Fuzzy-rough set and fuzzy ID3 decision approaches to knowledge discovery in datasets

A. S. Salama\textsuperscript{1} * O. G. Elbarbary \textsuperscript{1}

(1) Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.

1 Corresponding author. Email address: dr_asalamaa@yahoo.com Tel: +201091396647

Copyright 2012 ©A. S. Salama and O. G. Elbarbary. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract
Fuzzy rough sets are the generalization of traditional rough sets to deal with both fuzziness and vagueness in data. The existing researches on fuzzy rough sets mainly concentrate on the construction of approximation operators. Less effort has been put on the knowledge discovery in datasets with fuzzy rough sets. This paper mainly focuses on knowledge discovery in datasets with fuzzy rough sets. After analyzing the previous works on knowledge discovery with fuzzy rough sets, we introduce formal concepts of attribute reduction with fuzzy rough sets and completely study the structure of attribute reduction.

Keywords: Fuzzy Sets; Fuzzy Rough Sets; Rough Sets; Knowledge Discovery; Decision ID3.

1 Introduction

The concept of rough sets was originally proposed in [46, 47] as a mathematical tool for handling imprecision, vagueness and uncertainty in information systems. This theory has sufficiently been demonstrated to have its usefulness and adaptability in successfully solving a variety of problems [1, 2, 4, 5, 10, 11, 12, 13, 14, 16, 18, 21, 22]. One important application of rough set theory is that of attributes reduction in datasets. Given a dataset with coded attribute values, it is possible to find a subset of the original attributes that contains the same information as the original one. The concept of attribute reduction can be viewed as the strongest and most important results in rough set theory to discriminate itself from other theories. Rough sets approach of attribute reduction can be used as a purely structural method for self-dipping dimensionality using information contained within the dataset and preserving the meaning of the features. However, as mentioned in [9],
the values of attributes could be both of symbolic and real-valued. The conventional rough set theory will have difficulty in managing such real-valued attributes. One way to solve this problem is to code the attribute values previously [19, 20, 23], and create a new dataset with symbolic attribute values. However, information loss is found with this method. Another approach is the use of fuzzy rough sets. Fuzzy rough sets summarize the related but distinct concepts of fuzziness and indiscernibility, both occurring as a result of uncertainty existed in knowledge. In fuzzy rough sets [3, 6, 7, 8, 15, 17, 25] a fuzzy similarity relation is engaged to characterize the degree of similarity between two objects instead of the equivalence relation used in the crisp rough sets. The degree of similarity of two objects takes values on the unit interval. If the degree of similarity is 1, then they are indiscernible. They are discernible if the degree of similarity degree is 0. If the degree of similarity takes a value between 0 and 1, then these two objects are similar to a certain degree. However, different fuzzy similarity relation may construct different degree of similarity, so information loss is also possible. Awkward fuzzy similarity relations will bring large information loss. This problem could be handled by defining a reasonable fuzzy similarity relation.

In traditional set Theory, membership of an object belonging to a set can only be one of two values: 0 or 1. An object either belongs to a set completely or it does not belong at all. No partial membership is allowed. Crisp sets handle black and white concepts well, such as "chairs", "ships" and "trees", where little ambiguity exists. They are not sufficient, however, to realistically describe vague concept. In our daily lives, there are countless vague concepts leads to appearance of fuzzy information or fuzzy data. One of them, it may be due to the imprecision of real data. For instance, a sensor data may be a distribution, rather than a precise value. Second, fuzzy information can arise from subjective judgments. For example, a database containing information about real estate for family housing may need to describe the quality of public schools, the safety of the neighborhood, the estimated appreciation of housing price, and so on. Representing this information using precise values would often fail to capture the soft boundaries between qualitative descriptions such as poor, fair, good, excellent, etc. Thus; it is important to find a data model that can represent and manipulate fuzzy information. Third, the information that a user is interested in may not be more precise. For example, a college senior may be interested in finding a university that has a good graduate engineering program and low living costs. The meaning of "good" and "low" in the previous sentence is imprecise. Formulating this query using thresholds such as "an annual living cost is less than x dollars" will exclude those universities whose annual living cost is slightly above x, but whose graduate engineering school is excellent. In other words, representing an imprecise query using a precise formalism is likely to miss information that the user wishes to obtain.

The existing researches on fuzzy rough sets are mainly concentrated on the approximations of fuzzy sets. These researches have been studied and discussed completely in [26]-[41]. In [24] a pioneering work on attributes reduction with fuzzy rough sets is proposed. Formal concepts of fuzzy-rough attributes reduction were introduced and an algorithm to compute a reduction was developed by using the dependence function.
2 Fundamentals of rough sets models

Rough set theory [46, 47] is still a new approach to vagueness and data mining. Similarly to fuzzy set theory it is not an alternative to classical set theory but it is embedded in it. Rough set theory can be viewed as a specific implementation of vagueness, i.e., imprecision in this approach is expressed by a boundary region of a set, and not by a partial membership, like in fuzzy set theory.

Rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations.

Let us describe this problem more precisely. Suppose we are given a set of objects $U$ called the universe and an indiscernibility relation $R \subseteq U \times U$, representing our lack of knowledge about elements of $U$. For the sake of simplicity we assume that $R$ is an equivalence relation. Let $X$ be a subset of $U$. We want to characterize the set $X$ with respect to $R$. To this end we will need the basic concepts of rough set theory given below.

1. The lower approximation of a set $X$ with respect to $R$ is the set of all objects, which can be for certain classified as $X$ with respect to $R$ (are certainly $X$ with respect to $R$).
2. The upper approximation of a set $X$ with respect to $R$ is the set of all objects which can be possibly classified as $X$ with respect to $R$ (are possibly $X$ in view of $R$).
3. The boundary region of a set $X$ with respect to $R$ is the set of all objects, which can be classified neither as $X$ nor as not $X$ with respect to $R$.

Now we are ready to give the definition of rough sets.

1. Set $X$ is crisp (exact with respect to $R$), if the boundary region of $X$ is empty.
2. Set $X$ is rough (inexact with respect to $R$), if the boundary region of $X$ is nonempty.

Thus a set is rough (imprecise) if it has a nonempty boundary region; otherwise the set is crisp (precise).

The approximations and the boundary region can be defined more precisely. To this end we need some additional notation.

The equivalence class of $R$ determined by element $x$ will be denoted by $R(x)$. The indiscernibility relation in certain sense describes our lack of knowledge about the universe. Equivalence classes of the indiscernibility relation, called granules generated by $R$, represent an elementary portion of knowledge we are able to perceive due to $R$. Thus in view of the indiscernibility relation, in general, we are an able to observe individual objects but we are forced to reason only about the accessible granules of knowledge.

Formal definitions of approximations and the boundary region are as follows:

1. $R$-lower approximation of $X$, $\overline{R}(X) = \bigcup_{x \in U} \{R(x) : R(x) \subseteq X\}$.
2. $R$-upper approximation of $X$, $\overline{\overline{R}}(X) = \bigcup_{x \in U} \{R(x) : R(x) \cap X \neq \emptyset\}$.
3. $R$-boundary region of $X$, $RBND_R(X) = \overline{R}(X) - \overline{\overline{R}}(X)$. 
As we can see from the definition approximations are expressed in terms of granules of knowledge. The lower approximation of a set is the union of all granules which are entirely included in the set; the upper approximation is the union of all granules which have non-empty intersection with the set; the boundary region of the set is the difference between the upper and the lower approximation.

