An Overview on the Approximation Quality Based on Rough-Fuzzy Hybrids

Van-Nam Huynh, Tu-Bao Ho and Yoshiteru Nakamori

Abstract The so-called measure of approximation quality plays an important role in many applications of rough set based data analysis. In this chapter, we provide an overview on various extensions of approximation quality based on rough-fuzzy and fuzzy-rough sets, along with highlighting their potential applications as well as future directions for research in the topic.

1 Introduction

After nearly twenty years since the introduction of fuzzy sets theory [51], Pawlak [33] introduced the notion of a rough set as a new mathematical tool to deal with the approximation of a concept in the context of incomplete information. Basically, while a fuzzy set models the ill-definition of the boundary of a concept often described linguistically, a rough set characterizes a concept by its lower and upper approximations due to indiscernibility between objects arose because of incompleteness of available knowledge. Since its inception, the rough set theory has been proven to be of substantial importance in many areas of application [34, 39, 45].

During the last decades, many attempts to establish the relationships between the two theories, to compare each to the other, and to simultaneously hybridize them have been made, e.g., [10, 30, 31, 35, 40, 46, 47, 49, 50]. Among these lines of research, rough fuzzy hybridization has emerged as a promising new paradigm for decision-making related applications [17, 18, 31, 32], data analysis [22, 25] and many others. This is due to rough-fuzzy hybrids can encapsulate two distinct aspects of imperfection of knowledge being vagueness and indiscernibility, which may simultaneously occur in many situations of practical application [10].

On the other hand, one of issues of great practical importance in data analysis is discovering dependencies between attributes in datasets. In rough set theory, the notion of **approximation quality** (also called **degree of dependency**) is often used to evaluate the classification success of attributes in terms of a numerical evaluation of the dependency properties generated by these attributes. Particularly, it has been used as a useful tool, for instance, for discovering data dependencies and for semantics-preserving feature reduction using only the given data without any additional information as required by other theories [13, 25, 34]. This chapter aims

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2007

at providing an overview on various extensions of approximation quality based on rough-fuzzy hybrids, along with highlighting their potential applications and future directions for research in the topic as well.

The structure of the rest of this chapter is as follows. Section 2 briefly introduces necessary notions of fuzzy sets and rough sets. Section 3 recalls Pawlak's notions of approximation quality and significance of attributes. In Sect. 4, the notions of rough fuzzy sets and fuzzy rough sets are reviewed in relation to their applications in practice. Sect. 5 devotes to an overview on rough-fuzzy hybrids based extensions of approximation quality, accompanying with illustrative examples. Finally, some concluding remarks and future work are presented in Sect. 6.

2 Basic of Rough Sets and Fuzzy Sets

In this section we briefly recall basic notions of fuzzy sets and rough sets. For the purpose of this paper, it is sufficient to consider the finite version of universes of discourse.

2.1 Fuzzy Sets

Let $\mathbb{U}$ be a finite and non-empty set called universe of discourse. A fuzzy set F of $\mathbb{U}$ is a mapping $\mu_F : \mathbb{U} \longrightarrow [0, 1]$, where for each $x \in \mathbb{U}$ we call $\mu_F(x)$ the membership degree of x in F.

Given a number $\alpha \in (0, 1]$, the α-cut, or α-level set, of F is defined as follows

$$F_\alpha = \{x \in \mathbb{U} | \mu_F(x) \geq \alpha\}$$

which is a subset of $\mathbb{U}$. Let us denote $\text{rng}(\mu_F) = \mu_F(\mathbb{U}) \setminus \{0\}$ and assume that $\text{rng}(\mu_F) = \{\alpha_1, \ldots, \alpha_n\}$, where $\alpha_i > \alpha_{i+1}$, for $i = 1, \ldots, n-1$. Then the membership function μ_F can be expressed as [12]

$$\mu_F(x) = \sum_{x \in F_{\alpha_i}} (\alpha_i - \alpha_{i+1}) \tag{1}$$

Clearly, $\alpha_1 = 1$ if F is normal, i.e. $\exists x$ such that $\mu_F(x) = 1$. This representation of a fuzzy set is considered as providing a probability based semantics for membership function of fuzzy sets, where $m_i = (\alpha_i - \alpha_{i+1})$, with $\alpha_{n+1} = 0$ by convention, can be viewed as the probability that F_{α_i} stands as a crisp representative of F. Then $\{(F_{\alpha_i}, m_i) | i = 1, \ldots, n\}$ is usually referred to as a finitely discrete **consonant random set**, or **body of evidence** [41]. Note that the normalization assumption of F insures the body of evidence does not contain the empty set. This view of fuzzy sets has been also used in [2] to introduce the so-called **mass assignment** of a fuzzy set, with relaxing the normalization assumption of fuzzy sets. Namely, the mass assignment of F, denoted by m_F, is a probability distribution on $2^{\mathbb{U}}$ defined by

$$\begin{aligned} m_F(\emptyset) &= 1 - \alpha_1, \\ m_F(F_{\alpha_i}) &= m_i, \text{ for } i = 1, \ldots, n. \end{aligned} \tag{2}$$

2.2 Rough Sets

Pawlak's theory of rough sets begins with the notion of an approximation space, which is a pair $\langle \mathbb{U}, R \rangle$, where $\mathbb{U}$ is a non-empty set (the universe of discourse) and R an equivalence relation on U, i.e., R is reflexive, symmetric, and transitive. The relation R decomposes the set $\mathbb{U}$ into disjoint classes in such a way that two elements x, y are in the same class iff $(x, y) \in R$. If two elements x, y in $\mathbb{U}$ belong to the same equivalence class, we say that x and y are indistinguishable. For $X \in 2^{\mathbb{U}}$, in general it may not be possible to describe X precisely in $\langle \mathbb{U}, R \rangle$. One may then characterize X by a pair of lower and upper approximations defined as follows [33]

$$\underline{R}(X) = \{x \in \mathbb{U} | [x]_R \subseteq X\}; \qquad \overline{R}(X) = \{x \in \mathbb{U} | [x]_R \cap X \neq \emptyset\}$$

where $[x]_R$ stands for the equivalence class of x by R. The pair $(\underline{R}(X), \overline{R}(X))$ is the representation of an ordinary set X in the approximation space $\langle U, R \rangle$ or simply called the rough set of X.

In the context of rough set based data analysis, the equivalence relation in an approximation space is often interpreted via the notion of information systems. An **information system** $\mathcal{I}$ is a pair $\mathcal{I} = \langle \mathbb{U}, \mathcal{A} \rangle$, where $\mathbb{U}$ is a set of objects, $\mathcal{A}$ is a set of attributes, and each attribute $a \in \mathcal{A}$ associated with the set of attribute values V_a is understood as a mapping $a : \mathbb{U} \to V_a$. An information system is called a **decision system** if assuming that the set of attributes $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ and $\mathcal{C} \cap \mathcal{D} \neq \emptyset$, where $\mathcal{C}$ is the set of **conditional attributes** and $\mathcal{D}$ is the set of **decision attributes**. Given an information system $\mathcal{I}$, each subset P of the attribute set $\mathcal{A}$ induces an equivalence relation $\text{IND}(P)$ called P*-indiscernibility relation* as follows

$$\text{IND}(P) = \{(x, y) \in \mathbb{U}^2 | a(x) = a(y), \text{ for all } a \in P\}$$

and $\text{IND}(P) = \bigcap_{a \in P} \text{IND}(\{a\})$. If $(x, y) \in \text{IND}(P)$ we then say that objects x and y are indiscernible with respect to attributes in P. In other words, we cannot distinguish x from y, and vice versa, in terms of attributes in P. Note that the partition of $\mathbb{U}$ generated by $\text{IND}(P)$, denoted by $\mathbb{U}/\text{IND}(P)$, can be calculated in terms of those partitions generated by single attributes in P as follows [24]

$$\mathbb{U}/\text{IND}(P) = \bigotimes_{a \in P} \mathbb{U}/\text{IND}(\{a\}) \tag{3}$$

where

$$\mathcal{X} \otimes \mathcal{Y} = \{X \cap Y | X \in \mathcal{X}, Y \in \mathcal{Y}, X \cap Y \neq \emptyset\}$$

For simplicity of notation, from now on we use the same notation P to denote the equivalence relation induced from a set P of attributes, instead of IND(P).

