Yao, Y.Y. and Wong, S.K.M. A decision theoretic framework for approximating concepts,
International Journal of Man-machine Studies , Vol. 37, No. 6, pp. 793-809, 1992

A Decision Theoretic Framework for Approximating Concepts

Y.Y. Yao and S.K.M. Wong

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

Abstract

This paper explores the implications of approximating a concept based on the Bayesian decision procedure, which provides a plausible unification of the fuzzy set and rough set approaches for approximating a concept. We show that if a given concept is approximated by *one* set, the same result given by the α-cut in the fuzzy set theory is obtained. On the other hand, if a given concept is approximated by *two* sets, we can derive both the algebraic and probabilistic rough set approximations. Moreover, based on the well known principle of maximum (minimum) entropy, we give a useful interpretation of fuzzy intersection and union. Our results enhance the understanding and broaden the applications of both fuzzy and rough sets.

1. Introduction

The issues of representing and inferring a *concept* are of fundamental importance in the design of intelligent systems. In many applications, a class (subset) of objects can be interpreted as representing a concept. Suppose each object in the universe of discourse is described or characterized by the values of a set of attributes. Then the problem of identifying a concept is reduced to the problem of determining whether an object belongs to a particular subset based on the description of the object. If the descriptions of the individual objects are sufficient and precise enough with respect to a given concept, one can unambiguously describe the class, a subset of objects, representing the concept. However, the available knowledge in many practical situations is often incomplete and imprecise (Pawlak, Wong & Ziarko, 1988). Under such circumstances, it may not be possible to determine the exact membership of every object in the universe. Instead, one may provide an approximate characterization of a subset of objects based on their attribute values.

Many methods were proposed to deal with the approximation of a concept. For example, the well known fuzzy set theory characterizes a concept *approximately* by a membership function with a range between 0 and 1. Another approach is based on the rough set theory which provides the lower and upper approximations of a concept. The relationship between these approaches was studied by many authors (Dubois & Prade, 1990, 1987; Wygralak, 1989; Wong & Ziarko, 1987; Pawlak, 1985). The focus of this paper is to explore the implications of approximating a concept within the Bayesian decision theoretic framework, which provides a plausible unification of the fuzzy set and rough set approaches for approximating a concept. We believe that our formulation broadens the applications of rough sets for machine learning and decision-making. Moreover, based on the well known principle of maximum (minimum) entropy, our approach provides a useful interpretation for the different versions of fuzzy intersection and union.

In Section 2, we suggest two methods to approximate a concept of interest. The first leads to the notion of α-cut in the fuzzy set theory, and the second leads to a generalization of the rough set model. In Section 3, we discuss how to combine concepts, and finally comment on the fuzzy intersection and union.

2. Approximation of Concepts

The objective in this section is to develop a decision theoretic framework to approximate a concept which can be interpreted as a *label* of a class of objects (Hisdal, 1988a; Wong & Ziarko, 1987).

We assume that each object in the universe of discourse is labeled by a set of attribute values. Let $A = \{A_1, \ldots, A_n\}$ be a set of attributes, and let $V_1, \ldots, V_n$ be the domains of these attributes. Each object is described by one of the elements in the Cartesian product, $V_1 \times \ldots \times V_n$, which is referred to as the knowledge representation space. When two objects have the same description, we say that they are indistinguishable in such a knowledge representation system. Figure 2.1 depicts a knowledge representation space and a concept w characterized by a set of objects. The problem is to define this subset w in terms of the descriptions of the objects. In other words, given an object described by $\mathbf{x}$, one would like to decide if this object belongs to w or not. For example, if all the objects with the same description belong to w (see $\mathbf{x}_1$ in Figure 2.1), we have a deterministic decision rule. That is, we can conclude unambiguously that any object with description $\mathbf{x}_1$ definitely belongs to w. However, if not all the objects with the same description belong to w (see $\mathbf{x}_2$ in Figure 2.1), we have a non-deterministic decision rule. In this case, it is not possible to decide on the membership of a object with description $\mathbf{x}_2$. That is, this object may or may not be a member of w. Therefore, in many situations we may not be able to precisely define a concept w based on the descriptions of the objects alone. Instead, one can only approximately define the set w according to some external decision criteria.

There are many ways to approximate a concept. The approach we adopt here is based on the Bayesian decision theory. It is perhaps clearer first to consider the following example to illustrate the practical needs to classify objects based on incomplete information. Suppose a physician wants to identify those patients, from a group of patients, who have contracted a particular disease w based only on the available symptoms of the individual patients. If symptom $\mathbf{x}$ suggests that a patient definitely has contracted the disease w, then the physician can immediately provide the patient with the proper treatment without undue delay. However, in many cases

Yao, Y.Y. and Wong, S.K.M. A decision theoretic framework for approximating concepts,
International Journal of Man-machine Studies , Vol. 37, No. 6, pp. 793-809, 1992

Figure 2.1: A concept w in the knowledge representation space

the symptoms may not be sufficient for the physician to judge if a person suffers from the disease w. Assume that we are in an emergency situation (i.e., disease w is life threatening). Under these circumstances, the physician must consider how likely a patient would have contracted the disease w given symptom $\mathbf{x}$, and consider the consequences (costs) of not treating a patient who has the disease and treating a patient who does not have the disease. For a particular disease w, let $POS(w)$ denote the the set of patients who will receive treatment, and let $NEG(w)$ denote the set of patients who will not receive treatment immediately. That is, w is approximated by $POS(w)$. However, in some situations, it may be more appropriate to divide the patients into three groups $POS(w)$, $NEG(w)$, and $DOU(w)$ instead of just two groups. In this case, patients in $POS(w)$ require immediate treatment and patients in $NEG(w)$ do not require the treatment. The set $DOU(w)$ represents those patients who need further examination before a proper decision can be made by the physician.

In the following subsections, based on the Bayesian decision procedure we will show that if a concept is approximated by *one* set, the result of such an approximation is the same as the α-cut in the fuzzy set theory. On the other hand, if a concept is approximated by *two* sets, we will demonstrate that the rough set model is in fact a special case of our approach.

2.1. The Bayesian decision procedure

For completeness, we briefly review here the basic notions of the Bayesian decision procedure pertinent to the approximation of a concept (Duda & Hart, 1973).

Let $\Omega = \{w_1, \ldots, w_s\}$ be a finite set of s states of nature, and $A = \{a_1, \ldots, a_m\}$ be a finite set of m possible actions. Let $P(w_j|\mathbf{x})$ be the probability of an object in state w_j given that the object is described by $\mathbf{x}$. In the following discussions, we assume that these conditional probabilities $P(w_j|\mathbf{x})$ are known.