It is interesting to compare definitions of classical sets, fuzzy sets and rough sets. Classical set is a primitive notion and is defined intuitively or axiomatically. Fuzzy sets are defined by employing the fuzzy membership function, which involves advanced mathematical structures, numbers and functions. Rough sets are defined by approximations. Thus this definition also requires advanced mathematical concepts.

Approximations have the following properties:

1. \( R(X) \subseteq X \subseteq \overline{R}(X) \),
2. \( R(\phi) = \overline{R}(\phi) = \phi; R(U) = \overline{R}(U) = U \),
3. \( R(X \cup Y) = R(X) \cup R(Y) \),
4. \( R(X \cap Y) = R(X) \cap R(Y) \),
5. \( R(X \cup Y) \supseteq R(X) \cup R(Y) \),
6. \( \overline{R}(X \cap Y) \subseteq \overline{R}(X) \cap \overline{R}(Y) \),
7. \( X \subseteq Y \rightarrow R(X) \subseteq R(Y) \& \overline{R}(X) \subseteq \overline{R}(Y) \),
8. \( R(-X) = \overline{R}(X) \),
9. \( \overline{R}(-X) = -R(X) \),
10. \( R(\overline{R}(X)) = \overline{R}((\overline{R}(X)) = \overline{R}(X) \),
11. \( \overline{R}(\overline{R}(X)) = R(\overline{R}(X)) = \overline{R}(X) \).

It is easily seen that approximations are in fact interior and closure operations in a topology generated by data. Thus fuzzy set theory and rough set theory require completely different mathematical setting.

Rough sets can be also defined employing, instead of approximation, rough membership function \([46]\)

\[ \mu^R_X : U \rightarrow [0, 1] \text{ where } \mu^R_X(x) = \frac{|X \cap R(x)|}{|R(x)|} \text{ and } |X| \text{ denotes the cardinality of } X. \]

The rough membership function expresses conditional probability that \( x \) belongs to \( X \) given \( R \) and can be interpreted as a degree that \( x \) belongs to \( X \) in view of information about \( x \) expressed by \( R \).

The rough membership function can be used to define approximations and the boundary region of a set, as shown below:

\[ \underline{R}(X) = \{ x \in U : \mu^R_X(x) = 1 \}, \]
\[ \overline{R}(X) = \{ x \in U : \mu^R_X(x) > 0 \}, \]
\[ RBNDR(X) = \{ x \in U : 0 < \mu^R_X(x) < 1 \}. \]

It can be shown that the membership function has the following properties \([5]\):
1. \( \mu_R^X(x) = 1 \iff x \in R(X) \),
2. \( \mu_R^X(x) = 0 \iff x \in U - R(X) \),
3. \( 0 < \mu_R^X(x) < 1 \iff x \in R(BND_R(X)) \),
4. \( \mu_{R - X}^U(x) = 1 - \mu_R^X(x) \) for any \( x \in U \),
5. \( \mu_{R \cup Y}^X(x) \geq \max(\mu_R^X(x), \mu_Y^X(x)) \) for any \( x \in U \),
6. \( \mu_{R \cap Y}^X(x) \leq \min(\mu_R^X(x), \mu_Y^X(x)) \) for any \( x \in U \).

From the above properties it follows that the rough membership differs essentially from the fuzzy membership. For instance properties (5) and (6) show that the membership for union and intersection of sets, in general, cannot be computed – as in the case of fuzzy sets – from their constituent membership. Thus formally the rough membership is a generalization of fuzzy membership. Besides, the rough membership function, in contrast to fuzzy membership function, has a probabilistic flavor.

Now we can give two definitions of rough sets. Set \( X \) is rough with respect to \( R \) if \( R(X) \neq \emptyset \).
Set \( X \) rough with respect to \( R \) if for some \( x \), \( 0 < \mu_R^X(x) < 1 \).

### 3 Basics of fuzzy set models

Zadeh proposed a new approach to vagueness called fuzzy set theory [48]. In his approach an element can belong to a set to a degree \( k \) (\( 0 \leq k \leq 1 \)), in contrast to classical set theory where an element must definitely belong or not to a set. E.g., in classical set theory one can be definitely ill or healthy, whereas in fuzzy set theory we can say that someone is ill (or healthy) in 60 percent (i.e., in the degree 0.6). Of course, at once the question arises where we get the value of degree from. This issue raised a lot of discussion, but we will refrain from considering this problem here.

Thus the fuzzy membership function can be presented as: \( \mu_X^U(x) \in [0,1] \) where, \( X \) is a set and \( x \) is an element.

Let us observe that the definition of fuzzy set involves more advanced mathematical concepts, real numbers and functions, whereas in classical set theory the notion of a set is used as a fundamental notion of whole mathematics and is used to derive any other mathematical concepts, e.g., numbers and functions. Consequently fuzzy set theory cannot replace classical set theory, because, in fact, the theory is needed to define fuzzy sets.

The fuzzy membership function has the following properties.

1. \( \mu_{U - X}(x) = 1 - \mu_X^U(x) \) for any \( x \in U \),
2. \( \mu_{X \cup Y}^U(x) = \max(\mu_X^U(x), \mu_Y^U(x)) \) for any \( x \in U \),
3. \( \mu_{X \cap Y}^U(x) = \min(\mu_X^U(x), \mu_Y^U(x)) \) for any \( x \in U \).

That means that the membership of an element to the union and intersection of sets is uniquely determined by its membership to constituent sets. This is a very nice property and allows very simple operations on fuzzy sets, which is a very important feature both theoretically and practically.
Fuzzy set theory and its applications developed very extensively over the last years and attracted attention of practitioners, logicians and philosophers worldwide.

We use a capital letter and tilde (e., $\tilde{A}$) to represent a fuzzy set in this thesis. If an element is denoted by $x \in X$, where $X$ is the universe of discourse, the membership function of the fuzzy set $\tilde{A}$ is mathematically expressed as $\mu_{\tilde{A}}(x)$, $\mu_{\tilde{A}}$ or simply $\mu$.

For the above age example, $X = [0, 130]$ letting $\tilde{A}$ denote fuzzy set ”young”, we can represent its membership function by $\mu_{\tilde{A}}(x)$, where $x \in X$.

The most basic operators on fuzzy sets [2] are the union, intersection and complement. These are fuzzy extensions of their crisp counterparts, ensuring that if they are applied to the crisp sets, the results of their application will be identical to crisp union, intersection and complement.

The intersection (t-norm) of two fuzzy sets, $\tilde{A}$ and $\tilde{B}$, is specified by a binary operation on the unit interval; that is a function of the form:

For each element $x$ in the universe, this function takes as its arguments the memberships of $x$ in the fuzzy sets $\tilde{A}$ and $\tilde{B}$, and yields the membership grade of the element in the set constituting the intersection of $\tilde{A}$ and $\tilde{B}$:

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = t[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)].$$

The following axioms must hold for the operator $t$ to be considered a $t$-norm for all $x$, $y$ and $z$ in the range $[0, 1]$:

1. $t(x, 1) = x$ (Boundary condition).
2. $y \leq z \rightarrow t(x, y) \leq t(x, z)$ (Monotonicity).
3. $t(x, y) = t(y, x)$ (Commutatively).
4. $t(x, t(y, z)) = t(t(x, y), z)$ (Associatively).