3 Pawlak's Approximation Quality

As mentioned in [13], one of the strengths of rough set theory is the fact that all its parameters are directly obtained from the given data. That is, in rough set theory the numerical value of imprecision is calculated by making use of the granularity structure of the data only, while other uncertainty theories like Dempster-Shafer theory [41] or fuzzy set theory [26] require probability assignments and membership values respectively.

In [34], Pawlak firstly introduces two numerical characterizations of imprecision of a subset X in the approximation space $\langle \mathbb{U}, P \rangle$: *accuracy* and *roughness*. Accuracy of X, denoted by $\alpha_P(X)$, is simply the ratio of the number of objects in its lower approximation to that in its upper approximation; namely

$$\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \tag{4}$$

where $|\cdot|$ denotes the cardinality of a set. Then the roughness of X, denoted by $\rho_P(X)$, is defined by subtracting the accuracy from 1 as

$$\rho_P(X) = 1 - \alpha_P(X) = 1 - \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \tag{5}$$

Note that the lower is the roughness of a subset, the better is its approximation. In [48], Yao has interpreted Pawlak's accuracy measure in terms of a classic distance measure based on sets, called *Marczewski-Steinhaus (MS) metric* [27], which is defined by

$$D_{MS}(X, Y) = \frac{|X \cup Y| - |X \cap Y|}{|X \cup Y|} = 1 - \frac{|X \cap Y|}{|X \cup Y|}$$

Using MS metric, the roughness measure of a set X in $\langle \mathbb{U}, P \rangle$ is the distance between its lower and upper approximations.

Suppose now that two views of universe $\mathbb{U}$ are given, which may come from two different subsets P and Q of attributes, by means of associated equivalence relations. Then an interesting question arises to be how well the knowledge from one view can be expressed by that from the other. In other words, we are concerned here with the issue of measuring dependencies between attributes. This issue is very important in many tasks of data analysis. In rough set theory, the so-called **approximation quality measure** γ [34] is often used for such a situation to describe the degree of partial dependency between attributes.

Particularly, let P and Q be equivalence relations over $\mathbb{U}$, then the approximation quality of Q by P, also called *degree of dependency*, is defined by

$$\gamma_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|\mathbb{U}|} \tag{6}$$

where

$$\mathrm{POS}_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}(X) \tag{7}$$

is called the **positive region** of the partition $\mathbb{U}/Q$ with respect to P. We then say that Q depends on P in a degree $k = \gamma_P(Q)$ $(0 \leq k \leq 1)$ and denote as $P \Rightarrow_k Q$. If $k = 1$, Q totally depends on P; if $0 < k < 1$, Q partially (or roughly) depends on P, and if $k = 0$, Q is totally independent from P.

Note that the approximation quality $\gamma_P(Q)$ can be also represented in terms of accuracy as follows

$$\gamma_P(Q) = \sum_{X \in \mathbb{U}/Q} \frac{|\overline{P}(X)|}{|\mathbb{U}|} \alpha_P(X) \tag{8}$$

Then, $\gamma_P(Q)$ is regarded as the weighted mean of the accuracies of approximation of sets $X \in \mathbb{U}/Q$ by P [13].

Another issue of great practical importance is that of identifying how significant a specific attribute (or a group of attributes) is in respect of the classification power. This information is captured by calculating the change in dependency when an attribute is removed from the set of considered conditional attributes. In particularly, we can measure the significance of an attribute $a \in P$ with respect to the classification induced from Q by the difference

$$\sigma_P(Q, a) = \gamma_P(Q) - \gamma_{P \setminus \{a\}}(Q) \tag{9}$$

This measure expresses how influence on the quality of approximation if we drop the attribute a from P. The higher the change in dependency, the more significant the attribute is. If the significance is 0, the attribute is dispensable. A subset S of P is called a Q-reduct of P (or a reduct of P with respect to Q) if $\gamma_S(Q) = \gamma_P(Q)$.

In [13], the authors have also used the MS metric to re-interpret the rough approximation quality γ and ascertain its statistical significance. The approximation quality measure and its extended variants have been extensively studied and used in many applications, especially in feature selection, e.g., [4, 9, 22, 23, 24, 25, 43, 44] and ranking problems, e.g., [14, 15, 16].

4 Rough-Fuzzy Hybrids

As argued by Dubois and Prade [10], rough sets and fuzzy sets capture two distinct aspects of imperfection of knowledge: indiscernibility and vagueness, that may be simultaneously present in a given application. Therefore, it is necessary to find out hybrid models which combines these notions for knowledge representation and integration in such situations. Among many possibilities for rough-fuzzy hybridization, the most typical ones are to fuzzify sets to be approximated and/or to fuzzify the equivalence relation in an approximation space [10, 11]. The first case allows to obtain rough approximations of fuzzy sets which results in the so-called **rough fuzzy sets**; while the second case allows to obtain approximations of (fuzzy) sets by means of fuzzy similarity relations resulting in the so-called **fuzzy rough sets**.

4.1 Rough Fuzzy Sets

Given an approximation space $\langle \mathbb{U}, P \rangle$. Let F be a fuzzy set in $\mathbb{U}$ with the membership function μ_F. The upper and lower approximations $\overline{P}(F)$ and $\underline{P}(F)$ of F by P are fuzzy sets in the quotient set $\mathbb{U}/P$ with membership functions defined by, for each $F_i \in \mathbb{U}/P$,

$$\mu_{\overline{P}(F)}(F_i) = \sup_{x \in F_i} \{\mu_F(x)\}$$

$$\mu_{\underline{P}(F)}(F_i) = \inf_{x \in F_i} \{\mu_F(x)\}$$

The pair $(\underline{P}(F), \overline{P}(F))$ is then called a rough fuzzy set [11].

Furthermore, the rough fuzzy set $(\underline{P}(F), \overline{P}(F))$ naturally induces two fuzzy sets $P^*(F)$ and $P_*(F)$ in $\mathbb{U}$ with membership functions are defined respectively as follows

$$\mu_{P^*(F)}(x) = \mu_{\overline{P}(F)}([x]_P) \text{ and } \mu_{P_*(F)}(x) = \mu_{\underline{P}(F)}([x]_P) \tag{10}$$

That is, $P^*(F)$ and $P_*(F)$ are fuzzy sets with constant membership degree on the equivalence classes of $\mathbb{U}$ by P, and for any $x \in \mathbb{U}$, $\mu_{P^*(F)}(x)$ (respectively, $\mu_{P_*(F)}(x)$) can be viewed as the degree to which x possibly (respectively, definitely) belongs to the fuzzy set F [3]. Conceptually, the pair $(P_*(F), P^*(F))$ can be viewed as "extension" of rough fuzzy set $(\underline{P}(F), \overline{P}(F))$.

Rough fuzzy sets could find many applications in practical situations where a fuzzy classification or a fuzzy concept must be approximated by available knowledge expressed in terms of a Pawlak's approximation space, for instance as in pattern recognition and image analysis problems [1, 3, 5, 6, 7, 36, 37, 38, 42].

4.2 Fuzzy Rough Sets

Let us consider another extension of rough sets corresponding to the second case mentioned above. In this extension, instead of equipping the universe $\mathbb{U}$ with an equivalence relation P, we consider a fuzzy similarity relation R, i.e., a fuzzy set R of $\mathbb{U}^2$, such that the properties of reflexivity ($\mu_R(x, x) = 1$), symmetry ($\mu_R(x, y) = \mu_R(y, x)$), and $\wedge$-transitivity of the form

$$\mu_R(x, z) \geq \mu_R(x, y) \wedge \mu_R(y, z)$$

are holded [52]. In order to define fuzzy rough approximation operators, the counterpart of equivalence classes called fuzzy equivalence classes must be defined first. According to Zadeh [52], the fuzzy equivalence class $[x]_R$ of objects close to x is defined by

$$\mu_{[x]_R}(y) = \mu_R(x, y), \forall y \in \mathbb{U} \tag{11}$$

Interestingly, this definition degenerates to the usual definition of equivalence classes when R is a non-fuzzy relation. Furthermore, Höhle [19] also proposed a definition of what should be a fuzzy equivalence class X by means of the following axioms

(i) μ_X is normalized, i.e. $\exists x, \mu_X(x) = 1$,
(ii) $\mu_X(x) \wedge \mu_R(x, y) \leq \mu_X(y)$,
(iii) $\mu_X(x) \wedge \mu_X(y) \leq \mu_R(x, y)$.