Let $\lambda(a_i|w_j)$ denote the loss for taking action a_i when the state is w_j. For an object with description $\mathbf{x}$, suppose action a_i is taken. Since $P(w_j|\mathbf{x})$ is the probability that the true state is w_j given $\mathbf{x}$, the expected loss associated with taking action a_i is given by:

$$R(a_i|\mathbf{x}) = \sum_{j=1}^{s} \lambda(a_i|w_j) P(w_j|\mathbf{x}). \tag{2.1}$$

The quantity $R(a_i|\mathbf{x})$ is also called the conditional risk. Given description $\mathbf{x}$, a decision rule is a function $\tau(\mathbf{x})$ that specifies which action to take. That is, for every $\mathbf{x}$, $\tau(\mathbf{x})$ assumes one of the actions, $a_1, \ldots, a_m$. The overall risk $\mathbf{R}$ is the expected loss associated with a given decision rule. Since $R(\tau(\mathbf{x})|\mathbf{x})$ is the conditional risk associated with action $\tau(\mathbf{x})$, the overall risk is defined by:

$$\mathbf{R} = \sum_{\mathbf{x}} R(\tau(\mathbf{x})|\mathbf{x}) P(\mathbf{x}), \tag{2.2}$$

where the summation is over the entire knowledge representation space. If $\tau(\mathbf{x})$ is chosen so that $R(\tau(\mathbf{x})|\mathbf{x})$ is as small as possible for every $\mathbf{x}$, the overall risk $\mathbf{R}$ is minimized. Thus, the Bayesian decision procedure can be formally stated as follows. For every $\mathbf{x}$, compute the conditional risk $R(a_i|\mathbf{x})$ for $i = 1, \ldots, m$ defined by equation (2.1), and then select the action for which the conditional risk is minimum. If more than one action minimizes $R(a_i|\mathbf{x})$, any convenient tie-breaking rule can be used.

2.2. Approximation by a single set

Suppose only a single set is used to approximate a concept w. In this case, the entire universe is partitioned into two regions, a positive region $POS(w)$ and a negative region $NEG(w)$. The positive region $POS(w)$ is considered as an approximation of w. Let $\neg w$ denote the complement (negation) of w. Here, we have only two states, namely, $\Omega = \{w, \neg w\}$ and two actions $A = \{a_1, a_2\}$, where a_1 and a_2 denote the actions of deciding $POS(w)$ and deciding $NEG(w)$, respectively.

Let $\lambda_{11} = \lambda(a_1|w)$ be the cost of deciding $POS(w)$ when the object actually belongs to w, and let $\lambda_{12} = \lambda(a_1|\neg w)$ be the cost of deciding $POS(w)$ when the object does not belong to w. Likewise, let $\lambda_{21} = \lambda(a_2|w)$ and $\lambda_{22} = \lambda(a_2|\neg w)$ be the cost of deciding $NEG(w)$ when the object actually belongs and does not belong to w, respectively. Given an object with description $\mathbf{x}$, the expected loss $R(a_i|\mathbf{x})$ associated with taking action a_i can be expressed as:

$$\begin{aligned} R(a_1|\mathbf{x}) &= \lambda_{11} P(w|\mathbf{x}) + \lambda_{12} P(\neg w|\mathbf{x}), \\ R(a_2|\mathbf{x}) &= \lambda_{21} P(w|\mathbf{x}) + \lambda_{22} P(\neg w|\mathbf{x}). \end{aligned} \tag{2.3}$$

By applying the Bayesian decision procedure, we obtain the following minimum-risk decision rules:

(P) Decide $POS(w)$ if $R(a_1|\mathbf{x}) \leq R(a_2|\mathbf{x})$;

(N) Decide $NEG(w)$ if $R(a_2|\mathbf{x}) \leq R(a_1|\mathbf{x})$.

Note that it is necessary to choose a tie-breaking rule to differentiate actions producing the same risk.

Let us consider a loss function with $\lambda_{11} < \lambda_{21}$ and $\lambda_{22} < \lambda_{12}$. That is, the loss of classifying an object belonging to w into the positive region $POS(w)$ is less than the loss of classifying it into the negative region $NEG(w)$; whereas the loss of classifying an object not belonging to w into the positive region $POS(w)$ is greater than the loss of classifying it into the negative region $NEG(w)$. With this loss function and the fact that $P(w|\mathbf{x}) + P(\neg w|\mathbf{x}) = 1$, the above decision rules can be expressed as:

$$\begin{aligned} &\text{(P)} \quad \text{Decide} \quad POS(w) \quad \text{if} \quad P(w|\mathbf{x}) \geq \alpha; \\ &\text{(N)} \quad \text{Decide} \quad NEG(w) \quad \text{if} \quad P(w|\mathbf{x}) \leq \alpha, \end{aligned}$$

where

$$\alpha = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}. \tag{2.4}$$

Based on the assumptions, $\lambda_{11} < \lambda_{21}$ and $\lambda_{22} < \lambda_{12}$, it is not difficult to verify that $0 < \alpha < 1$.

When the risk of deciding $POS(w)$ or $NEG(w)$ is the same, suppose $POS(w)$ is selected. With this tie-breaking criterion, the decision rules become:

$$\begin{aligned} &\text{(P1)} \quad \text{Decide} \quad POS(w) \quad \text{if} \quad P(w|\mathbf{x}) \geq \alpha; \\ &\text{(N1)} \quad \text{Decide} \quad NEG(w) \quad \text{if} \quad P(w|\mathbf{x}) < \alpha. \end{aligned}$$

Thus, the positive and negative regions can be expressed explicitly in terms of the parameter α, namely:

$$\begin{aligned} POS(w, \alpha) &= \bigcup_{P(w|\mathbf{x}) \geq \alpha} [\mathbf{x}], \\ NEG(w, \alpha) &= \bigcup_{P(w|\mathbf{x}) < \alpha} [\mathbf{x}], \end{aligned} \tag{2.5}$$

where $[\mathbf{x}]$ denotes the set of objects described by $\mathbf{x}$. The positive region $POS(w, \alpha)$ is indeed an approximation of w:

$$Apr(w, \alpha) = POS(w, \alpha) = \bigcup_{P(w|\mathbf{x}) \geq \alpha} [\mathbf{x}]. \tag{2.6}$$

In fact, one may choose another tie-breaking criterion: if the risk of deciding $POS(w)$ or $NEG(w)$ is the same, we decide $NEG(w)$. In this case, we obtain the following decision rules:

$$\text{(P2)} \quad \text{Decide} \quad POS(w) \quad \text{if} \quad P(w|\mathbf{x}) > \alpha;$$
$$\text{(N2)} \quad \text{Decide} \quad NEG(w) \quad \text{if} \quad P(w|\mathbf{x}) \leq \alpha.$$

The corresponding positive and negative regions, $POS(w, \alpha_+)$ and $NEG(w, \alpha_+)$, are defined by:

$$\begin{aligned} POS(w, \alpha_+) &= \bigcup_{P(w|\mathbf{x})>\alpha} [\mathbf{x}], \\ NEG(w, \alpha_+) &= \bigcup_{P(w|\mathbf{x})\leq\alpha} [\mathbf{x}]. \end{aligned} \tag{2.7}$$