The following are examples of $t$-norm that are often used as fuzzy intersections:

1. $t(x, y) = \min(x, y)$ (Standard intersection).
2. $t(x, y) = x.y$ (Algebraic product).
3. $t(x, y) = \max(0, x + y - 1)$ (Bounded difference).

The fuzzy union ($t$-co-norm or $s$-norm) of two fuzzy sets $\tilde{A}$ and $\tilde{B}$ is specified by a function:

$$S : [0, 1] \times [0, 1] \rightarrow [0, 1]$$

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = S[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)].$$

A fuzzy union is a binary operation that satisfies at least the following axioms for all $x$, $y$ and $z$ in $[0, 1]$.

1. $S(x, 0) = x$(boundary condition)
2. $y \leq z \rightarrow S(x, z) \leq S(x, z)$(monotonic)
3. $S(x, y) = S(y, x)$(commutative)
4. \( S(x, S(y, z)) = S(S(x, y), Z) \) (associative)

The following are examples of \( t \)-co-norms that are often used as fuzzy unions:

1. \( S(x, y) = \max(x, y) \) (Standard union)
2. \( S(x, y) = x + y - xy \) (Algebraic sum)
3. \( S(x, y) = \min(1, x + y) \) (Bounded sum)

The most popular interpretation of fuzzy union and intersection is the max/min interpretation, primarily due to its ease of computation.

In classical set theory, there are binary logic operators AND (i.e., Intersection), OR (i.e., Union), NOT (I.e., Complement), and so on. The corresponding fuzzy logic operators exist in fuzzy set theory. Fuzzy logic AND and OR operations are used in fuzzy controllers and models unlike the binary AND and OR operators whose operations are uniquely defined, their fuzzy counterparts are non unique. Numerous fuzzy logic AND operators and OR operators have been proposed, some of them purely from the mathematical point of view. To a large extent, the Zadeh fuzzy AND operator, product fuzzy AND operator, the Zadeh OR operator and the Lukasiewicz OR operator have been found to be most useful for fuzzy control and modeling. Their definitions are as follows:

1. Zadeh fuzzy logic AND operator: \( \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \).
2. Product fuzzy logic AND operator: \( \mu_{A \cap B}(x) = \mu_A(x) \times \mu_B(x) \).
3. Zadeh fuzzy logic OR operator: \( \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \).
4. Lukasiewicz fuzzy logic OR operator: \( \mu_{A \cup B}(x) = \min(\mu_A(x) + \mu_B(x), 1) \).

Fuzzification [48] is a mathematical procedure for converting an element in the universe of discourse into the membership value of the fuzzy set. Suppose that fuzzy set \( \tilde{A} \) is defined on \([a, b]\) that is, the universe of discourse is \([a, b]\); for any \( x \in [a, b] \), the result of fuzzification is simply \( \mu_{\tilde{A}}(x) \).

Defuzzification [44] is a mathematical process used to convert a fuzzy set or fuzzy sets to a real number. It is a necessary step because fuzzy sets generated by fuzzy inference in fuzzy rules must be somehow mathematically combined to come up with one single number as the output of a fuzzy controller or model.

In the same way that crisp equivalence classes are central to rough sets, fuzzy equivalence classes are central to the fuzzy-rough set approach [46, 47]. For typical RSAR applications, this means that the decision values and the conditional values may all be fuzzy. The concept of crisp equivalence classes can be extended by the inclusion of a fuzzy similarity relation \( S \) on the universe, which determines the extent to which two elements are similar in \( S \). For example, if \( \mu_S(x, y) = 0.9 \), then objects \( x \) and \( y \) are considered to be quite similar. The usual properties of reflexivity (\( \mu_S(x, x) = 1 \)), symmetry (\( \mu_S(x, y) = \mu_S(y, x) \)) and transitivity (\( \mu_S(x, z) \geq \mu_S(x, y) \land \mu_S(y, z) \)) hold.

Using the fuzzy similarity relation, the fuzzy equivalence class \([x]_S \) for objects close to \( x \) can be defined:

\[ \mu_{[x]_S}(y) = \mu_S(x, y). \]

The following axioms should hold for a fuzzy equivalence class \( F \) [8]:

7
1. \( \exists x, \mu_F(x) = 1, \)

2. \( \mu_F(x) \land \mu_S(x, y) \leq \mu_F(y), \)

3. \( \mu_F(x) \land \mu_F(y) \leq \mu_S(x, y). \)

The first axiom corresponds to the requirement that an equivalence class is non-empty. The second axiom states that elements in \( y \)'s neighborhood are in the equivalence class of \( y \). The final axiom states that any two elements in \( F \) are related via \( S \) obviously, this definition degenerates to the normal definition of equivalence classes when \( S \) is non-fuzzy.

The family of normal fuzzy sets produced by a fuzzy partitioning of the universe of discourse can play the role of fuzzy equivalence classes [5]. Consider the crisp partitioning of a universe, \( U \) by the attributes in \( Q : U/Q = \{1,3,6\}, \{2,4,5\} \). This contains two equivalence classes namely \{1,3,6\} and \{2,4,5\} that can be thought of as degenerated fuzzy sets, with those elements belonging to the class possessing a membership of one, zero or otherwise. For the first class for instance, the objects 2,4 and 5 have a membership of zero. Extending this to the case of fuzzy equivalence classes is straightforward: objects can be allowed to assume membership values, with respect to any given class, in the interval \([0,1]\). \( U/Q \) is not restricted to crisp partitions only; fuzzy partitions are equally acceptable [9].

4 Fuzzy-rough set model

The RSAR process described previously can only operate effectively with datasets containing discrete values. Additionally, there is no way of handling noisy data. As most datasets contain real-valued attributes, it is necessary to perform a discrimination step beforehand. This is typically implemented by fuzzification techniques [10], enabling linguistic labels to be associated with attribute values. It also aids uncertainty modeling by allowing the possibility of the membership of a value to more than one fuzzy label. However, membership degrees of attribute values to fuzzy sets are not exploited in the process of dimensionality reduction. By using fuzzy -rough sets [5, 11], it is possible to use this information to better guide attribute selection.

From the literatures, the fuzzy P-lower and P-upper approximations are defined as [5]:

\[
\mu_{P_X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \forall i
\]

\[
\mu_{\overline{P}_X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \forall i.
\]

Where \( F_i \) denotes a equivalence class belonging to \( U/P \). Note that although the universe of discourse in attribute selection is finite, this is not the case in general, hence the use of \( \sup \) and \( \inf \). These definitions diverge a little from the crisp upper and lower approximations, as the memberships of individual objects to the approximations are not explicitly available. As a result of this, the fuzzy lower and upper approximations are in redefined as:

\[
\mu_{\overline{P}_X}(x) = \sup_{F \in U/P} \min(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_{F_i}(y), \mu_X(y)\})
\]

\[
\mu_{P_X}(F_i) = \sup_{F \in U/P} \min(\mu_{F_i}(x), \sup_{y \in U} \min(\mu_F(y), \mu_X(y)))
\]
In implementation, not only all \( y \in U \) need to be considered only those where \( \mu_F(y) \) is non-zero, i.e. where object \( y \) is a fuzzy member of (fuzzy)equivalence class \( F \). The pair \(< PX, PX >\) is called a fuzzy-rough set. For this particular attribute selection method, the upper approximation is not used, thought this may be useful for other methods.