Then, according to [10], a fuzzy set $[x]_R$ as in (11) is a fuzzy equivalence class in the sense of Höhle.

The family of fuzzy equivalence classes $\{[x]_R | x \in \mathbb{U}\}$, also denoted by $\mathbb{U}/R$, forms a "fuzzy partition" of $\mathbb{U}$. Also, a more direct way is to define a family $\mathcal{F} = \{F_1, \ldots, F_n\}$ of normal fuzzy sets of $\mathbb{U}$, with $m < |\mathbb{U}|$, which covers $\mathbb{U}$ sufficiently in the following sense

$$\inf_{x \in \mathbb{U}} \max_i \mu_{F_i}(x) > 0$$

Further, a disjointness property between F_i's can be requested as

$$\forall i, j, \sup_{x \in \mathbb{U}} \min\{\mu_{F_i}(x), \mu_{F_j}(x)\} < 1$$

In the literature, a stronger restriction is often adopted

$$\sum_{i=1}^{n} \mu_{F_i}(x) = 1 \tag{12}$$

for any $x \in \mathbb{U}$. Then $\mathcal{F}$ plays the role of the family of fuzzy equivalence classes induced from a similarity relation R, i.e., $\mathcal{F} = \mathbb{U}/R$.

Given a fuzzy approximate space $\langle \mathbb{U}, R \rangle$, a fuzzy set F can be approximated by means of the fuzzy partition $\mathbb{U}/R$ in terms of an R-upper and an R-lower approximation $\overline{R}(F)$ and $\underline{R}(F)$ as follows [10]

$$\mu_{\overline{R}(F)}(F_i) = \sup_{x \in \mathbb{U}} \min\{\mu_{F_i}(x), \mu_F(x)\} \tag{13}$$

$$\mu_{\underline{R}(F)}(F_i) = \inf_{x \in \mathbb{U}} \max\{1 - \mu_{F_i}(x), \mu_F(x)\} \tag{14}$$

for any $F_i \in \mathbb{U}/R$. The pair $(\underline{R}(F), \overline{R}(F))$ is then called a fuzzy rough set. When F_i's are crisp, i.e., R is an equivalence relation, we obtain the rough approximation of F which results in a rough fuzzy set defined previously.

As noted in [24], these definitions given in (13)–(14) differ a little from the crisp rough approximations, as the memberships of individual objects to the approximations are not explicitly available. As a result of this, fuzzy rough approximations are redefined as fuzzy sets of $\mathbb{U}$ [24] by

$$\mu_{\overline{R}(F)}(x) = \sup_{F_i \in \mathbb{U}/R} \min\left(\mu_{F_i}(x), \sup_{y \in \mathbb{U}} \min\{\mu_{F_i}(y), \mu_F(y)\}\right) \tag{15}$$

$$\mu_{\underline{R}(F)}(x) = \sup_{F_i \in \mathbb{U}/R} \min\left(\mu_{F_i}(x), \inf_{y \in \mathbb{U}} \max\{1 - \mu_{F_i}(y), \mu_F(y)\}\right) \tag{16}$$

These definitions have been often used in application of fuzzy rough sets to dimensionality reduction [22, 23, 24, 25, 44].

Remark 1. Note that (15)–(16) can be viewed as the "extension" of the fuzzy rough set $(\underline{R}(F), \overline{R}(F))$, which was defined in [10] making use of the knowledge of fuzzy similarity relation R directly, instead of fuzzy equivalence classes induced by R. Particularly, according to Dubois and Prade [10], we have

$$\mu_{\overline{R}(F)}(x) = \sup_{y \in \mathbb{U}} \mu_F(x) * \mu_R(x, y) \tag{17}$$

$$\mu_{\underline{R}(F)}(x) = \inf_{y \in \mathbb{U}} \mu_R(x, y) \to \mu_F(y) \tag{18}$$

where $*$ is a t-norm and $\to$ is an S-implication operator. However, in practical applications of fuzzy rough sets in data analysis, the knowledge of fuzzy similarity relation R may not be available, but a fuzzy linguistic partition of attribute domain which plays the role of the family of fuzzy equivalence classes is often pre-assumed. This practically explains why (15)–(16) is often used in application.

For a more general and comprehensive treatment of fuzzy rough sets, the readers can refer, e.g., to [10, 11, 40, 49].

5 Approximation Quality Based on Rough-Fuzzy Sets

As we have mentioned previously, rough fuzzy sets arise naturally when we want to approximate a fuzzy set or a fuzzy classification by means of the available knowledge expressed in terms of an approximation space $\langle \mathbb{U}, P \rangle$.

The first case may often occur in, for example, problems of image analysis, where $\mathbb{U}$ denotes a gray image or feature space and $\mathbb{U}/P$ is a partition of $\mathbb{U}$, a fuzzy set F can be viewed to represent ill-defined pattern classes or some imprecise image property such as brightness, darkness, smoothness, etc [3, 7]. In such a situation, roughness (or accuracy) of a fuzzy set F may be used to provide the information of how well its approximation is in $\langle \mathbb{U}, P \rangle$. Regarding to this, Banerjee and Pal [3] have proposed a roughness measure for fuzzy sets and have discussed the issue of how to use this measure in tasks of image analysis.

The second case may come up in a natural way when a linguistic classification must be expressed by means of already existing knowledge P. For example, let us consider two attributes "experience" and "salary" in a database of employees. Then the attribute "experience" may take values in a finite set of labels such as **good**, **poor**, **very good**, etc., and the attribute "salary" may have numerical values. Then it is natural to intuitively infer a "partial" dependence between "experience" and "salary" as (the better the experience, the higher the salary). However, such a dependency could not be expressed in terms of traditional data dependencies, because there may be different employees having the same value of "experience" but different salaries, even in small magnitude. Therefore, it is necessary and useful to look for measures such as the approximation quality that may support us as numerical characteristics to realize partial dependency between attributes in such situations.

5.1 Roughness of a Fuzzy Set

Banerjee and Pal's Approach

In [3], Banerjee and Pal have proposed a roughness measure for fuzzy sets in a given approximation space. Essentially, this measure of roughness of a fuzzy set depends on parameters that are designed as thresholds of definiteness and possibility in membership of the objects in $\mathbb{U}$ to the fuzzy set.

More explicitly, let us be given an approximation space $\langle \mathbb{U}, P \rangle$ and a fuzzy set F in $\mathbb{U}$. We now consider parameters α, β such that $0 < \beta \leq \alpha \leq 1$. The α-cut $P_*(F)_\alpha$ and β-cut $P^*(F)_\beta$ of fuzzy sets $P_*(F)$ and $P^*(F)$, respectively, are called to be the α-*lower approximation* and the β-*upper approximation* of F in $\langle \mathbb{U}, P \rangle$, respectively. Then a roughness measure of the fuzzy set F with respect to parameters α, β, with $0 < \beta \leq \alpha \leq 1$, and the approximation space $\langle \mathbb{U}, P \rangle$ is defined by

$$\rho_P^{\alpha,\beta}(F) = 1 - \frac{|P_*(F)_\alpha|}{|P^*(F)_\beta|}$$

By the assumption made on parameters, we have

1. $0 \leq \rho_P^{\alpha,\beta}(F) \leq 1$.
2. If F is a fuzzy set such that there is a member x in each equivalence class of $\mathbb{U}/P$ with $\mu_F(x) < \alpha$, then $\rho_P^{\alpha,\beta}(F) = 1$.
3. If F is a definable fuzzy set, i.e., μ_F is a constant function on each equivalence class of $\mathbb{U}/P$ and $\alpha = \beta$, then $\rho_P^{\alpha,\beta}(F) = 0$.