Thus, one can define another approximation of w as:

$$Apr(w, \alpha_+) = POS(w, \alpha_+) = \bigcup_{P(w|\mathbf{x})>\alpha} [\mathbf{x}]. \tag{2.8}$$

For any real number $\alpha \in (0, 1)$ computed from equation (2.4) with the loss function satisfying $\lambda_{11} < \lambda_{21}$ and $\lambda_{22} < \lambda_{12}$, the following properties hold:

$$\text{(I1)} \quad POS(w, \alpha) = NEG(\neg w, (1-\alpha)_+)$$
$$\text{(I2)} \quad POS(w, \alpha_+) = NEG(\neg w, 1-\alpha)$$
$$\text{(I3)} \quad NEG(w, \alpha) = POS(\neg w, (1-\alpha)_+)$$
$$\text{(I4)} \quad NEG(w, \alpha_+) = POS(\neg w, 1-\alpha)$$

Obviously, the two approximations $Apr(w, \alpha)$ and $Apr(\neg w, (1-\alpha)_+)$ complement each other.

The approximation of a concept w by one set can also be analyzed from the fuzzy set point of view (Kandel, 1986; Dubois & Prade, 1980). Let $\mathbf{x}_o$ denote the description of an object o. If we cannot judge, relying on the description $\mathbf{x}_o$ alone, whether the object o is a member of the set w, we can define a fuzzy set to characterize such a membership. Based on the conditional probabilities $P(w|\mathbf{x})$, a membership function $\mu_{\tilde{w}}$ with respect to w can be defined as:

$$\mu_{\tilde{w}}(o) = P(w|\mathbf{x}_o). \tag{2.9}$$

Thus, for any real number $\alpha \in (0,1)$, we can construct an approximation $\tilde{w}_\alpha$ of the fuzzy set $\tilde{w}$ as follows:

$$\tilde{w}_\alpha = \{o|\mu_w(o) \geq \alpha\} = \{o|P(w|\mathbf{x}_o) \geq \alpha\} = \bigcup_{P(w|\mathbf{x})\geq\alpha} [\mathbf{x}], \qquad (2.10)$$

which is called an α-cut of the fuzzy set $\tilde{w}$. Similarly, another approximation of $\tilde{w}$ can be defined as:

$$\tilde{w}_{\alpha_+} = \{o|\mu_w(o) > \alpha\} = \{o|P(w|\mathbf{x}_o) > \alpha\} = \bigcup_{P(w|\mathbf{x})>\alpha} [\mathbf{x}], \qquad (2.11)$$

which is known as a *strong* α-cut of $\tilde{w}$.

It is clear that by comparing the fuzzy set approach with our method, $Apr(w,\alpha)$ is equivalent to the α-cut of $\tilde{w}$, while $Apr(w,\alpha_+)$ is equivalent to the strong α-cut of $\tilde{w}$. In general, a membership function is not necessarily defined in terms of the descriptions of the objects and the conditional probabilities involved. The membership function defined by equation (2.9) represents a special kind of fuzzy sets. The use of probability for defining fuzzy set membership has been a controversial issue (Hisdal, 1988a, 1988b; Zadeh, 1978). The connections between fuzzy set and probability have been investigated by many authors (Hisdal, 1988a, 1988b; Wong & Ziarko, 1987; Giles, 1982, 1976; Gaines, 1978; Hersh & Caramazza, 1976; Ruspini, 1969). In fact, Hisdal (1988a, 1988b) argued quite convincingly that one should not rule out the possibility that there is a useful probabilistic interpretation of fuzzy membership functions. Based on the probabilistic interpretation, our analysis provides a plausible justification for using the α-cuts as approximations. Each α-cut, defined by a cost function, represents a different level of approximation of w. Which level of approximation is deemed suitable depends of course on the application itself.

2.3. Lower and upper approximations

Instead of using one set, one may use *two* sets to approximate a given concept w. For this purpose, we partition the universe into three regions, the positive region $POS(w)$, the negative region $NEG(w)$, and the doubtful region $DOU(w)$. The concept w can be characterized by a lower approximation $POS(w)$ and an upper approximation $POS(w) \cup DOU(w)$ (Yao, Wong & Lingras, 1990). In this case, the set of states remains $\Omega = \{w, \neg w\}$, but the set of actions becomes $A = \{a_1, a_2, a_3\}$, where a_1, a_2, and a_3 represent the three actions, deciding $POS(w)$, deciding $NEG(w)$, and deciding $DOU(w)$, respectively.

Let $\lambda(a_i|w)$ denote the loss incurred for taking action a_i when an object in fact belongs to w, and let $\lambda(a_i|\neg w)$ denote the loss incurred when the object actually belongs to $\neg w$. $P(w|\mathbf{x})$ and $P(\neg w|\mathbf{x})$ are the probabilities that an object with description $\mathbf{x}$ belongs to w and $\neg w$, respectively. Thus, the expected loss $R(a_i|\mathbf{x})$ associated with taking the individual actions can be expressed as:

$$\begin{aligned}
R(a_1|\mathbf{x}) &= \lambda_{11}P(w|\mathbf{x}) + \lambda_{12}P(\neg w|\mathbf{x}), \\
R(a_2|\mathbf{x}) &= \lambda_{21}P(w|\mathbf{x}) + \lambda_{22}P(\neg w|\mathbf{x}), \\
R(a_3|\mathbf{x}) &= \lambda_{31}P(w|\mathbf{x}) + \lambda_{32}P(\neg w|\mathbf{x}), \qquad (2.12)
\end{aligned}$$

where $\lambda_{i1} = \lambda(a_i|w)$, $\lambda_{i2} = \lambda(a_i|\neg w)$, and $i = 1, 2, 3$. The Bayesian decision procedure leads to the following minimum-risk decision rules:

(P′) Decide $POS(w)$ if $R(a_1|\mathbf{x}) \leq R(a_2|\mathbf{x})$ and $R(a_1|\mathbf{x}) \leq R(a_3|\mathbf{x})$;

(N′) Decide $NEG(w)$ if $R(a_2|\mathbf{x}) \leq R(a_1|\mathbf{x})$ and $R(a_2|\mathbf{x}) \leq R(a_3|\mathbf{x})$;

(D′) Decide $DOU(w)$ if $R(a_3|\mathbf{x}) \leq R(a_1|\mathbf{x})$ and $R(a_3|\mathbf{x}) \leq R(a_2|\mathbf{x})$.

Since $P(w|\mathbf{x}) + P(\neg w|\mathbf{x}) = 1$, the above decision rules can be simplified such that only the probabilities $P(w|\mathbf{x})$ are involved. Thus, we can classify any object with description $\mathbf{x}$ based only on the probabilities $P(w|\mathbf{x})$ and the given loss function λ_{ij} ($i = 1, 2, 3$; $j = 1, 2$).