It can be seen that these definitions degenerate to traditional rough sets when all equivalence classes are crisp. It is useful to think of the crisp lower approximation as characterized by the following membership function:

\[
\mu_{PX}(x) = \begin{cases} 
1, & x \in F, F \subseteq X \\
0, & \text{otherwise}
\end{cases}
\]

This states that an object \( x \) belongs to the \( P \)-lower approximation of \( X \) if it belongs to an equivalence class that is a subset of \( X \). Obviously, the behavior of the fuzzy lower approximation must be exactly that of the crisp definition for crisp situations. This is indeed the case as the fuzzy lower approximation may be rewritten as:

\[
\mu_{PX}(x) = \sup_{F \in U/P} \min(\mu_F(x), \inf_{y \rightarrow \mu_X(y)} \mu_F(y)).
\]

Where \( \rightarrow \) stands for fuzzy implication (using the conventional min-max interpretation). In the crisp case \( \mu_{F(x)} \) and \( \mu_X(x) \) will take values from \{0,1\}. Hence, it is clear that the only time \( \mu_{PX}(x) \) will be zero is when at least one object in its equivalence class \( F \) fully belongs to \( F \) but not to \( X \). This is exactly the same as the definition for the crisp lower approximation. Similarly, the definition for the \( P \)-upper approximation can be established.

Fuzzy-Rough set-based Attribute Selection (FRAS) builds on the notion of fuzzy lower approximation to enable reduction of datasets containing real valued attributes. As will be shown, the process becomes identical to the crisp approach when dealing with nominal well-defined attribute.

The crisp positive region in traditional rough set theory is defined as the union of the lower approximations. By the extension principal [12], the membership of an object \( x \in U \), belonging to the fuzzy positive region can be defined by \( \mu_{POS(P)(Q)}(x) = \sup_{X \in U/Q} \mu_{PX}(x) \).

Object \( x \) will not belong to the positive region only if the equivalence class it belongs to is not a constituent of the positive region. This is equivalent to the crisp version where objects belong to the positive region only if their underlying equivalence class does so. Similarly, the negative and boundary regions can be defined.

Using the definition of the fuzzy positive region, the new dependency function can be defined as follows:

\[
\gamma_P(Q) = \frac{\mid \mu_{POS(P)(Q)}(x) \mid}{\mid U \mid} = \frac{\sum_{x \in U} \mu_{POS(P)(Q)}(x)}{\mid U \mid}.
\]

As with crisp rough sets, the depending of \( Q \) on \( P \) is the proportion of objects that are discernible out of the entire dataset. In the present approach, this corresponds to determining the fuzzy cardinality of \( \mu_{POS(P)(Q)}(x) \) divided by the total number of objects in the universe.

If a function \( \mu_{POS(P)(Q)}(x) \) is defined which returns 1 if the object \( x \) belongs to the positive region, 0 otherwise, then the above definition may be rewritten as:

\[
\gamma_P(Q) = \frac{\sum_{x \in U} \mu_{POS(P)(Q)}(x)}{\mid U \mid}.
\]
If the fuzzy rough reduction process is to be useful, it must be able to deal with multiple attribute finding the dependency between various subsets of the original attribute set. For example, it may be necessary to be able to determine the degree of dependency of the decision attribute with respect to \( P = \{a, b\} \). In the crisp case, \( U/P \) contains sets of objects grouped together that are indiscernible according to both attributes \( a \) and \( b \). In the fuzzy case, objects may belong to many equivalence classes, so the Cartesian product of \( U/IND(\{a\}) \) and \( U/IND(\{b\}) \) must be considered in determining \( U/P \). In general, \( U/P = \otimes \{a \in P : U/IND(\{a\})\} \) where \( A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \).

Each set in \( U/P \) denotes an equivalence class. For example, if \( P = \{a, b\} \), \( U/IND(\{a\}) = \{N_a, Z_a\} \) and \( U/IND(\{b\}) = \{N_b, Z_b\} \), then \( U/P = \{N_a \cap N_b, N_a \cap Z_b, Z_a \cap N_b, Z_a \cap Z_b\} \). The extent to which an object belongs to such an equivalence class is therefore calculated by using the conjunction of constituent fuzzy equivalence classes, say \( F_i, i = 1, 2, ..., n; \mu_{F_1} \cap \mu_{F_n}(x) = \min(\mu_{F_1}(x), \mu_{F_2}(x), ..., \mu_{F_n}(x)) \).

A problem may arise when this approach is compared to the crisp approach. In conventional RSAR, a reduct is defined as a subset \( R \) of the attributes which have the same information content as the full attribute set \( A \). In terms of the dependency function this means that the values \( \gamma(R) \) and \( \gamma(A) \) are identical and equal to 1 if the dataset is consistent. However, in the fuzzy-rough approach this is not necessarily the case as the uncertainty encountered when objects belong to many fuzzy equivalence classes results in a reduced total dependency.

A possible way of combating this would be to determine the degree of dependency of a set of decision attribute \( D \) upon the full feature set and use this as the denominator rather than \( |U| \), allowing \( \gamma \) to reach 1. With this issues in mind, a new \textit{QUICKREDUCT} algorithm has been developed. It employs the new dependency function \( \gamma' \) to choose which features to add to current reduct candidate in the same way as the original \textit{QUICKREDUCT}
process. The algorithm terminates when the addition of any remaining attribute does not increase the dependency.

To illustrate the operation of FRAS, we presenting the following example and apply the FRAS on it. In crisp RSAR, the dataset would be discredited using the non-fuzzy sets. However, in the fuzzy-rough approach membership degrees are used in calculating the fuzzy lower approximations and fuzzy positive regions. But it is not easy to replace a continuous domain with a discrete one. This requires some partition and clustering. It is also very difficult to define the boundary of the continuous attributes, for our example, how we define weather the traffic-jam is long or short? Can we say that the traffic-jam of 3 Km is long, and 2.9 Km is short? Can we say it is cool when the weather is 9, and it is mild for 10? Table 2 is the fuzzy representation of the sample data for the Table 1.

<table>
<thead>
<tr>
<th>Day</th>
<th>Weather</th>
<th>Rain</th>
<th>Traffic-jam</th>
<th>Go to zoo</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>32</td>
<td>3</td>
<td>7.5</td>
<td>No</td>
</tr>
<tr>
<td>D2</td>
<td>33</td>
<td>4.5</td>
<td>6.8</td>
<td>No</td>
</tr>
<tr>
<td>D3</td>
<td>30</td>
<td>2.5</td>
<td>8.3</td>
<td>Yes</td>
</tr>
<tr>
<td>D4</td>
<td>24</td>
<td>1.5</td>
<td>9</td>
<td>Yes</td>
</tr>
<tr>
<td>D5</td>
<td>3</td>
<td>2.5</td>
<td>3.8</td>
<td>Yes</td>
</tr>
<tr>
<td>D6</td>
<td>1</td>
<td>5</td>
<td>4.2</td>
<td>No</td>
</tr>
<tr>
<td>D7</td>
<td>8</td>
<td>4</td>
<td>2.7</td>
<td>Yes</td>
</tr>
<tr>
<td>D8</td>
<td>12</td>
<td>3</td>
<td>6.7</td>
<td>No</td>
</tr>
<tr>
<td>D9</td>
<td>5</td>
<td>2</td>
<td>3.5</td>
<td>Yes</td>
</tr>
<tr>
<td>D10</td>
<td>12</td>
<td>2.5</td>
<td>4.1</td>
<td>Yes</td>
</tr>
<tr>
<td>D11</td>
<td>15</td>
<td>6</td>
<td>2.3</td>
<td>Yes</td>
</tr>
<tr>
<td>D12</td>
<td>22</td>
<td>5</td>
<td>7.3</td>
<td>Yes</td>
</tr>
<tr>
<td>D13</td>
<td>32</td>
<td>2.5</td>
<td>2.6</td>
<td>Yes</td>
</tr>
<tr>
<td>D14</td>
<td>25</td>
<td>4</td>
<td>10.3</td>
<td>No</td>
</tr>
</tbody>
</table>

The following is the membership functions for the attributes weather, rain and traffic-jam.