Note that while the third statement seems interesting as it says that the measure $\rho_P^{\alpha,\beta}(\cdot)$ inherits a property of Pawlak's roughness measure, the second one may not be well-justified. Furthermore, the following property of $\rho_P^{\alpha,\beta}(\cdot)$ proved in [3] may be also undesired, unless the support of a constant fuzzy set, i.e. its strong 0-cut, is definable in the approximation space.

Proposition 1. *If F is a constant fuzzy set, say $\mu_F(x) = \delta$, for all $x \in \mathbb{U}$, then $\rho_P^{\alpha,\beta}(F) = 0$, with the exception when $\beta < \delta < \alpha$, in which $\rho_P^{\alpha,\beta}(F) = 1$.*

Properties of the measure $\rho_P^{\alpha,\beta}(\cdot)$ and its potential applications in the field of pattern recognition have been reported and mentioned in [3], and more recently in [53].

An Alternative Approach

In [20], the authors have introduced a parameter-free measure of roughness of a fuzzy set that in fact is a generalization of Pawlak's notion of roughness measure and avoids the undesirable properties held by Banerjee and Pal's roughness measure as mentioned above. Basically, this approach is based on the random set based representation of a fuzzy set and defines its roughness as the weighted mean of roughness measures of its crisp representatives.

In particularly, let $\text{rng}(\mu_F)$ and m_F be the range of the membership function μ_F and the mass assignment of F, respectively. Recall that in this representation of fuzzy set F, for each $\alpha \in \text{rng}(\mu_F)$, $m_F(F_\alpha)$ is viewed as the probability that F_α stands as a crisp representative of F. Under such a representation, the roughness measure of F with respect to the approximation space $\langle \mathbb{U}, P \rangle$ is defined as follows

$$\hat{\rho}_P(F) = \sum_{\alpha \in \text{rng}(\mu_F)} m_F(F_\alpha)\left(1 - \frac{|\underline{P}(F_\alpha)|}{|\overline{P}(F_\alpha)|}\right) \equiv \sum_{\alpha \in \text{rng}(\mu_F)} m_F(F_\alpha)\rho_P(F_\alpha) \qquad (19)$$

Remark 2. With this definition of roughness, we have

- $0 \leq \hat{\rho}_P(F) \leq 1$.
- $\hat{\rho}_P(\cdot)$ is a natural extension of Pawlak's roughness measure for fuzzy sets, i.e., if F is a crisp subset of U then $\hat{\rho}_P(F) = \rho_P(F)$.
- F is a definable fuzzy set, i.e., if $\underline{P}(F) = \overline{P}(F)$, if and only if $\hat{\rho}_P(F) = 0$.

Table 1 The approximations of the fuzzy set μ_{small}

	{0,2,4}	{1,3,5}	{6,8,10}	{7,9}
μ_{small_*}	0.25	0	0	0
μ_{small^*}	1	1	0	0

Let us consider a simple example depicting the introduced notions.

Example 1. Suppose we are given an approximation space $\langle \mathbb{U}, P \rangle$, where $U = \{0, 1, 2, \ldots, 10\}$ and P is such that

$$\mathbb{U}/P = \{\{0, 2, 4\}, \{1, 3, 5\}, \{6, 8, 10\}, \{7, 9\}\}$$

Let us consider a linguistic value *small* whose membership function is defined by

u	0	1	2	3	4	5	6	7	8	9	10
$\mu_{small}(u)$	1	1	0.75	0.5	0.25	0	0	0	0	0	0

The approximations of the fuzzy set μ_{small} in $\langle \mathbb{U}, P \rangle$ are given in Table 1. Then we obtain the mass assignment for the linguistic value *small*, and approximations of its focal sets given in Table 2.

Using Banerjee and Pal's notion, we obtain

$$\rho_P^{\alpha,\beta}(small) = \begin{cases} 1 & \text{for } \alpha > 0.25 \\ 0.5 & \text{for } 0.25 \geq \alpha > 0 \end{cases}$$

where the constraint $\alpha \geq \beta > 0$ is always assumed. On the other hand, the roughness by (19) yields

$$\hat{\rho}_P(small) = \sum_{\alpha \in \text{rng}(\mu_{small})} m_{small}(small_\alpha)(1 - \frac{|\underline{P}(small_\alpha)|}{|\overline{P}(small_\alpha)|}) = 0.875$$

Let $P^*(F)$ and $P_*(F)$ be fuzzy sets of $\mathbb{U}$ induced from the rough fuzzy set $(\underline{P}(F), \overline{P}(F))$ as in preceding section. Denote

Table 2 Mass assignment for *small* and approximations of its focal sets

$\text{rng}(\mu_{small})$	1	0.75	0.5	0.25
$small_\alpha$	{0, 1}	{0, 1, 2}	{0, 1, 2, 3}	{0, 1, 2, 3, 4}
$m_{small}(small_\alpha)$	0.25	0.25	0.25	0.25
$\underline{P}(small_\alpha)$	∅	∅	∅	{0, 2, 4}
$\overline{P}(small_\alpha)$	{0, 1, 2, 3, 4, 5}	{0, 1, 2, 3, 4, 5}	{0, 1, 2, 3, 4, 5}	{0, 1, 2, 3, 4, 5}

$$\mathrm{rng}(\mu_{P_*(F)}) \cup \mathrm{rng}(\mu_{P^*(F)}) = \{\omega_1, \ldots, \omega_p\}$$

such that $\omega_i > \omega_{i+1} > 0$ for $i = 1, \ldots, p-1$. Obviously, $\{\omega_1, \ldots, \omega_p\} \subseteq \mathrm{rng}(\mu_F)$. With these notations, the following holds [20]

Lemma 1. *For any* $1 \leq j \leq p$, *if there exists* $\alpha_i, \alpha_{i'} \in \mathrm{rng}(\mu_F)$ *such that* $\omega_{j+1} < \alpha_i < \alpha_{i'} \leq \omega_j$ *then we have* $F_{\alpha_i} \approx_P F_{\alpha'_i}$, *i.e.* $\underline{P}(F_{\alpha_i}) = \underline{P}(F_{\alpha'_i})$ *and* $\overline{P}(F_{\alpha_i}) = \overline{P}(F_{\alpha'_i})$, *and so* $\rho_R(F_{\alpha_i}) = \rho_R(F_{\alpha'_i})$.

Further, the following lemma is due to Dubois and Prade [10]

Lemma 2. *For any* $\alpha \in (0, 1]$, *we have*

$$P^*(F)_\alpha = \overline{P}(F_\alpha) \text{ and } P_*(F)_\alpha = \underline{P}(F_\alpha)$$

It then follows from Lemmas 1 and 2 that $\hat{\rho}_P(F)$ can be represented in terms of level sets of fuzzy sets $P_*(F)$ and $P^*(F)$ as the following proposition shows.

Proposition 2. $\hat{\rho}_P(F) = \sum_{j=1}^{p} (\omega_j - \omega_{j+1})(1 - \frac{|P_*(F)_{\omega_j}|}{|P^*(F)_{\omega_j}|})$, *where* $\omega_{p+1} = 0$, *by convention.*

Example 2. Let us continue with the approximation space $\langle U, P\rangle$ and the fuzzy set *small* given in Example 1. We have

$$\mathrm{rng}(\mu_{small}) = \{1, 0.75, 0.5, 0.25\}$$

By Table 1, we obtain

$$\mathrm{rng}(\mu_{P_*(small)}) \cup \mathrm{rng}(\mu_{P^*(small)}) = \{1, 0.25\}$$

which makes a partition of $\mathrm{rng}(\mu_{small})$ as $\{\{1, 0.75, 0.5\}, \{0.25\}\}$. It is easily to see that Table 2 illustrates for Lemma 1, and by Proposition 2 we get

$$\hat{\rho}_R(small) = (1 - 0.25)(1 - \frac{P_*(small)_1}{P^*(small)_1}) + 0.25(1 - \frac{P_*(small)_{0.25}}{P^*(small)_{0.25}}) = 0.875$$

which coincides with that given in Example 1.

Similar to the case of roughness of a crisp set, we have also the following proposition [20].