Consider a special kind of loss functions with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$. That is, the loss of classifying an object o belonging to w into the positive region $POS(w)$ is less than or equal to the loss of classifying o into the doubtful region $DOU(w)$, and both of these losses are strictly less than the loss of classifying o into the negative region $NEG(w)$. We obtain the reverse order of losses by classifying an object that does not belong to w. For this type of loss functions, the minimum-risk decision rules (P′)-(D′) can be written as:

$$\begin{array}{lll} (\mathrm{P}') & \text{Decide } POS(w) & \text{if } P(w|\mathbf{x}) \geq \beta \text{ and } P(w|\mathbf{x}) \geq \gamma; \\ (\mathrm{N}') & \text{Decide } NEG(w) & \text{if } P(w|\mathbf{x}) \leq \gamma \text{ and } P(w|\mathbf{x}) \leq \delta; \\ (\mathrm{D}') & \text{Decide } DOU(w) & \text{if } \delta \leq P(w|\mathbf{x}) \leq \beta; \end{array}$$

where

$$\begin{aligned} \beta &= \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})}, \\ \gamma &= \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}, \\ \delta &= \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}. \end{aligned} \tag{2.13}$$

From the assumptions, $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, it follows that $\beta \in (0, 1]$, $\gamma \in (0, 1)$, and $\delta \in [0, 1)$. Note that the parameters λ_{ij} should satisfy the condition $\delta \leq \beta$; otherwise, this approximation problem would be reduced to the one discussed in Section 2.2. It should be emphasized that the decision rules (P′)-(D′) depend only on the input parameters β, γ, and δ computed from the λ_{ij}'s directly supplied by the user or expert.

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

2.3.1. Case (i) $\delta < \beta$

In this case, we have $\delta < \gamma < \beta$. When the risk of deciding $POS(w)$ or $DOU(w)$ is the same, we decide $POS(w)$; if the risk of deciding $NEG(w)$ or $DOU(w)$ is the same, we decide $NEG(w)$. With these tie-breaking criteria, we obtain from (P′)-(D′) a simpler set of decision rules:

$$
\begin{array}{lll}
\text{(P}'\text{1)} & \text{Decide } POS(w) & \text{if } P(w|\mathbf{x}) \geq \beta; \\
\text{(N}'\text{1)} & \text{Decide } NEG(w) & \text{if } P(w|\mathbf{x}) \leq \delta; \\
\text{(D}'\text{1)} & \text{Decide } DOU(w) & \text{if } \delta < P(w|\mathbf{x}) < \beta.
\end{array}
$$

The positive, negative, and doubtful regions can be explicitly expressed in terms of the pair of parameters δ and β, namely:

$$
\begin{aligned}
POS(w, \beta, \delta) &= \bigcup_{P(w|\mathbf{x}) \geq \beta} [\mathbf{x}], \\
NEG(w, \beta, \delta) &= \bigcup_{P(w|\mathbf{x}) \leq \delta} [\mathbf{x}], \\
DOU(w, \beta, \delta) &= \bigcup_{\delta < P(w|\mathbf{x}) < \beta} [\mathbf{x}], \qquad (2.14)
\end{aligned}
$$

where $[\mathbf{x}]$ denotes the set of objects described by $\mathbf{x}$. We can now define the lower and upper approximations $\underline{Apr}(w, \beta, \delta)$ and $\overline{Apr}(w, \beta, \delta)$ of w as:

$$
\begin{aligned}
\underline{Apr}(w, \beta, \delta) &= POS(w, \beta, \delta) = \bigcup_{P(w|\mathbf{x}) \geq \beta} [\mathbf{x}], \\
\overline{Apr}(w, \beta, \delta) &= POS(w, \beta, \delta) \cup DOU(w, \beta, \delta) = \bigcup_{P(w|\mathbf{x}) > \delta} [\mathbf{x}]. \qquad (2.15)
\end{aligned}
$$

The *algebraic* approximations of a concept introduced in the rough set model (Pawlak, 1984, 1982) can be easily derived from the lower and upper approximations $\underline{Apr}(w, \beta, \delta)$ and $\overline{Apr}(w, \beta, \delta)$. Consider the following loss function:

$$
\lambda_{12} = \lambda_{21} = 1, \qquad \lambda_{11} = \lambda_{22} = \lambda_{31} = \lambda_{32} = 0. \qquad (2.16)
$$

This means that there is a unit cost if an object belonging to w is classified into the negative region or if an object not belonging to w is classified into the positive region; otherwise there is no cost. For such a loss function, we obtain from equation (2.13) that $\beta = 1$ and $\delta = 0$. Hence, according to equation (2.15), we have:

$$
\underline{Apr}(w, 1, 0) = \bigcup_{P(w|\mathbf{x}) = 1} [\mathbf{x}],
$$

$$\overline{Apr}(w,1,0) = \bigcup_{P(w|\mathbf{x})>0} [\mathbf{x}]. \tag{2.17}$$

Suppose the probabilities $P(w|\mathbf{x})$ can be estimated from the cardinalities of $w \cap [\mathbf{x}]$ and $[\mathbf{x}]$, namely, $P(w|\mathbf{x}) = |w \cap [\mathbf{x}]|/|[\mathbf{x}]|$. In this case, $\underline{Apr}(w,1,0)$ and $\overline{Apr}(w,1,0)$ can be expressed as:

$$\begin{aligned} \underline{Apr}(w,1,0) &= \bigcup_{[\mathbf{x}]\subseteq w} [\mathbf{x}], \\ \overline{Apr}(w,1,0) &= \bigcup_{[\mathbf{x}]\cap w\neq\emptyset} [\mathbf{x}]. \end{aligned} \tag{2.18}$$

These are exactly the lower and upper approximations of w defined in the algebraic theory of rough sets (Pawlak, 1982). The results given here suggest that the algebraic rough set model can be viewed as a special case of our decision theoretic approach.

2.3.2. **Case (ii)** $\delta = \beta$

Since $\delta = \beta$, we have $\delta = \gamma = \beta$. Here we adopt a different set of tie-breaking rules. Whenever the risk of classifying an object into $POS(w)$ or $DOU(w)$ is the same, we decide $DOU(w)$. If the risk of classifying an object into $NEG(w)$ or $DOU(w)$ is the same, we decide $DOU(w)$. In this case, the decision rules (P′)-(D′) can be written as:

(P′2) Decide $POS(w)$ if $P(w|\mathbf{x}) > \gamma$;

(N′2) Decide $NEG(w)$ if $P(w|\mathbf{x}) < \gamma$;

(D′2) Decide $DOU(w)$ if $P(w|\mathbf{x}) = \gamma$.

