, R. and Koczkodaj, W.W. (editors) *Computing and Information -- Proceedings of the International Conference on Computing and Information, ICCI'89,* pp. 255-261, North-Holland, Amsterdam.

Pawlak, Z. (1982). Rough sets, *International Journal of Information and Computer Sciences,* 11, pp. 145-172.

Pawlak, Z. (1984). Rough classification, *International Journal of Man-Machine Studies,* 20, pp. 469-483.

Pawlak, Z., Slowinski, K., & Slowinski, R. (1986). Rough classification of patients after highly selective vagotomy for duodendal ulcer, *International Journal of Man-Machine Studies,* 24, pp. 413-433.

Pawlak, Z., Wong, S.K.M., & Ziarko, W. (1988). Rough sets: probabilistic versus deterministic approach, *International Journal of Man-Machine Studies,* 29, pp. 81-95.

Pettorossi, A., Ras, Z.W., & Zemankova, M. (1987). On learning with imperfect teachers, in Ras, Z.W. and Zemankova, M. (editors) *Methodologies for Intelligent Systems,* pp. 256-263, North-Holland, New York.

Srinivasan, P. (1989). Approximation for information retrieval, in Ras, Z.W. (editor) *Methodologies for Intelligent Systems 4,* pp. 128-136, North-Holland, New York.

Winkler, R.L. (1972). *Introduction to Bayesian Inference and Decision,* Holt, Rinehart and Winston, Inc., New York.

Wong, S.K.M., & Ziarko, W. (1986). INFER -- an adaptive decision support system, *Proceedings of the Sixth International Workshop on Expert Systems and Their Applications,* Vol. 1, pp. 713-726, Avignon, France.

Wong, S.K.M., & Ziarko, W. (1987). Comparison of the probabilistic approximate classification and the fuzzy set model, *Fuzzy Sets and Systems,* 21, pp. 357-362.

Wong, S.K.M., Ziarko, W., & Ye, R.L. (1987). Comparison of the rough set and statistical methods in inductive learning, *International Journal of Man-Machine Studies,* 24, pp. 53-72.