Proposition 3. *If fuzzy sets* F *and* G *in* U *are roughly equal in* $\langle U, R\rangle$, *then we have* $\hat{\rho}_R(F) = \hat{\rho}_R(G)$.

5.2 Approximation Quality of a Fuzzy Classification

Let P and Q be two equivalence relations over universal set $\mathbb{U}$. As mentioned above, P and Q may be induced respectively by sets of attributes applied to objects in $\mathbb{U}$. Then the approximation quality $\gamma_P(Q)$ of Q by P defined by (6) can be rewritten as

$$\gamma_P(Q) = \frac{1}{|\mathbb{U}|} \sum_{X \in \mathbb{U}/Q} |\underline{P}(X)| \tag{20}$$

In [34], Pawlak also defines the so-called *approximation accuracy* of Q by P, which extends the approximation accuracy of sets, by

$$\alpha_P(Q) = \frac{\sum_{X \in \mathbb{U}/Q} |\underline{P}(X)|}{\sum_{X \in \mathbb{U}/Q} |\overline{P}(X)|} \tag{21}$$

which is easily represented in terms of accuracies of sets as follows

$$\alpha_P(Q) = \sum_{X \in \mathbb{U}/Q} \frac{|\overline{P}(X)|}{\sum_{Y \in \mathbb{U}/Q} |\overline{P}(Y)|} \alpha_P(X)$$

That is, the approximation accuracy of a classification can be regarded as the convex sum of accuracies of its classes.

Furthermore, as mentioned in [34], the measure of approximation quality $\gamma_P(Q)$ does not capture how this partial dependency is actually distributed among classes of $\mathbb{U}/Q$. To capture this information we need the so-called *precision measure* $\pi_P(X)$, for $X \in \mathbb{U}/Q$, defined by

$$\pi_P(X) = \frac{|\underline{P}(X)|}{|X|} \tag{22}$$

Clearly, we have $\pi_R(X) \geq \alpha_P(X)$, for any $X \in \mathbb{U}/Q$. The two measures $\gamma_P(Q)$ and $\pi_P(X)$, $X \in \mathbb{U}/Q$, give us full information about the "classification power" of the knowledge P with respect to the classification $\mathbb{U}/Q$.

Now let us consider a fuzzy classification $\widetilde{Q}$ of $\mathbb{U}$ instead of a crisp one Q, i.e., $\mathbb{U}/\widetilde{Q}$ is a fuzzy partition of $\mathbb{U}$. This situation may naturally occur when a linguistic classification must be approximated in terms of already existing knowledge P. For example, assume that we have a personnel database given as $\mathbb{D} =$ PERSONNEL[ID; $Name$; $Position$; $Salary$], and attribute **Position** induces an approximation space $\langle \mathbb{D}, \text{IND}(Position) \rangle$. Given a linguistic description on the attribute **Salary**, say '$high$', it defines a fuzzy set on $\mathbb{D}$ denoted by $\mathbb{D}_{high}$. Then the accuracy of the fuzzy set $\mathbb{D}_{high}$, namely

$$\hat{\alpha}_{\text{IND}(Position)}(\mathbb{D}_{high}) = 1 - \hat{\rho}_{\text{IND}(Position)}(\mathbb{D}_{high})$$

may express the degree of completeness of our knowledge about the statement "Salary is $high$", given the granularity of $\mathbb{D}/\mathrm{IND}(Position)$. Further, a linguistic classification, say $\{low, medium, high\}$, may be imposed on the attribute **Salary** that induces a fuzzy partition of $\mathbb{D}$. Now one may want to measure a degree of dependency between "knowledge on attribute **Salary** expressed linguistically" and "knowledge on attribute **Position**".

In such a situation, guided by (20)–(21) and the random set based interpretation of a fuzzy set, the approximation quality and accuracy of a fuzzy classification $\widetilde{Q}$ by a crisp classification P can be defined [20, 21] as

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{1}{|\mathbb{U}|} \sum_{X \in \mathbb{U}/\widetilde{Q}} \sum_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\underline{P}(X_\alpha)| \tag{23}$$

and

$$\hat{\alpha}_P(\widetilde{Q}) = \frac{\sum\limits_{X \in \mathbb{U}/\widetilde{Q}} \sum\limits_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\underline{P}(X_\alpha)|}{\sum\limits_{X \in \mathbb{U}/\widetilde{Q}} \sum\limits_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\overline{P}(X_\alpha)|} \tag{24}$$

respectively, where for $X \in \mathbb{U}/\widetilde{Q}$, m_X stands for the mass assignment of X.

On the other hand, for each fuzzy class $X \in \mathbb{U}/\widetilde{Q}$, viewing $\underline{P}(X)$ as the induced fuzzy set $P_*(X)$ of $\mathbb{U}$ (refer to (10)) defined by

$$\mu_{P_*(X)}(x) = \mu_{\underline{P}(X)}([x]_P)$$

we can then define a counterpart of (7) for $\mathrm{POS}_P(\widetilde{Q})$ as a fuzzy set of $\mathbb{U}$ by

$$\mu_{\mathrm{POS}_P(\widetilde{Q})}(x) = \max_{X \in \mathbb{U}/\widetilde{Q}} \mu_{P_*(X)}(x) \tag{25}$$

Thus, guided by (6), another extension of the approximation quality can be also defined as

$$\hat{\gamma}'_P(\widetilde{Q}) = \frac{|\mathrm{POS}_P(\widetilde{Q})|}{|\mathbb{U}|} = \frac{\sum\limits_{x \in \mathbb{U}} \mu_{\mathrm{POS}_P(\widetilde{Q})}(x)}{|\mathbb{U}|} \tag{26}$$

Similarly, rewriting (21) as

$$\alpha_P(Q) = \frac{|\bigcup\limits_{X \in \mathbb{U}/Q} \underline{P}(X)|}{\sum_{X \in \mathbb{U}/Q} |\overline{P}(X)|}$$

suggests another extension of approximation accuracy of $\widetilde{Q}$ by P defined by

$$\hat{\alpha}'_P(\widetilde{Q}) = \frac{|\mathrm{POS}_P(\widetilde{Q})|}{\sum_{X\in\mathbb{U}/\widetilde{Q}} |\overline{P}(X)|} = \frac{\sum_{x\in\mathbb{U}} \mu_{\mathrm{POS}_P(\widetilde{Q})}(x)}{\sum_{X\in\mathbb{U}/\widetilde{Q}} \sum_{\alpha\in\mathrm{rng}(\mu_X)} m_X(X_\alpha)|\overline{P}(X_\alpha)|} \tag{27}$$

It is worth noting [20] here that the approximation quality and accuracy of $\widetilde{Q}$ by P defined by (23)–(24) can be respectively represented as

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{1}{|\mathbb{U}|} \sum_{X\in\mathbb{U}/\widetilde{Q}} |P_*(X)| = \frac{1}{|\mathbb{U}|} \sum_{X\in\mathbb{U}/\widetilde{Q}} \sum_{x\in\mathbb{U}} \mu_{P_*(X)}(x) \tag{28}$$

$$\hat{\alpha}_P(\widetilde{Q}) = \frac{\sum_{X\in\mathbb{U}/\widetilde{Q}} |P_*(X)|}{\sum_{X\in\mathbb{U}/\widetilde{Q}} |P^*(X)|} = \frac{\sum_{X\in\mathbb{U}/\widetilde{Q}} \sum_{x\in\mathbb{U}} \mu_{P_*(X)}(x)}{\sum_{X\in\mathbb{U}/\widetilde{Q}} \sum_{x\in\mathbb{U}} \mu_{P^*(X)}(x)} \tag{29}$$

which interestingly turn out to be natural extensions of (20) and (21), respectively, for the crisp case.

Clearly, two different, but equivalent, representations of $\gamma_P(Q)$ and $\alpha_P(Q)$ lead to various different extensions in the fuzzy case. Therefore, the natural question arises is that what extension should be used in practice. Theoretically, it seems difficult to give a satisfactory answer to the question, however, an appropriate selection could be made on the basis of experimental evaluations as usual for a given application.

In the following we consider a simple example to illustrate discussed extensions.

Example 3. Let us consider a relation in a relational database as shown in Table 3 (this database is a variant of that found in [8]).