We know that in most areas, the space of the temperature factor $x$ is approximately between -50 and 50. Then the membership functions of the fuzzy set hot, mild and cool $\mu_x$ may be defined separately as follows:

Attribute weather $\mu_{cool}(x) = \begin{cases} 
1 & x < 0 \\
1 - \frac{x}{15} & 0 \leq x \leq 15 \\
0 & x > 15 
\end{cases}$,  

$\mu_{mild}(x) = \begin{cases} 
0 & x < 5 \\
\frac{x}{15} - \frac{1}{3} & 5 \leq x \leq 20 \\
1 & 20 \leq x \leq 30 \\
-\frac{x}{5} + 7 & 30 \leq x \leq 35 \\
0 & x > 35 
\end{cases}$, $\mu_{hot}(x) = \begin{cases} 
0 & x > 25 \\
\frac{x}{15} - 2.5 & 25 \leq x \leq 35 \\
1 & x > 35 
\end{cases}$.
Also we can define the membership functions of the attribute rain as follows:

\[
\mu_{\text{weak}}(x) = \begin{cases} 
1 & x < 3 \\
2.5 - \frac{x}{2} & 0 \leq x \leq 15 \\
0 & x > 15
\end{cases} \quad , \quad \mu_{\text{strong}}(x) = \begin{cases} 
0 & x < 3 \\
\frac{x}{5} - 0.6 & 3 \leq x \leq 8 \\
1 & x > 8
\end{cases}
\]

Also we can define the membership functions of the attribute traffic-jam as follows:

\[
\mu_{\text{short}}(x) = \begin{cases} 
1 & x < 3 \\
1.5 - \frac{x}{6} & 3 \leq x \leq 9 \\
0 & x > 9
\end{cases} \quad , \quad \mu_{\text{long}}(x) = \begin{cases} 
0 & x < 5 \\
\frac{x}{10} - 0.5 & 5 \leq x \leq 15 \\
1 & x > 15
\end{cases}
\]

As the example above, we have partitioned the sample set into different intervals. The partition is complete (each domain value belongs to at least one subset) and inconsistent domain value can be found in more than one subset.

**Example Traffic-jam**

1. If traffic-jam is 3Km, the value of membership function long is null.

2. If traffic-jam is 3Km, the value of the membership function short is one.

From the previous membership functions for all attributes, we can calculate the membership values for all datasets and it is recorded in Table 2 below.

<table>
<thead>
<tr>
<th>Weather</th>
<th>Rain</th>
<th>Traffic-jam</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hot</td>
<td>Mild</td>
<td>Cool</td>
</tr>
<tr>
<td>D1</td>
<td>0.7</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>D2</td>
<td>0.8</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>D3</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>D8</td>
<td>0</td>
<td>0.47</td>
<td>1</td>
</tr>
<tr>
<td>D9</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D10</td>
<td>0</td>
<td>0.47</td>
<td>1</td>
</tr>
<tr>
<td>D11</td>
<td>0</td>
<td>0.47</td>
<td>0</td>
</tr>
<tr>
<td>D12</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D13</td>
<td>0.7</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>D14</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sum</td>
<td>2.7</td>
<td>7.41</td>
<td>6</td>
</tr>
</tbody>
</table>

Using fuzzy sets and setting \(\text{weather} = \{\text{hot, mild, cool}\}\), \(\text{rain} = \{\text{strong, weak}\}\), \(\text{traffic-jam} = \{\text{long, short}\}\) and \(\text{D} = \{\text{a, b}\}\), the following equivalence classes are obtained:

\(U/A = \{\text{hot, mild, cool}\}, U/B = \{\text{strong, weak}\}, U/C = \{\text{long, short}\}\) and \(U/Q = \{\text{a, b}\}\), where \(a = \{1, 2, 6, 8, 14\}\) and \(b = \{3, 4, 5, 7, 9, 10, 11, 12, 13\}\).
The first step is to calculate the lower approximations of the sets $A$, $B$ and $C$ using the equation: $\mu_A(x) = \sup_{F \in U/Q} \min(\mu_F(x), \inf_{Y \subseteq U} \max\{1 - \mu_F(y), \mu_A(y)\})$ and using the membership function in the last table using $A$ to approximate $Q$, for the two decision values $Yes$ and $No$. For the first decision value $No$ equivalence class $X = \{1, 2, 6, 8, 14\}$ and then $\mu_{AX}(x) = \sup_{F \in U/Q} \min(\mu_F(x), \inf_{Y \subseteq U} \max\{1 - \mu_F(y), \mu_X(y)\})$ must be calculated. Consider the first fuzzy equivalence class of $A$ (the hot class) then we need to calculate $\min(\mu_{hot}(x), \inf_{Y \subseteq U} \max\{1 - \mu_{hot}(y), \mu_X(y)\})$ for all objects. Then from Table 2 we can calculate the following:

$$\max \{1 - \mu_{hot}(1), \mu_X(1)\} = \max(0.3, 1) = 1,$$
$$\max \{1 - \mu_{hot}(2), \mu_X(2)\} = \max(0.2, 1) = 1,$$
$$\max \{1 - \mu_{hot}(3), \mu_X(3)\} = \max(0.5, 0) = 0.5,$$
$$\max \{1 - \mu_{hot}(4), \mu_X(4)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(5), \mu_X(5)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(6), \mu_X(6)\} = \max(1, 1) = 1,$$
$$\max \{1 - \mu_{hot}(7), \mu_X(7)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(8), \mu_X(8)\} = \max(1, 1) = 1,$$
$$\max \{1 - \mu_{hot}(9), \mu_X(9)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(10), \mu_X(10)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(11), \mu_X(11)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(12), \mu_X(12)\} = \max(1, 0) = 1,$$
$$\max \{1 - \mu_{hot}(13), \mu_X(13)\} = \max(0.3, 0) = 0.3,$$
$$\max \{1 - \mu_{hot}(14), \mu_X(14)\} = \max(1, 1) = 1.$$