Accordingly:

$$\begin{aligned} POS(w,\gamma,\gamma) &= \bigcup_{P(w|\mathbf{x})>\gamma} [\mathbf{x}], \\ NEG(w,\gamma,\gamma) &= \bigcup_{P(w|\mathbf{x})<\gamma} [\mathbf{x}], \\ DOU(w,\gamma,\gamma) &= \bigcup_{P(w|\mathbf{x})=\gamma} [\mathbf{x}]. \end{aligned} \tag{2.19}$$

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

$$\begin{aligned} \underline{Apr}(w,\gamma,\gamma) &= POS(w,\gamma,\gamma) = \bigcup_{P(w|\mathbf{x})>\gamma} [\mathbf{x}], \\ \overline{Apr}(w,\gamma,\gamma) &= POS(w,\beta,\delta) \cup DOU(w,\gamma,\gamma) = \bigcup_{P(w|\mathbf{x})\geq\gamma} [\mathbf{x}]. \end{aligned} \tag{2.20}$$

Consider a special loss function:

$$\lambda_{12} = \lambda_{21} = 1, \quad \lambda_{31} = \lambda_{32} = 1/2, \quad \lambda_{11} = \lambda_{22} = 0. \tag{2.21}$$

That is, a unit cost is incurred if the system classifies an object belonging to w into the negative region or an object not belonging to w is classified into the positive region; half of a unit cost is incurred if any object is classified into the doubtful region. For other cases, there is no cost. Substituting these λ_{ij}'s into equation (2.13), we obtain $\delta = \beta = \gamma = 1/2$. It is interesting to note that by replacing γ by 1/2 in equation (2.20), we arrive at the same results obtained by Pawlak, Wong, and Ziarko (1988). In particular, $\underline{Apr}(w, 1/2, 1/2)$ and $\overline{Apr}(w, 1/2, 1/2)$ are identical to their *probabilistic* lower and upper approximations of w. We have thus demonstrated that our approach based on the Bayesian decision theory is a generalization of the probabilistic rough set model as well.

Given $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, for any pair of real numbers β, δ (defined by equation (2.13)) satisfying the condition $\delta \leq \beta$, the following properties hold:

(II1) $POS(w,\beta,\delta) = NEG(\neg w, 1-\delta, 1-\beta)$

(II2) $NEG(w,\beta,\delta) = POS(\neg w, 1-\delta, 1-\beta)$

(II3) $DOU(w,\beta,\delta) = DOU(\neg w, 1-\delta, 1-\beta)$

Thus, $(\underline{Apr}(w,\beta,\delta), \overline{Apr}(w,\beta,\delta))$ and $(\underline{Apr}(\neg w, 1-\delta, 1-\beta), \overline{Apr}(\neg w, 1-\delta, 1-\beta))$ complement each other.

It is perhaps worth mentioning here that in general one can apply, if necessary, the same procedure to approximate a concept by more than two sets by introducing an appropriate cost function.

3. Combination of Concepts

In the previous section, we have shown how the decision procedure can be used to approximate a *single* concept represented by a subset of objects in the universe of discourse. In many applications, it is often necessary to consider a number of concepts. For example, in the medical diagnostic problem discussed in Section 2, a physician may be interested in identifying the patients who have contracted both diseases w and w'. The physician may also be interested in identifying the patients who have contracted either disease w or disease w'. These two categories of patients can be represented by the intersection $w \cap w'$ and the union $w \cup w'$, respectively. If the conditional probabilities $P(w \cap w'|\mathbf{x})$ or $P(w \cup w'|\mathbf{x})$ are known, we can compute, for instance, the approximations $Apr(w \cap w', \alpha)$ and $Apr(w \cup w', \alpha)$ as defined by equation (2.6) in Section 2.2.

A probability function satisfies the inequalities:

$$
\begin{aligned}
&P(w \cap w'|\mathbf{x}) \leq P(w|\mathbf{x}) \ , \qquad && P(w \cap w'|\mathbf{x}) \leq P(w'|\mathbf{x}) \ , \\
&P(w \cup w'|\mathbf{x}) \geq P(w|\mathbf{x}) \ , \qquad && P(w \cup w'|\mathbf{x}) \geq P(w'|\mathbf{x}) \ . \qquad (3.1)
\end{aligned}
$$

Based on these inequalities, it can be easily verified that the following properties hold:

(O1) $Apr(w \cap w', \alpha) \subseteq Apr(w, \alpha) \cap Apr(w', \alpha)$

(O2) $Apr(w \cup w', \alpha) \supseteq Apr(w, \alpha) \cup Apr(w', \alpha)$

(O3) $Apr(w \cap w', \alpha_+) \subseteq Apr(w, \alpha_+) \cap Apr(w', \alpha_+)$

(O4) $Apr(w \cup w', \alpha_+) \supseteq Apr(w, \alpha_+) \cup Apr(w', \alpha_+)$.

These properties can be considered as a generalized version of those satisfied by an α-cut in the fuzzy set theory (Kandel, 1986; Dubois & Prade, 1980). On the other hand, when the given concept is approximated by two sets as discussed in Section 2.3, one can show that:

(T1) $\underline{Apr}(w \cap w', \beta, \delta) \subseteq \underline{Apr}(w, \beta, \delta) \cap \underline{Apr}(w', \beta, \delta)$

$$\text{(T2)} \quad \underline{Apr}(w \cup w', \beta, \delta) \supseteq \underline{Apr}(w, \beta, \delta) \cup \underline{Apr}(w', \beta, \delta)$$
$$\text{(T3)} \quad \overline{Apr}(w \cap w', \beta, \delta) \subseteq \overline{Apr}(w, \beta, \delta) \cap \overline{Apr}(w', \beta, \delta)$$
$$\text{(T4)} \quad \overline{Apr}(w \cup w', \beta, \delta) \supseteq \overline{Apr}(w, \beta, \delta) \cup \overline{Apr}(w', \beta, \delta)$$

In fact, these properties subsume those given by Pawlak, Wong, and Ziarko (1988) in the probabilistic rough set model.

In many situations, one may not know some of the conditional probabilities $P(w \cap w'|\mathbf{x})$ or $P(w \cup w'|\mathbf{x})$. However, one may assume that the probabilities $P(w|\mathbf{x})$ and $P(w'|\mathbf{x})$ are known. Since $P(w \cup w'|\mathbf{x}) = P(w|\mathbf{x}) + P(w'|\mathbf{x}) - P(w \cap w'|\mathbf{x})$, then the task of computing the approximations of $w \cap w'$ and $w \cup w'$ is reduced to the estimation of the joint probabilities $P(w \cap w'|\mathbf{x})$ based on the given marginal probabilities $P(w|\mathbf{x})$ and $P(w'|\mathbf{x})$. There exist a number of methods to estimate the joint probabilities from the marginal ones (Lingras, Wong & Yao, 1990; Gokhale & Kullback, 1978; Ku & Kullback, 1969; Chow & Liu, 1968; Brown, 1959). The methods we adopt here for estimating $P(w \cap w'|\mathbf{x})$ are based on the principle of *maximum* entropy (Wise & Henrion, 1986; Shore & Johnson, 1980; Jaynes, 1979) and the principle of *minimum* entropy (Klir & Folger, 1988; Watanabe, 1985).