Let P be the set of attributes **D** (degree) and **E** (experience). We then obtain an approximation space $\langle \mathbb{U}, P\rangle$, where $\mathbb{U} = \{1, \ldots, 16\}$, with the corresponding partition

Table 3 Relation in a relational database

ID	Degree	Experience	Salary	ID	Degree	Experience	Salary
1	Ph.D.	good	63K	9	M.S.	poor	41K
2	Ph.D.	very poor	47K	10	M.S.	very good	68K
3	M.S.	good	53K	11	M.S.	good	50K
4	B.S.	very poor	26K	12	B.S.	very poor	23K
5	B.S.	poor	29K	13	M.S.	good	55K
6	Ph.D.	very poor	50K	14	M.S.	good	51K
7	B.S.	poor	35K	15	Ph.D.	good	65K
8	M.S.	poor	40K	16	M.S.	very good	64K

August 2, 2007 Time: 06:14pm t1-v1.1

AU: Please provide citation for Fig 1.

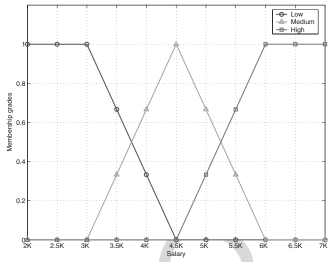

Fig. 1 A linguistic partition of attribute *salary*

$$\mathbb{U}/P = \{\{1, 15\}, \{2, 6\}, \{3, 11, 13, 14\},\ \{4, 12\}, \{5, 7\}, \{8, 9\}, \{10, 16\}\}$$

Further, consider now for example a linguistic classification over attribute **S** (salary), i.e. $\widetilde{Q} = \{\mathrm{S}\}$, with membership functions of linguistic classes **Low**, **Medium**, **High** graphically depicted as in Fig. 2. Then the linguistic classification induces a fuzzy partition $\mathbb{U}/\widetilde{Q}$ whose membership functions of fuzzy classes are shown in Table 4.

Then approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in the approximation space $\langle \mathbb{U}, P \rangle$ are given in Table 5.

Using (28) and (29) we obtain

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{13.46}{16} = 0.84, \text{ and } \hat{\alpha}_P(\widetilde{Q}) = \frac{13.46}{18.21} = 0.739$$

respectively. That is, we have the following partial dependency in the database

$$\{\mathrm{D}, \mathrm{E}\} \Rightarrow_{0.84} \mathrm{S} \tag{30}$$

Note that making use of (26) and (27) gives us

Table 4 Induced fuzzy partition of $\mathbb{U}$ based on *salary*

$\mathbb{U}$	μ_{Low}	μ_{Medium}	μ_{High}	$\mathbb{U}$	μ_{Low}	μ_{Medium}	μ_{High}
1	0	0	1	9	0.27	0.73	0
2	0	0.87	0.13	10	0	0	1
3	0	0.47	0.53	11	0	0.67	0.33
4	1	0	0	12	1	0	0
5	1	0	0	13	0	0.33	0.67
6	0	0.67	0.33	14	0	0.6	0.4
7	0.67	0.33	0	15	0	0	1
8	0.33	0.67	0	16	0	0	1

Table 5 The approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in $\langle \mathbb{U}, P \rangle$

X_i	{1, 15}	{2, 6}	{3, 11, 13, 14}	{4, 12}	{5, 7}	{8, 9}	{10, 16}
$\mu_{P_*(High)}$	1	0.13	0.33	0	0	0	1
$\mu_{P^*(High)}$	1	0.33	0.67	0	0	0	1
$\mu_{P_*(Medium)}$	0	0.67	0.33	0	0	0.67	0
$\mu_{P^*(Medium)}$	0	0.87	0.67	0	0.33	0.73	0
$\mu_{P_*(Low)}$	0	0	0	1	0.67	0.27	0
$\mu_{P^*(Low)}$	0	0	0	1	1	0.33	0

$$\hat{\gamma}'_P(\widetilde{Q}) = \frac{11.34}{16} = 0.709, \text{ and } \hat{\alpha}'_P(\widetilde{Q}) = \frac{11.34}{13.88} = 0.82$$

Now in order to show how the influence of, for example, attribute **E** on the quality of approximation, let us consider the partition induced by the relation $R = P\backslash\{\text{E}\} = \{\text{D}\}$ as follows

$$U/R = \{\{1, 2, 6, 15\}, \{3, 8, 9, 10, 11, 13, 14, 16\}, \{4, 5, 7, 12\}\}$$

Then we obtain approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in the approximation space $\langle \mathbb{U}, R \rangle$ given in Table 6.

Thus we have

$$\hat{\gamma}_R(\widetilde{Q}) = \hat{\gamma}_{P\backslash\{\text{E}\}}(\widetilde{Q}) = \frac{3.2}{16} = 0.2$$

Similarly, we also easily obtain

$$\hat{\gamma}_{P\backslash\{\text{D}\}}(\widetilde{Q}) = \frac{5.06}{16} = 0.316$$

As we can see, both attributes **D** and **E** are highly significant because without each of them the approximation quality $\hat{\gamma}_P(\widetilde{Q})$ changes considerably.

Table 6 The approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in $\langle U, R \rangle$

X_i	{1, 2, 6, 15}	{3, 8, 9, 10, 11, 13, 14, 16}	{4, 5, 7, 12}
$\mu_{R_*(High)}$	0.13	0	0
$\mu_{R^*(High)}$	1	1	0
$\mu_{R_*(Medium)}$	0	0	0
$\mu_{R^*(Medium)}$	0.87	0.73	0.33
$\mu_{R_*(Low)}$	0	0	0.67
$\mu_{R^*(Low)}$	0	0.33	1

6 Approximation Quality Based on Fuzzy-Rough Sets

Let us turn to a fuzzy approximation space $\langle \mathbb{U}, P\rangle$, where P is a fuzzy similarity relation over universe $\mathbb{U}$. This fuzzy similarity relation induces a fuzzy partition over $\mathbb{U}$ denoted by $\mathbb{U}/P$ as mentioned previously. Assume now that $\mathbb{U}/Q$ is another (fuzzy) partition of $\mathbb{U}$. In order to have a counterpart of (6) for the approximation quality of Q by P in this situation, one needs to define the fuzzy positive region $\mathrm{POS}_P(Q)$ which is regarded as a fuzzy set of $\mathbb{U}$. Then, once having defined the fuzzy positive region, an extension of the approximation quality of Q by P can be defined [24, 28] as follows

$$\hat{\gamma}_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|\mathbb{U}|} = \frac{\sum_{x\in\mathbb{U}} \mu_{\mathrm{POS}_P(Q)}(x)}{|\mathbb{U}|} \tag{31}$$

where Σ-count is used for the cardinality of a fuzzy set.

In the case that the knowledge of P is not given directly but, instead, a fuzzy partition $\mathbb{U}/P$ is predefined, Jensen and Shen [24, 25] have defined the membership function of fuzzy positive region $\mathrm{POS}_P(Q)$, for any object $x \in \mathbb{U}$, as

$$\mu_{\mathrm{POS}_P(Q)}(x) = \sup_{X\in\mathbb{U}/Q} \mu_{\underline{P}(X)}(x) \tag{32}$$

where the membership function $\mu_{\underline{P}(X)}(x)$ of fuzzy lower approximations can be defined by (16). Note that when $\mathbb{U}/P$ is a crisp partition, (31) is identical to (26) above. This approach has been successfully used for the task of feature reduction for crisp and real-valued datasets in various applications of data mining [22, 23, 24, 25, 44].

In particularly, regarding the issue of feature reduction in crisp and real-valued datasets, each real-valued attribute a is first associated with a fuzzy linguistic partition denoted by $\mathbb{U}/\{a\}$, then the fuzzy partition $\mathbb{U}/P$ induced by a set P of attributes defined over objects in $\mathbb{U}$ is defined as a fuzzy counterpart of (3) as follows

$$\mathbb{U}/P = \bigotimes_{a\in P} \mathbb{U}/\{a\} \tag{33}$$

where t-norm min is used for the fuzzy intersection. On the basis of these above extensions, a fuzzy-rough based method of attribute reduction described by the so-called fuzzy-rough QuickReduct algorithm has been proposed and applied to Web categorization in [24] and complex systems monitoring [25].