From the calculation above the smallest value is 0.3 hence:

$$\min(\mu_{hot}(x), \inf_{Y \subseteq U} \max\{1 - \mu_{hot}(y), \mu_X(y)\}) = \min(\mu_{hot}(x), 0.3)$$

which calculated again as follows:

$$\min(\mu_{hot}(1), 0.3) = \min(0.7, 0.3) = 0.3,$$
$$\min(\mu_{hot}(2), 0.3) = \min(0.8, 0.3) = 0.3,$$
$$\min(\mu_{hot}(3), 0.3) = \min(0.5, 0.3) = 0.3,$$
$$\min(\mu_{hot}(4), 0.3) = \min(0, 0.3) = 0,$$
$$\min(\mu_{hot}(5), 0.3) = \min(0, 0.3) = 0,$$
$$\min(\mu_{hot}(6), 0.3) = \min(0, 0.3) = 0,$$
$$\min(\mu_{hot}(7), 0.3) = \min(0, 0.3) = 0,$$
$$\min(\mu_{hot}(8), 0.3) = \min(0, 0.3) = 0,$$
\[
\min(\mu_{\text{hot}}(9), 0.3) = \min(0, 0.3) = 0,
\]
\[
\min(\mu_{\text{hot}}(10), 0.3) = \min(0, 0.3) = 0,
\]
\[
\min(\mu_{\text{hot}}(11), 0.3) = \min(0, 0.3) = 0,
\]
\[
\min(\mu_{\text{hot}}(12), 0.3) = \min(0, 0.3) = 0,
\]
\[
\min(\mu_{\text{hot}}(13), 0.3) = \min(0.7, 0.3) = 0.3,
\]
\[
\min(\mu_{\text{hot}}(14), 0.3) = \min(0, 0.3) = 0.
\]

Similarly for the fuzzy equivalence class mild of \(A\) then we need to calculate
\[
\min(\mu_{\text{mild}}(x), \inf_{Y \subseteq U} \max \{1 - \mu_{\text{mild}}(y), \mu_X(y)\})
\]
for all objects. Then from Table 2 we can calculate the following:
\[
\max \{1 - \mu_{\text{mild}}(1), \mu_X(1)\} = \max(0.4, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(2), \mu_X(2)\} = \max(0.6, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(3), \mu_X(3)\} = \max(0, 0) = 0,
\]
\[
\max \{1 - \mu_{\text{mild}}(4), \mu_X(4)\} = \max(0, 0) = 0,
\]
\[
\max \{1 - \mu_{\text{mild}}(5), \mu_X(5)\} = \max(1, 0) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(6), \mu_X(6)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(7), \mu_X(7)\} = \max(0.8, 0) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(8), \mu_X(8)\} = \max(0.53, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(9), \mu_X(9)\} = \max(1, 0) = 1,
\]
\[
\max \{1 - \mu_{\text{mild}}(10), \mu_X(10)\} = \max(0.53, 0) = 0.53,
\]
\[
\max \{1 - \mu_{\text{mild}}(11), \mu_X(11)\} = \max(0.33, 0) = 0.33,
\]
\[
\max \{1 - \mu_{\text{mild}}(12), \mu_X(12)\} = \max(0, 0) = 0,
\]
\[
\max \{1 - \mu_{\text{mild}}(13), \mu_X(13)\} = \max(0.4, 0) = 0.4,
\]
\[
\max \{1 - \mu_{\text{mild}}(14), \mu_X(14)\} = \max(0, 1) = 1.
\]

From the calculation above the smallest value is 0 then :
\[
\min(\mu_{\text{mild}}(x), \inf_{Y \subseteq U} \max \{1 - \mu_{\text{mild}}(y), \mu_X(y)\}) = 0 , \text{ also}
\]
\[
\min(\mu_{\text{cool}}(x), \inf_{Y \subseteq U} \max \{1 - \mu_{\text{cool}}(y), \mu_X(y)\}) = 0.
\]

Then the lower approximations of the set \(A\) for the decision value \(No\) are:
\[
\mu_{\Delta_{\{1,2,6,8,14\}}}(1) = 0.3,
\]
\[
\mu_{\Delta_{\{1,2,6,8,14\}}}(8) = 0,
\]
\[
\mu_{\Delta_{\{1,2,6,8,14\}}}(2) = 0.3,
\]
\[
\mu_{\Delta_{\{1,2,6,8,14\}}}(9) = 0,
\]
\[
\mu_{\Delta_{\{1,2,6,8,14\}}}(3) = 0.3,
\]
\[\mu_{A(1,2,6,8,14)}(10) = 0,\]
\[\mu_{A(1,2,6,8,14)}(4) = 0,\]
\[\mu_{A(1,2,6,8,14)}(11) = 0,\]
\[\mu_{A(1,2,6,8,14)}(5) = 0.3,\]
\[\mu_{A(1,2,6,8,14)}(12) = 0,\]
\[\mu_{A(1,2,6,8,14)}(6) = 0,\]
\[\mu_{A(1,2,6,8,14)}(13) = 0.3,\]
\[\mu_{A(1,2,6,8,14)}(7) = 0,\]
\[\mu_{A(1,2,6,8,14)}(14) = 0.\]

Now we need to calculate the lower approximations of the set \(A\) for the second decision value \(\text{Yes}\) equivalence class \(X = \{3, 4, 5, 7, 9, 10, 11, 12, 14\}\) and then \(\mu_{AX}(x) = \sup_{F \in U \cap Q} \min(\mu_{F}(x), \inf_{Y \subset U} \max \{1 - \mu_{F}(y), \mu_{X}\})\) must be calculated.

Then from Table 2 we can calculate the following:

\[
\max \{1 - \mu_{\text{hot}}(1), \mu_{X}(1)\} = \max(0.3, 0) = 0,
\]
\[
\max \{1 - \mu_{\text{hot}}(2), \mu_{X}(2)\} = \max(0.2, 0) = 0.2,
\]
\[
\max \{1 - \mu_{\text{hot}}(3), \mu_{X}(3)\} = \max(0.5, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(4), \mu_{X}(4)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(5), \mu_{X}(5)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(6), \mu_{X}(6)\} = \max(1, 0) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(7), \mu_{X}(7)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(8), \mu_{X}(8)\} = \max(1, 0) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(9), \mu_{X}(9)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(10), \mu_{X}(10)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(11), \mu_{X}(11)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(12), \mu_{X}(12)\} = \max(1, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(13), \mu_{X}(13)\} = \max(0.3, 1) = 1,
\]
\[
\max \{1 - \mu_{\text{hot}}(14), \mu_{X}(14)\} = \max(1, 0) = 1.
\]

From the calculation above the smallest value is 0.2 hence:

\[
\min(\mu_{\text{hot}}(x), \inf_{Y \subset U} \max \{1 - \mu_{\text{hot}}(y), \mu_{X}(y)\}) = \min(\mu_{\text{hot}}(x), 0.2) \text{ which calculated again as follows:}
\]

\[
\min(\mu_{\text{hot}}(1), 0.2) = \min(0.7, 0.2) = 0.2,
\]
\[
\min(\mu_{\text{hot}}(2), 0.2) = \min(0.8, 0.2) = 0.2,
\]
\[
\min(\mu_{\text{hot}}(3), 0.2) = \min(0.5, 0.2) = 0.2,
\]

15
Then the lower approximations of the set $A$ for the decision value Yes are:

$$
\min(\mu_{\text{hot}}(4), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(5), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(6), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(7), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(8), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(9), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(10), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(11), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(12), 0.2) = \min(0, 0.2) = 0,
$$

$$
\min(\mu_{\text{hot}}(13), 0.2) = \min(0.7, 0.2) = 0.2,
$$

$$
\min(\mu_{\text{hot}}(14), 0.2) = \min(0, 0.2) = 0.
$$

Similarly for attribute mild and cool, we have:

$$
\min(\mu_{\text{mild}}(x), 0) = 0,
$$

$$
\min(\mu_{\text{cool}}(x), 0.2) = 0.
$$

Then the lower approximations of the set $A$ for the decision value Yes are:

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(1) = 0.2,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(8) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(2) = 0.2,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(9) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(3) = 0.2,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(10) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(4) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(11) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(5) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(12) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(6) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(13) = 0.2,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(7) = 0,
$$

$$
\mu_{A\{2,3,4,5,7,9,10,11,13\}}(14) = 0.
$$

But $\mu_{\text{POS}_{A(Q)}(x)} = \sup_{x \in U/Q} \mu_{AX}(x)$ then, $\mu_{\text{POS}_{A(Q)}(1)} = 0.3$.
\[ \mu_{POS_{A(Q)}}(8) = 0, \]
\[ \mu_{POS_{A(Q)}}(2) = 0.3, \]
\[ \mu_{POS_{A(Q)}}(9) = 0, \]
\[ \mu_{POS_{A(Q)}}(3) = 0.3 \]
\[ \mu_{POS_{A(Q)}}(10) = 0, \]
\[ \mu_{POS_{A(Q)}}(4) = 0, \]
\[ \mu_{POS_{A(Q)}}(11) = 0, \]
\[ \mu_{POS_{A(Q)}}(5) = 0, \]
\[ \mu_{POS_{A(Q)}}(12) = 0 \]
\[ \mu_{POS_{A(Q)}}(6) = 0, \]
\[ \mu_{POS_{A(Q)}}(13) = 0.3, \]
\[ \mu_{POS_{A(Q)}}(7) = 0, \]
\[ \mu_{POS_{A(Q)}}(14) = 0. \]

The next step to determine the degree of dependency of \( Q \) on \( A \) is as follows:

\[ \gamma_A(Q) = \sum_{x \in U} \mu_{POS_{A(Q)}}(x) \frac{|x|}{|U|} = 1.2 \]
\[ 14 = 0.08. \]

Similarly for the attribute \( B \) and \( C \), we obtain the following:

\[ \gamma_B(Q) = \sum_{x \in U} \mu_{POS_{B(Q)}}(x) \frac{|x|}{|U|} = 2.1 \]
\[ 14 = 0.15, \]
\[ \gamma_C(Q) = \sum_{x \in U} \mu_{POS_{C(Q)}}(x) \frac{|x|}{|U|} = 3.48 \]
\[ 14 = 0.248. \]

From this results it can be seen that the attribute \( B \) will cause the greatest increase in dependency degree. This attribute is chosen and added to the potential reduct. This process iterates and we calculate the two dependency degree \( \gamma_{(A,C)}(Q) \)and \( \gamma_{(B,C)}(Q) \) calculated as follows.

\[ \gamma_{(B,C)}(Q) = \sum_{x \in U} \mu_{POS_{(B,C)(Q)}}(x) \frac{|x|}{|U|} = 24.25 \]
\[ 14 = 0.3, \]
\[ \gamma_{(A,C)}(Q) = \sum_{x \in U} \mu_{POS_{(A,C)(Q)}}(x) \frac{|x|}{|U|} = 3.73 \]
\[ 14 = 0.27. \]

Adding attribute \( B \) to the reduct candidate causes the larger increase of dependency, so the new candidate becomes \( \{B, C\} \).

Now, we need to add the attribute \( A \) to the subset \( \{B, C\} \) and calculate \( \gamma_{(A,B,C)}(Q) \)

\[ \gamma_{(A,B,C)}(Q) = \sum_{x \in U} \mu_{POS_{(A,B,C)(Q)}}(x) \frac{|x|}{|U|} = 4.4 \]
\[ 14 = 0.31. \]
5. Fuzzy ID3 Decision Approach

Decision tree induction methods such as ID3 generate a tree structure through recursively partitioning the attribute space until the whole decision space is completely partitioned into a set of non-overlapping subspaces. The original decision tree is restricted to attributes that take a discrete set of values. If continuous attributes are involved, they must be discredited appropriately. Two typical methods have been used for this purpose, one of which is to partition the attribute range into two intervals using a threshold, while another is to partition the attribute domain into several intervals using a set of cut points.

In both methods above, the cut points used in classical decision trees are usually crisp. Applications showed that these approaches could only work well for a disjoint class that can be separated with clearly defined boundaries. However, due to the existence of vague and imprecise information on real world problems, the class boundaries may not be defined clearly. In this case, the decision tree may produce high misclassification rates in testing even if they perform well in training. To overcome this drawback, several approaches have been proposed, including probability based method [42, 43, 45]. Another method to handle this problem use fuzzy set theory decision trees, the fuzzy reasoning process allows two or more rules to be simultaneously validated with gradual certainty and the end result will be the outcome of combining several results.

As with crisp decision trees fuzzy decision tree induction involves the recursive partitioning of training data in a top-down manner. The most informative attribute is selected at each stage and the remaining data are divided according to the values of the attribute. Partitioning continues until there are no more attribute to evaluate or if the examples in the current node belong to the same class.

The main difference between ID3 and fuzzy ID3 (FID3) is that FID3 is able to handle continuous attributes through the use of fuzzy sets. Fuzzy sets and logic allow language related uncertainties to be modeled and provide a symbolic framework for knowledge comprehensibility, unlike crisp decision tree induction on FDT do not use the original numerical attribute values directly in the tree. Instead, they use fuzzy sets generated either from a gasification process beforehand or expert defined partitions to construct comprehensible trees. As a result of this there are several key differences between FDT induction and the original crisp approaches:

1. Membership of objects, traditionally objects/examples belonged to nodes with a membership of \{0,1\}; now these memberships may take values from the interval [0,1]. On each node an example has a different membership degree to the current example set and this degree is calculated from the conjunctive combination of the membership degrees of the example to the fuzzy sets along the path to the node and its degrees of membership to the classes.

2. Measures of attribute significance, as fuzzy sets are used, the measures of significance should incorporate this membership information to decide which attribute from nodes within the tree. This is particularly important as the equality of the tree can be greatly reduced by a poor measure of attribute significance.

3. Fuzz tests, within nodes fuzzy tests are carried out to determine the membership degree of a feature value to a fuzzy set.

4. Stopping criteria, learning is usually terminated if all attributes are used on current path or if all objects in the current node belong to the same class. With fuzzy
trees objects can belong to any node with any degree of membership. As a result of this fuzzy tree tends to be larger in size which can lead to poorer generalization performance. An additional threshold can be introduced based on the attribute significance measure, to determine construction earlier in induction. For classification, the decision tree is converted to an equivalent result.

The formulas of the fuzzy entropy [39] for attributes and the information gain are a little bit different because of the data fuzzy expression. Their definitions are defined as follows respectively with the assumption dataset $S = \{x_1, x_2, ..., x_j\}$:

$$H_F(S, H) = -\sum_{i=1}^{C} \sum_{j}^{N} \frac{\mu_{ij}}{S} \log_2 \left( \sum_{j}^{N} \frac{\mu_{ij}}{S} \right), G_F(S, A) = H_F(S) - \sum_{V \subseteq A} \frac{|S_V|}{|S|} \times H_F(S_V, A)$$

where : $\mu_{ij}$ is the membership value of the $j^{th}$ pattern to the $i^{th}$ class.

$H_F(S)$ presents the entropy of the set $S$ of training examples in the node.

$|S_V|$ is the size of the subset $S_V \subseteq S$ of training examples $x_j$ with $V$ attributes.

$|S|$ presents the size of set $S$.

Define thresholds

If the learning of FDT stops until all the same data in each leaf node belongs to one class, it is poor in accuracy. In order to improve the accuracy, the learning must stopped early or termed pruning in general. As a result, two thresholds are defined as follows:

1. fuzziness control threshold $\theta_r$.