Let $\mathcal{P} = (p_1, \ldots, p_n)$ be a discrete joint probability distribution. The Shannon (1948) entropy function is defined by:

$$H(\mathcal{P}) = -\sum_{i=1}^{n} p_i \log p_i. \tag{3.2}$$

The entropy provides a measure of uncertainty conveyed by a given probability distribution. If the available information about $\mathcal{P}$ is insufficient, we can, for example, estimate the joint distribution by maximizing the entropy function under a set of constraints. The resulting probability distribution is said to be *maximally non-committal* or *minimally prejudiced* (Tribus, 1969). In other words, by the principle of maximum entropy, we adopt the most unbiased view in estimating the joint probabilities. In contrast, the principle of minimum entropy allows us to take the most biased view. Obviously, these two principles are not compatible with each other. The choice between them depends very much on the particular application. In any case, one cannot apply both principles simultaneously to the same problem.

To simplify subsequent discussions, let us define the following symbols:

$$\begin{array}{ll} a = P(w|\mathbf{x}), & b = P(w'|\mathbf{x}), \\ a_{11} = P(w \cap w'|\mathbf{x}), & a_{12} = P(w \cap \neg w'|\mathbf{x}), \\ a_{21} = P(\neg w \cap w'|\mathbf{x}), & a_{22} = P(\neg w \cap \neg w'|\mathbf{x}). \end{array} \tag{3.3}$$

The joint probabilities should satisfy the following constraints:

$$\begin{array}{ll} a_{11} + a_{12} = a & a_{21} + a_{22} = 1 - a \\ a_{11} + a_{21} = b & a_{12} + a_{22} = 1 - b \end{array} \tag{3.4}$$

These constraints can be conveniently represented by a contingency table:

	w'	$\neg w'$	
w	a_{11}	a_{12}	a
$\neg w$	a_{21}	a_{22}	$1-a$
	b	$1-b$	

For the probability distribution $\mathcal{P} = (a_{11}, a_{12}, a_{21}, a_{22})$, the entropy function is defined by:

$$H(\mathcal{P}) = -(a_{11} \log a_{11} + a_{12} \log a_{12} + a_{21} \log a_{21} + a_{22} \log a_{22}). \tag{3.5}$$

Based on the constraints (3.4), one can express all the other variables in terms of a_{11}, namely:

$$\begin{aligned} a_{12} &= a - a_{11}, \\ a_{21} &= b - a_{11}, \\ a_{22} &= 1 - (a+b) + a_{11}. \end{aligned} \tag{3.6}$$

By substituting these variables into equation (3.5), $H(\mathcal{P})$ can be written as:

$$\begin{aligned} H(\mathcal{P}) &= -a_{11} \log a_{11} - (a - a_{11}) \log(a - a_{11}) - (b - a_{11}) \log(b - a_{11}) - \\ & [1 - (a+b) + a_{11}] \log[1 - (a+b) + a_{11}]. \end{aligned} \tag{3.7}$$

From the probability theory, we have:

$$\max(0, P(w|\mathbf{x}) + P(w'|\mathbf{x}) - 1) \le P(w \cap w'|\mathbf{x}) \le \min(P(w|\mathbf{x}), P(w'|\mathbf{x})), \tag{3.8}$$

or

$$\max(0, a + b - 1) \le a_{11} \le \min(a, b). \tag{3.9}$$

That is, the interval $[\max(0, a + b - 1), \min(a, b)]$ defines the domain of the variable a_{11}.

By differentiating $H(\mathcal{P})$ with respect to a_{11} and setting it to zero, we obtain:

$$\frac{\mathrm{d}H(\mathcal{P})}{\mathrm{d}a_{11}} = \log \frac{(a - a_{11})(b - a_{11})}{a_{11}[1 - (a + b) + a_{11}]} = 0, \tag{3.10}$$

which gives the solution $a_{11} = ab$. Note that

$$\frac{\mathrm{d}H(\mathcal{P})}{\mathrm{d}a_{11}} > 0 \qquad \text{for} \quad a_{11} \in [\max(0, a + b - 1), ab), \tag{3.11}$$

$$\frac{\mathrm{d}H(\mathcal{P})}{\mathrm{d}a_{11}} < 0 \qquad \text{for} \quad a_{11} \in (ab, \min(a, b)]. \tag{3.12}$$

This means that the entropy function $H(\mathcal{P})$ defined by equation (3.7) increases in the interval $[\max(0, a+b-1), ab)$, decreases in $(ab, \min(a, b)]$, and obtains the maximum value at the point $a_{11} = ab$. It is therefore clear that the minimum of the entropy function occurs at one of the two boundary points, $\max(0, a + b - 1)$ and $\min(a, b)$.

Thus, according to the principle of maximum entropy, the joint conditional probability $a_{11} = P(w \cap w'|\mathbf{x})$ is equal to the product of $a = P(w|\mathbf{x})$ and $b = P(w'|\mathbf{x})$. The same result can be obtained by the probabilistic independence assumption. That is,

$$P(w \cap w'|\mathbf{x}) = P(w|\mathbf{x})P(w'|\mathbf{x}) , \tag{3.13}$$

$$\begin{aligned} P(w \cup w'|\mathbf{x}) &= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - P(w \cap w'|\mathbf{x}) \\ &= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - P(w|\mathbf{x})P(w'|\mathbf{x}) . \end{aligned} \tag{3.14}$$

The joint probabilities are summarized in the following table:

	w'	$\neg w'$	
w	ab	$a(1-b)$	a
$\neg w$	$(1-a)b$	$(1-a)(1-b)$	$1-a$
	b	$1-b$	

Recall that one can define a fuzzy set $\tilde{w}$ for a given concept w by the following membership function:

$$\mu_{\tilde{w}}(o) = P(w|\mathbf{x}_o), \tag{3.15}$$

where $\mathbf{x}_o$ denotes the description of an object o. Based on the estimation of the probabilities given by equations (3.13) and (3.14), and the fuzzy set membership function defined by equation (3.15), we have:

$$\begin{aligned}
\mu_{\tilde{w}\cap\tilde{w}'}(o) &= P(w \cap w'|\mathbf{x}_o) \\
&= P(w|\mathbf{x}_o)P(w'|\mathbf{x}_o) \\
&= \mu_{\tilde{w}}(o)\mu_{\tilde{w}'}(o) & (3.16)\\
\mu_{\tilde{w}\cup\tilde{w}'}(o) &= P(w \cup w'|\mathbf{x}_o) \\
&= P(w|\mathbf{x}_o) + P(w'|\mathbf{x}_o) - P(w|\mathbf{x}_o)P(w'|\mathbf{x}_o), \\
&= \mu_{\tilde{w}}(o) + \mu_{\tilde{w}'}(o) - \mu_{\tilde{w}}(o)\mu_{\tilde{w}'}(o). & (3.17)
\end{aligned}$$

We can therefore conclude that the principle of maximum entropy enables us to derive the *probabilistic-like* definition of fuzzy intersection and union (Bellman & Zadeh, 1970).