The following simple example taken from [24] will illustrate how these extensions work.

Example 4. Let us consider an example data set and fuzzy sets N and Z given in Fig. 2. Here, for illustrative simplicity, the fuzzy sets are viewed as fuzzy classes defined for all real-valued attributes.

Then we have the following partitions induced from corresponding individual attributes

Object	a	b	c	q
1	-.4	-.3	-.5	no
2	-.4	.2	-.1	yes
3	-.3	-.4	-.3	no
4	.3	-.3	0	yes
5	.2	-.3	0	yes
6	.2	0	0	no

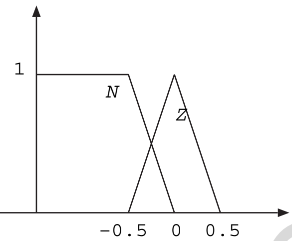

Fig. 2 Data set and corresponding fuzzy sets

$$\mathbb{U}/A = \{N_a, Z_a\},\ \mathbb{U}/B = \{N_b, Z_b\},$$
$$\mathbb{U}/C = \{N_c, Z_c\},\ \mathbb{U}/Q = \{\{1,3,6\},\{2,4,5\}\},$$

where $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $Q = \{q\}$ and membership functions of corresponding fuzzy classes are given in Table 7.

The following fuzzy partitions induced from subsets of conditional attributes are obtained by (33)

$$\begin{aligned}
\mathbb{U}/\{a,b\} &= \{N_a \cap N_b, N_a \cap Z_b, Z_a \cap N_b, Z_a \cap Z_b\},\\
\mathbb{U}/\{b,c\} &= \{N_b \cap N_c, N_b \cap Z_c, Z_b \cap N_c, Z_b \cap Z_c\},\\
\mathbb{U}/\{a,c\} &= \{N_a \cap N_c, N_a \cap Z_c, Z_a \cap N_c, Z_a \cap Z_c\},\\
\mathbb{U}/\{a,b,c\} &= \{N_a \cap N_b \cap N_c, N_c \cap N_b \cap Z_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c,\\
&\quad N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c\}
\end{aligned}$$

where $\cap = \min$. Using (16) and (31) respectively for calculating fuzzy lower approximations and the approximation quality, we obtain

$$\hat{\gamma}_A(Q) = \frac{2}{6}, \hat{\gamma}_B(Q) = \frac{2.4}{6}, \hat{\gamma}_C(Q) = \frac{1.6}{6}, \hat{\gamma}_{\{a,b\}}(Q) = \frac{3.4}{6}$$

$$\hat{\gamma}_{\{b,c\}}(Q) = \frac{3.2}{6}, \hat{\gamma}_{\{a,c\}}(Q) = \frac{3.2}{6}, \hat{\gamma}_{\{a,b,c\}}(Q) = \frac{3.4}{6}$$

Table 7 Membership functions of corresponding fuzzy classes

Object	*a*		*b*		*c*		*q*	
	N_a	Z_a	N_b	Z_b	N_c	Z_c	{1, 3, 6}	{2, 4, 5}
1	0.8	0.2	0.6	0.4	1.0	0.0	1.0	0.0
2	0.8	0.2	0.0	0.6	0.2	0.8	0.0	1.0
3	0.6	0.4	0.8	0.2	0.6	0.4	1.0	0.0
4	0.0	0.4	0.6	0.4	0.0	1.0	0.0	1.0
5	0.0	0.6	0.6	0.4	0.0	1.0	0.0	1.0
6	0.0	0.6	0.0	1.0	0.0	1.0	1.0	0.0

From these results it can be seen that the attribute c is not significant at all because removing it from the set of conditional attributes does not cause any change in the approximation quality, i.e. c is dispensable. Details on the fuzzy-rough QuickReduct algorithm as well as how it could be applied to generate the Q-reduct $\{a, b\}$ of $P = \{a, b, c\}$ for this example can be referred to [22, 24, 25].

In the study of fuzzy information systems, in which attribute values of object may be fuzzy (linguistic) values, Mieszkowicz-Rolka and Rolka [28] proposed to define a so-called compatibility relation over $\mathbb{U}$ induced from a set of attributes P as follows

$$\mu_P(x, y) = \min_{a \in P} \sup_{v \in V_a} \min(\mu_{f(x,a)}(v), \mu_{f(y,a)}(v)) \tag{34}$$

where V_a is the domain of attribute a; $f(x, a)$ and $f(y, a)$ are fuzzy values of x and y at attribute a, respectively. Using this definition of a fuzzy similarity relation, fuzzy lower approximations of fuzzy sets can be defined using (18) and then (31) can be also used to define the approximation quality in case of fuzzy information systems.

Similarly, as discussed in the preceding section, it is of interest to mention here that equivalent representation of the approximation quality $\gamma_P(Q)$ by (20) may also suggest another extension for $\hat{\gamma}_P(Q)$. However, due to overlapping of fuzzy lower approximations, in this case we may need to carry out some normalization. For example, we can normalize involved fuzzy similarity relations so that (12) is satisfied, then a fuzzy counterpart of (20) can be used to define an extension for $\hat{\gamma}_P(Q)$. Another possibility is that we can carry out a normalization after defining a fuzzy counterpart of (20), for instance, as follows

$$\hat{\gamma}_P(Q) = \frac{1}{|\mathbb{U}||\mathbb{U}/Q|} \sum_{X \in \mathbb{U}/Q} \sum_{x \in \mathbb{U}} \mu_{\underline{P}(X)}(x) \tag{35}$$

Intuitively, we may observe that if the fuzzy lower approximation of some (fuzzy) class in $\mathbb{U}/Q$ dominates all those of the others, it solely affects the approximation quality $\hat{\gamma}_P(Q)$ defined by (31), while others classes play no role. This situation does not occur in the crisp case because of the disjoint union. In such a situation, an extension for $\hat{\gamma}_P(Q)$ guided by (20) may be interesting to be considered since, in any case, it takes fuzzy lower approximations of all classes in $\mathbb{U}/Q$ into account. This, however, requires further research.

7 Conclusion and Future Work

The concepts of approximation quality essentially play an important role in practical applications of rough set theory. They supply numerical characterizations for measuring the dependency between attributes in databases and the accuracy of concept approximation using the given data alone and no additional information. At the same time, rough-fuzzy hybrids have emerged naturally due to the need of encapsulating the related but distinct concepts of vagueness and indiscernibility, both of which occur as a result of imperfection in knowledge. This review paper has focused on those extensions of approximation quality that make use of rough fuzzy and fuzzy rough sets. We have also discussed how different but equivalent representations of approximation quality in the (crisp) rough case may lead to various different extensions for rough-fuzzy cases. However, much research work should be done in the future to explore theoretical features as well as practical implications of these mentioned extensions.

Let us conclude here by pointing out some issues regarding the research topic discussed, which would be interestingly considered for further research:

- Exploiting practical applications of roughness measure for fuzzy sets, particularly in classification and image analysis problems as pointed out in [3, 42], as well as its generalization in a fuzzy approximation space.
- Apart from those having been well studied, formulating and investigating other extensions of the approximation quality, for example as mentioned in the preceding section, and conducting comparative experiments to verify their applicability in, for example, dimensionality reduction in comparison with known extensions as studied in [22, 23, 24, 25, 44].
- Using rough-fuzzy hybrids based extensions of the approximation quality in areas of decision analysis [16], case-based reasoning [32] and knowledge discovery [39].
- Studying extensions of approximation quality in variable precision fuzzy rough sets model [29, 54] and their applicability.

Acknowledgment The authors would like to thank two anonymous reviewers and Editors for their helpful comments and suggestions.