If the proportion of a dataset of a class $C_k$ is greater than or equal to a threshold $\theta_r$, stop expanding the tree. For example; if in sub-dataset the ratio of class 1 is 90%, class2 is 10% and $\theta_r$ is 85%, then stop expanding.

1. Leaf decision threshold $\theta_n$.

If the number of a dataset is less than a threshold $\theta_n$, stop expanding. For example, a dataset has 600 examples where $\theta_n$ is 2%. If the number of samples in a node is less than 12 (2% of 600), then stop expanding.

The level of these thresholds has great influences on the result of the tree. We define them in different levels in our experiment to find optimal values. Moreover, if there are no more attributes for classification, the algorithm does not create a new node.

5 Fuzzy ID3 algorithm

1. create a root node that has a set of fuzzy data with membership value 1.

2. If a node $t$ with a fuzzy set of data $D$ satisfies the following conditions, then it is a leaf node and assigned by the class name.

   (a) The proportion of a class $C_k$ is greater than or equal to $\theta_r$, $\frac{|D_{C_k}|}{|D|} \geq \theta_r$.

   (b) The number of a dataset is less than $\theta_n$.

   (c) There are no attributes for more classifications.
3. If a node $D$ does not satisfy the above conditions, then it is not a leaf node, and a new sub-node is generated as follows:

4. For $A_i \ (i = 1, 2, \ldots, l)$ calculated the information gain, and select the test attribute $A_{max}$ that maximizes them.

5. Divide $D$ into fuzzy subsets $D_1, D_2, \ldots, D_m$ according to $A_{max}$, where the membership value of the data in $D_j$ is the product of the membership value in $D$ and the value of $F_{max,j}$ of the value of $A_{max}$ in $D$.

6. Generate new nodes $t_1, t_2, \ldots, t_m$ for fuzzy subsets $D_1, D_2, \ldots, D_m$ and label the fuzzy sets $F_{max,j}$ to edges that connect between the nodes $t_j$ and $t$.

7. Replace $D$ by $D_j \ (j = 1, 2, \ldots, m)$ and repeat from 2 recursively.

We apply the ID3 approach on the fuzzy data in Table 2, first, we have to calculate the fuzzy entropy and information gain of the fuzzy dataset to expand the tree. In this case, we get the same result of the entropy of the ID3 $H_F(S) = 0.94$. Now, we do the calculation on the example set:

(1) Information gain of the weather attribute are:

$$H_F(\text{weather, cool}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0.918, H_F(\text{weather, mild}) = -\frac{4.94}{7.41} \log_2 \frac{4.94}{7.41} - \frac{2.47}{7.41} \log_2 \frac{2.47}{7.41} = 0.918,$$

$$H_F(\text{weather, hot}) = -\frac{1.2}{2.7} \log_2 \frac{1.2}{2.7} - \frac{1.5}{2.7} \log_2 \frac{1.5}{2.7} = 0.991,$$

then $G_F(S, \text{weather}) = 0.0098$.

(2) Information gain of the rain attribute are:

$$H_F(\text{rain, weak}) = -\frac{6.5}{9} \log_2 \frac{6.5}{9} - \frac{2.5}{9} \log_2 \frac{2.5}{9} = 0.852, H_F(\text{rain, strong}) = -\frac{1.1}{2.1} \log_2 \frac{1.1}{2.1} - \frac{1}{2.1} \log_2 \frac{1}{2.1} = 0.998,$$

then $G_F(S, \text{rain}) = 0.06$.

(3) Information gain of the traffic-jam attribute are:

$$H_F(\text{traffic, long}) = -\frac{0.96}{2.09} \log_2 \frac{0.96}{2.09} - \frac{1.13}{2.09} \log_2 \frac{1.13}{2.09} = 0.995, H_F(\text{traffic, short}) = -\frac{6.01}{7.81} \log_2 \frac{6.01}{7.81} - \frac{1.8}{7.81} \log_2 \frac{1.8}{7.81} = 0.779,$$

then $G_F(S, \text{traffic}) = 0.1154$.

The information gain of the attribute traffic-jam has the highest value. We use it to expand the tree. Generate two sub-nodes with the examples, where the membership values of these sub-nodes are the product of the original membership values at the root and the membership values of the attribute traffic-jam. The examples are omitted if its membership value is null.
Next, we have to calculate the proportion of the class \( C_k \). It is the quotient of the sum of membership values of class \( C_k \) to the sum of all the membership values.

Now we calculate the proportion of class \( N \) and \( Y \) for left and right nodes.

1. The proportion of class \( N \) for the left node long is:

\[
N : \frac{1.13}{2.09} = 55\%, \quad Y : \frac{0.96}{2.09} = 46\%.
\]

Also for the right node short:

\[
N : \frac{1.8}{7.81} = 23\%, \quad Y : \frac{6.01}{7.81} = 77\%.
\]

Calculation of the class membership:

1. Left node

\[
C_N = 0.25 + 0.18 + 0.17 + 0.53 = 1.13, \quad C_Y = 0.33 + 0.4 + 0.23 = 0.96.
\]

The total membership value: \( 1.13 + 0.96 = 2.09 \).

Proportion of class \( N \): 54%.

Proportion of class \( Y \): 46%.

1. Right node

\[
C_N = 0.25 + 0.37 + 0.8 + 0.38 = 1.8, \quad C_Y = 0.12 + 0.87 + 1 + 0.92 + 0.82 + 1 + 0.28 + 1 = 6.01.
\]

The total membership value: \( 1.8 + 6.01 = 7.81 \).

Proportion of class \( N \): 23%.

Proportion of class \( Y \): 77%.

In this case, if \( \theta_r \) is 95%, then both of the nodes expand; if \( \theta_r \) is 75%, the left node expands and the right node stops expanding.

6 Conclusion

One of the main contributions of this paper is in the area of fuzzy rough classifications. Based on topological spaces, we presented an underlying theory to explain how classifications in information systems may be performed using fuzzy rough set approach.

The main difference between ID3 and fuzzy ID3 (FID3) is that FID3 is able to handle continuous attributes through the use of fuzzy sets. Fuzzy sets and logic allow language related uncertainties to be modeled and provide a symbolic framework for knowledge comprehensibility, unlike crisp decision tree induction on FDT do not use the original numerical attribute values directly in the tree. Instead, they use fuzzy sets generated either from a fuzzification process beforehand or expert defined partitions to construct comprehensible trees.

We conclude that the intermingling of fuzzy rough sets in the construction of some approximation concepts will help to get results with abundant logical statements. That is discovering hidden relationships among data and, moreover, probably helps in producing accurate programs.
References

http://dx.doi.org/10.1016/j.asoc.2004.11.003

http://dx.doi.org/10.1016/j.patcog.2008.11.021

http://dx.doi.org/10.1016/j.ins.2006.11.013

http://dx.doi.org/10.1016/j.ins.2008.05.024

http://dx.doi.org/10.1016/j.ijar.2009.02.003

http://dx.doi.org/10.1016/j.omega.2007.04.003

http://dx.doi.org/10.1016/j.eswa.2008.06.019

http://dx.doi.org/10.1016/j.chaos.2004.09.107

http://dx.doi.org/10.1016/j.ins.2009.04.002

http://dx.doi.org/10.1016/j.patrec.2007.10.026

http://dx.doi.org/10.1016/j.eswa.2008.12.026


http://dx.doi.org/10.1016/S0019-9958(65)90241-X