As mentioned before, in some applications one can use the principle of minimum entropy instead to estimate the joint probabilities. Suppose the minimum of entropy function occurs at the point $a_{11} = \max(0, a+b-1)$. In this case, we obtain:

$$\begin{aligned}
P(w \cap w'|\mathbf{x}) &= \max(0, P(w|\mathbf{x}) + P(w'|\mathbf{x}) - 1), & (3.18)\\
P(w \cup w'|\mathbf{x}) &= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - P(w \cap w'|\mathbf{x}) \\
&= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - \max(0, P(w|\mathbf{x}) + P(w'|\mathbf{x}) - 1) \\
&= \min(1, P(w|\mathbf{x}) + P(w'|\mathbf{x})). & (3.19)
\end{aligned}$$

The estimated joint probabilities are summarized in the following table:

	w'	$\neg w'$	
w	$\max(0, a+b-1)$	$\min(a, 1-b)$	a
$\neg w$	$\min(1-a, b)$	$\max(0, 1-a-b)$	$1-a$
	b	$1-b$	

Again, based on definition (3.15) and the above joint probabilities, one can define another version of fuzzy intersection and union as follows:

$$\begin{aligned} \mu_{\tilde{w}\cap\tilde{w}'}(o) &= P(w \cap w'|\mathbf{x}_o) \\ &= \max(0, P(w|\mathbf{x}_o) + P(w'|\mathbf{x}_0) - 1) \\ &= \max(0, \mu_{\tilde{w}}(o) + \mu_{\tilde{w}'}(o) - 1), \end{aligned} \tag{3.20}$$

$$\begin{aligned} \mu_{\tilde{w}\cup\tilde{w}'}(o) &= P(w \cup w'|\mathbf{x}_o) \\ &= \min(1, P(w|\mathbf{x}_o) + P(w'|\mathbf{x}_o)) \\ &= \min(1, \mu_{\tilde{w}}(o) + \mu_{\tilde{w}'}(o)). \end{aligned} \tag{3.21}$$

These are, in fact, the *bold* intersection and union suggested by Giles (1976).

Now suppose the entropy function has the minimum at the point $a_{11} = \min(a, b)$. It follows:

$$P(w \cap w'|\mathbf{x}) = \min(P(w|\mathbf{x}), P(w'|\mathbf{x})), \tag{3.22}$$

$$\begin{aligned} P(w \cup w'|\mathbf{x}) &= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - P(w \cap w'|\mathbf{x}) \\ &= P(w|\mathbf{x}) + P(w'|\mathbf{x}) - \min(P(w|\mathbf{x}), P(w'|\mathbf{x})) \\ &= \max(P(w|\mathbf{x}), P(w'|\mathbf{x})). \end{aligned} \tag{3.23}$$

The joint probabilities are summarized below:

	w'	$\neg w'$	
w	$\min(a, b)$	$\max(0, a-b)$	a
$\neg w$	$\max(0, b-a)$	$\min(1-a, 1-b)$	$1-a$
	b	$1-b$	

In this case, we obtain the classical definition of fuzzy intersection and union (Zadeh, 1965):

$$\begin{aligned} \mu_{\tilde{w}\cap\tilde{w}'}(o) &= P(w \cap w'|\mathbf{x}_o) \\ &= \min(P(w|\mathbf{x}_o), P(w'|\mathbf{x}_o)) \\ &= \min(\mu_{\tilde{w}}(o), \mu_{\tilde{w}'}(o)) \qquad (3.24) \\ \mu_{\tilde{w}\cup\tilde{w}'}(o) &= P(w \cup w'|\mathbf{x}_o) \\ &= \max(P(w|\mathbf{x}_o), P(w'|\mathbf{x}_o)) \\ &= \max(\mu_{\tilde{w}}(o), \mu_{\tilde{w}'}(o)) \qquad (3.25) \end{aligned}$$

By comparing the last two tables, it becomes evident that Giles' definition given by equations (3.20) and (3.21), and Zadeh's definition given by equations (3.24) and (3.25) are dual definitions.

Let us summarize the results obtained so far. We have demonstrated that the probabilistic-like definition of fuzzy intersection and union can be obtained by applying the maximum entropy principle, which represents the most unbiased view in estimating the joint probabilities. By applying the minimum entropy principle, we arrive at Giles' and Zadeh's definitions of fuzzy intersection and union. Both of these definitions represent the most biased view, but the emphases are different. The classical definition assumes that w and w' are most correlated, whereas the bold intersection and union assume the opposite (Wise & Henrion, 1986). Obviously, the probabilistic-like definition is a *median* between Giles' and Zadeh's definitions (Dubois & Prade, 1980).

It should perhaps be emphasized that our analysis presents a different view for interpreting fuzzy set connectives. The axiomatic approach for justifying Zadeh's definition is based on a mathematical structure (Fung & Fu, 1975; Bellman & Giertz, 1973). However, the physical interpretation of such a structure is not entirely clear (Giles, 1988). Another justification of fuzzy set connectives is based on the notion of triangular norms and conorms, which shows that all three definitions discussed in this paper are special cases (Bonissone, 1987; Bonissone & Decker, 1986; Dubois

& Prade, 1982). Such a justification does not identify the situations under which a particular definition is applicable, because each of these definitions satisfies the conditions for triangular norms and conorms. On the contrary, our analysis shows that the probabilistic-like definition follows naturally from the principle of maximum entropy representing the most unbiased view, while Zadeh's and Giles' definitions follow from the principle of minimum entropy representing the most biased views.

If the joint probabilities given by equation (3.13) or (3.18) are used, the resulting approximations $Apr(w \cap w', \alpha)$, $Apr(w \cup w', \alpha)$, $Apr(w \cap w', \alpha_+)$, and $Apr(w \cup w', \alpha_+)$ satisfy properties (O1)-(O2) as well; the approximations $\underline{Apr}(w \cap w', \beta, \delta)$, $\underline{Apr}(w \cup w', \beta, \delta)$, $\overline{Apr}(w \cap w', \beta, \delta)$, and $\overline{Apr}(w \cup w', \beta, \delta)$ satisfy properties (T1)-(T4). However, when the joint probabilities are computed from equation (3.22), the corresponding approximations satisfy the following properties instead:

$$\begin{aligned}
&(\mathrm{O}'1) \quad Apr(w \cap w', \alpha) = Apr(w, \alpha) \cap Apr(w', \alpha)\\
&(\mathrm{O}'2) \quad Apr(w \cup w', \alpha) = Apr(w, \alpha) \cup Apr(w', \alpha)\\
&(\mathrm{O}'3) \quad Apr(w \cap w', \alpha_+) = Apr(w, \alpha_+) \cap Apr(w', \alpha_+)\\
&(\mathrm{O}'4) \quad Apr(w \cup w', \alpha_+) = Apr(w, \alpha_+) \cup Apr(w', \alpha_+)
\end{aligned}$$

and

$$\begin{aligned}
&(\mathrm{T}'1) \quad \underline{Apr}(w \cap w', \beta, \delta) = \underline{Apr}(w, \beta, \delta) \cap \underline{Apr}(w', \beta, \delta)\\
&(\mathrm{T}'2) \quad \underline{Apr}(w \cup w', \beta, \delta) = \underline{Apr}(w, \beta, \delta) \cup \underline{Apr}(w', \beta, \delta)\\
&(\mathrm{T}'3) \quad \overline{Apr}(w \cap w', \beta, \delta) = \overline{Apr}(w, \beta, \delta) \cap \overline{Apr}(w', \beta, \delta)\\
&(\mathrm{T}'4) \quad \overline{Apr}(w \cup w', \beta, \delta) = \overline{Apr}(w, \beta, \delta) \cup \overline{Apr}(w', \beta, \delta)
\end{aligned}$$