References

1. A. Apostolico, E. R. Caianiello, E. Fischetti, S. Vitulano, C-Calculus: an elementary approach to some problems in pattern recognition, **Pattern Recognition 19** (1978) 375–387.
2. J. F. Baldwin, The management of fuzzy and probabilistic uncertainties for knowledge based systems, in: S.A. Shapiro (Ed.), *The Encyclopaedia of AI*, Wiley, New York, 1992, pp. 528–537.
3. M. Banerjee, S. K. Pal, Roughness of a fuzzy set, *Information Sciences* **93** (1996) 235–246.
4. R. B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, *Pattern Recognition Letters* **26** (2005) 965–975.
5. E.R. Caianiello, A. Petrosino, Neural networks, fuzziness and image processing, in V. Cantoni (Ed.): **Machine and Human Perception: Analogies and Divergences**, Plenum Press, 1994, pp. 355–370.
6. M. Ceccarelli, A. Petrosino, A parallel fuzzy scale-space approach to the unsupervised texture separation, **Pattern Recognition Letters 23** (2002) 557–567.
7. M. Ceccarelli, F. Musacchia, A. Petrosino, A fuzzy scale-space approach to feature-based image representation and retrieval, in M. De Gregorio et al. (Eds.): **Brain, Vision, and Artificial**

Intelligence, BVAI'2005, Naples, Italy, 2005, LNCS **3704**, Springer-Verlag, pp. 377–385.

8. S. M. Chen, C. M. Huang, Generating weighted fuzzy rules from relational database systems for estimating null values using genetic algorithms, *IEEE Transactions on Fuzzy Systems* **11** (2003) 495–506.
9. A. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorisation, **Applied Artificial Intelligence** 15 (9) (2001) 843–873.
10. D. Dubois, H. Prade, Putting fuzzy sets and rough sets together, in R. Słowiński (Ed.), *Intelligent Decision Support*, Kluwer Academic, Dordrecht, 1992, pp. 203–232.
11. D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems* **17** (1990) 191–209.
12. D. Dubois, H. Prade, Fuzzy sets, possibility and measurement, *European Journal of Operational Research* **40** (1989) 135–154.
13. G. Gediga, I. Düntsch, Rough approximation quality revisited, *Artificial Intelligence* **132** (2001) 219–234.
14. G. Gediga, I. Düntsch, Approximation quality for sorting rules, **Computational Statistics & Data Analysis 40** (2002) 499–526.
15. S. Greco, B. Matarazzo, R. Słowiński, Rough set handling of ambiguity, in H.-J. Zimmermann (Ed.), **Proceedings of the Sixth European Congress on Intelligent Techniques and Soft Computing** (EUFIT98), Mainz Wissenschaftsverlag, Aachen, 1998, pp. 3–14.
16. S. Greco, B. Matarazzo, R. Słowiński, Rough sets theory for multicriteria decision analysis, *European Journal of Operational Research* **129** (2001) 1–47.
17. S. Greco, M. Inuiguchi, R. Slowinski, Fuzzy rough sets and multiple-premise gradual decision rules, *International Journal of Approximate Reasoning* **41** (2006) 179–211.
18. S. Greco, M. Inuiguchi, R. Slowinski, A new proposal for fuzzy rough approximations and gradual decision rule representation, in J. F. Peters et al. (Eds.), **Transactions on Rough Sets II: Rough Sets and Fuzzy Sets**, LNCS **3135**, 2004, pp. 319–342.
19. U. Höhle, Quotients with respect to similarity relations, *Fuzzy Sets and Systems* **27** (1988) 31–44.
20. V. N. Huynh, Y. Nakamori, A rouhgness measure for fuzzy sets, *Information Sciences* **173** (2005) 255–275.
21. V. N. Huynh, T. Murai, T. B. Ho, Y. Nakamori, An extension of rough approximation quality to fuzzy classification, in D. Slezak et al.(Eds.) *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNCS **3641**, 2005, pp. 512–521.
22. R. Jensen, **Combining Rough and Fuzzy Sets for Feature Selection**, PhD Thesis, University of Edinburgh, UK, 2005.
23. R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization, **Fuzzy Sets and Systems 149** (2005) 5–20.
24. R. Jensen, Q. Shen, Fuzzy-rough attribute reduction with application to web categorization, **Fuzzy Sets and Systems 141** (2004) 469–485.
25. R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches, **IEEE Transactions on Knowledge and Data Engineering 16** (2004) 1457–1471.
26. R. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall PTR, Upper Saddle River, NJ, 1995.
27. E. Marczewski, H. Steinhaus, On a certain distance of sets and the corresponding distance of functions, *Colloquium Mathematicum* **6** (1958) 319–327.
28. A. Mieszkowicz-Rolka, L. Rolka, Fuzziness in information systems, **Electronic Notes in Theoretical Computer Science 82** (4) (2003) 1–10.
29. A. Mieszkowicz-Rolka, L. Rolka, Remarks on approximation quality in variable precision fuzzy rough sets model, in S. Tsumoto et al. (Eds.), RSCTC'2004, LNAI **3066**, Springer-Verlag, 2004, pp. 402–411.
30. T. Murai, H. Kanemitsu, M. Shimbo, Fuzzy sets and binary-proximity-based rough sets, *Information Sciences* **104** (1–2) (1998) 49–80.
31. S. K. Pal, A. Skowron (Eds.), *Rough Fuzzy Hybridization: New Trends in Decision Making*, Springer Verlag, Singapore, 1999.

32. S. K. Pal, P. Mitra, Case generation using rough sets with fuzzy representation, **IEEE Transactions on Knowledge and Data Engineering 16** (2004) 292–300.
33. Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences* **11** (1982) 341–356.
34. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, MA, 1991.
35. Z. Pawlak, Rough sets and fuzzy sets, *Fuzzy Sets and Systems* **17** (1985) 99–102.
36. A. Petrosino, Rough fuzzy sets and unsupervised neural learning: applications in computer vision. In: A. Bonarini et al. (Eds.), **New trends in Fuzzy Logic**. World Scientific, Singapore, 1996, pp. 166–176.
37. A. Petrosino, M. Ceccarelli, M., Unsupervised texture discrimination based on rough fuzzy sets and parallel hierarchical clustering, **Proceedings of IEEE International Conf. on Pattern Recognition**, IEEE-CS Press, Silver Spring 2000, pp. 1100–1103.
38. A. Petrosino, G. Salvi, Rough fuzzy set based scale space transforms and their use in image analysis, *International Journal of Approximate Reasoning* **41** (2006) 212–228.
39. L. Polkowski, A. Skowron (Eds.), *Rough Sets in Knowledge Discovery*, Parts 1 & 2, Physica-Verlag, Heidelberg-New York, 1998.
40. A. M. Radzikowska, E. E. Kerre, A comparative study of fuzzy rough sets, **Fuzzy Sets and Systems 126** (2002) 137–155.
41. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).
42. M. Sarkar, Rough–fuzzy functions in classification, *Fuzzy Sets and Systems* **132** (2002) 353–369.
43. Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, **Pattern Recognition 35** (2002) 2425–2438.
44. Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, **Pattern Recognition 37** (2004) 1351–1363.
45. R. Słowiński (Ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic, Dordrecht/Boston/London, 1992.
46. W. Z. Wu, W. X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, *Information Sciences* **159** (2004) 233–254.
47. M. Wygralak, Rough sets and fuzzy sets: some remarks on interrelations, *Fuzzy Sets and Systems* **29** (1989) 241–243.
48. Y. Y. Yao, Information granulation and rough set approximation, *International Journal of Intelligent Systems* **16** (2001) 87–104.
49. Y. Y. Yao, Combination of rough and fuzzy sets based on alpha-level sets, in T. Y. Lin & N. Cercone (Eds.), *Rough Sets and Data Mining: Analysis of Imprecise Data* (Kluwer Academic Publishers, Boston/London/Dordrecht, 1997) 301–321.
50. Y. Y. Yao, A comparative study of fuzzy sets and rough sets, *Information Sciences* **109** (1998) 227–242.
51. L. A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965) 338–353.
52. L. A. Zadeh, Similarity relations and fuzzy orderings, *Information Sciences* **3** (1971) 177–200.
53. H. Zhang, H. Liang, D. Liu, Two new operators in rough set theory with applications to fuzzy sets, *Information Sciences* **166** (2004) 147–165.
54. W. Ziarko, Variable precision rough set model, **Journal of Computer and System Sciences 46** (1993) 39–59.