This implies that the *inclusion* ($\subseteq$ or $\supseteq$) in (O1)-(O4) and (T1)-(T4) becomes an *equality* only if maximum correlation between w and w' is assumed. It should be noted that the properties (O′1)-(O′4) about α-cuts (Kandel, 1986; Dubois & Prade, 1980) hold only if one adopts Zadeh's definition of fuzzy intersection and union.

4. Conclusion

Yao, Y.Y. and Wong, S.K.M. A decision theoretic framework for approximating concepts,
International Journal of Man-machine Studies , Vol. 37, No. 6, pp. 793-809, 1992

We have proposed in this paper a decision theoretic framework for approximating concepts, which provides a plausible unification of the fuzzy set and rough set approaches. We have explicitly shown that if a given concept is approximated by one set, the same result given by the α-cut in the fuzzy set theory is obtained. On the other hand, if a given concept is approximated by two sets, we can derive both the algebraic and probabilistic rough set approximations. Moreover, based on the well known principle of maximum (minimum) entropy, we give a useful interpretation of the fuzzy intersection and union. We believe that our results enhance the understanding and broaden the applications of both fuzzy and rough sets.

References

Bellman, R.E. & Giertz, M. (1973). On the analytic formalism of theory of fuzzy set. *Information Science*, **5**, 149-157.

Bellman, R.E. & Zadeh, L.A. (1970). Decision-making in fuzzy environment. *Management Science,* **17**, 141-164.

Bonissone, P.P. (1987). Summarizing and propagating uncertain information with triangular norms. *International Journal of Approximate Reasoning*, **1**, 71-101.

Bonissone, P.P. & Decker, K.S. (1986). Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity. In L.N. Kanal & J.F. Lemmer, Eds., *Uncertainty in Artificial Intelligence*, 217-247, New York: North-Holland.

Brown, D.T. (1959). A note on approximations to discrete probability distributions. *Information and Control,* **2**, 386-392.

Chow, C.K. & Liu, C.N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory,* **IT-14**, 462-467.

Dubois, D. & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, **17**, 191-209.

Dubois, D. & Prade, H. (1987). Twofold fuzzy sets and rough sets – some issues in knowledge representation. *Fuzzy Sets and Systems,* **23**, 3-18.

Dubois, D. & Prade, H. (1982). A class of fuzzy measures based on triangular norms: a general framework for the combination of uncertain information. *International Journal of General Systems*, **8**, 43-61.

Dubois, D. & Prade, H. (1980). *Fuzzy Sets and Systems – Theory and Applications.* New York: Academic Press.

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

Fung, L.W. & Fu, S.K. (1975). An axiomatic approach to rational decision-making in a fuzzy environment. In L.A. Zadeh, K.S. Fu, K. Tanaka & M. Shimura, Eds., *Fuzzy Sets and Their Application to Cognitive and Decision Process*, 227-256, New York: Academic Press.

Gaines, B.R. (1978). Fuzzy and probability uncertainty logics. *Information and Control*, *38*, 154-169.

Giles, R. (1988). The concept of grade of membership. *Fuzzy Sets and Systems*, **25**, 297-323.

Giles, R. (1982). Semantics for fuzzy reasoning. *International Journal of Man-machine Studies*, **17**, 401-415.

Giles, R. (1976). Lukasiewicz logic and fuzzy theory. *International Journal of Man-Machine Studies,* **8**, 313-327.

Gokhale, D.V. & Kullback, S. (1978). *The Information in Contingency Tables.* New York: Marcel Dekker, Inc..

Hersh, H.M. & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General*, **105**, 254-276.

Hisdal, E. (1988a). Are grades of membership probabilities? *Fuzzy Sets and Systems*, **25**, 325-348.

Hisdal, E. (1988b). The philosophical issues raised by fuzzy set theory. *Fuzzy Sets and Systems*, **25**, 349-356.

Jaynes, E.T. (1979). Where do we stand on maximum entropy? In R.D. Levine & M. Tribus, Eds. *The Maximum Entropy Formalism.* Cambridge, Massachusetts: The MIT Press.

Kandel, A. (1986). *Fuzzy Mathematical Techniques with Applications.* Massachusetts: Addison-Wesley Publishing Company.

Klir, G. J. & Golger, T. A. (1988). *Fuzzy sets, Uncertainty, and Information.* Englewood Cliffs, New Jersey: Prentice Hall.

Ku, H.H. & Kullback, S (1969). Approximating discrete probability distributions. *IEEE Transactions on Information Theory,* **IT-15**, 444-447.

Lingras, P., Wong, S.K.M. & Yao, Y.Y. (1990). Probability estimation in intelligent information systems. *Proceedings of the IASTED International Conference, Expert Systems: Theory and Applications,* 49-51.

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

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

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information and Sciences,* **11**, 341-356.

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

Ruspini, E. (1969). A new approach to clustering. *Information and Control*, **8**, 338-353.

Shannon, C.E. (1948). The mathematical theory of communication. *The Bell System and Technical Journal,* **27**, 379-423, 623-656.

Shore, J.E. & Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory,* **IT-26**, 26-37.

Tribus, M. (1985). *Rational Descriptions, Decisions, and Designs.* Oxford: Pergamon.

Watanabe, S. (1985). *Pattern Recognition: Human and Mechanical.* New York: Wiley.

Wise, B.P. & Henrion, M. (1986). A framework for computing uncertain inference systems to probability. In L.N. Kanal & J.F. Lemmer, Eds. *Uncertainty in Artificial Intelligence,* 69-83, New York: North-Holland.

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

Wygralak, M. (1989). Rough sets and fuzzy sets – some remarks on interrelations. *Fuzzy Sets and Systems,* **29**, 241-243.

Yao, Y.Y., Wong, S.K.M. & Lingras, P. (1990). A decision-theoretic rough set model. In Z.W. Ras, M. Zemankova & M.L. Emrich, Eds. *Methodologies for Intelligent Systems, 5,* 17-24, New York: North-Holland.

Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3-28.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control,* **8**, 338-353.