Feature Selection with Fuzzy Decision Reducts
Cornelis, Chris; Hurtado Martín, Germán; Jensen, Richard; Izak, Dominik

Published in:
Rough Sets and Knowledge Technology
DOI:
10.1007/978-3-540-79721-0_41
Publication date:
2008

Citation for published version (APA):
Feature Selection with Fuzzy Decision Reducts

Chris Cornelis\textsuperscript{1}, Germán Hurtado Martín\textsuperscript{1,2},
Richard Jensen\textsuperscript{3}, and Dominik Ślęzak\textsuperscript{4}

\textsuperscript{1} Dept. of Mathematics and Computer Science, Ghent University, Gent, Belgium
\{Chris.Cornelis,German.HurtadoMartin\}@UGent.be
\textsuperscript{2} Dept. of Industrial Sciences, Hogeschool Gent, Gent, Belgium
\textsuperscript{3} Dept. of Computer Science, The University of Wales, Aberystwyth, UK, rkj@aber.ac.uk
\textsuperscript{4} Infobright Inc., Toronto, Canada, slezak@infobright.com

Abstract. In this paper, within the context of fuzzy rough set theory, we generalize the classical rough set framework for data-based attribute selection and reduction, based on the notion of fuzzy decision reducts. Experimental analysis confirms the potential of the approach.

Keywords: fuzzy sets, rough sets, decision reducts, classification

1 Introduction

Rough set theory \cite{Pawlak1982} is well-suited to semantics-preserving data dimensionality reduction, i.e.: to omit attributes (features) from decision systems without sacrificing the ability to discern between objects belonging to different concepts or classes. A minimal set of attributes that preserves the decision making power of the original system is called a decision reduct.

Traditionally, discernibility is modeled by an equivalence relation in the set of objects: objects are indiscernible w.r.t. a given set of attributes $B$ if they have the same values for all attributes in $B$. This works well for qualitative data, in particular if the number of distinct values for each attribute is limited and there is no particular relationship among them. Quantitative data, however, involve continuous (i.e., real-valued) attributes like age, speed or length, and are tied to a natural scale of closeness, loosely expressing that the closer the attribute values of two objects are, the less discernible they are. While the standard methodology can be tailored to handle them by applying discretization, it is more natural to consider a notion of approximate equality between objects. Formally, such a notion can be modeled by means of a fuzzy relation \cite{Zadeh1971} in the set of objects.

Guided by this principle, the original rough set framework for data-based attribute selection and reduction can be generalized (see e.g. \cite{Cornelis2007,Cornelis2008,Cornelis2009}). This paper differs from previous research efforts by the introduction of the concept of a fuzzy decision reduct: conceptually, this is a weighted version of its classical counterpart that assigns to each attribute subset the degree to which it preserves the predictive ability of the original decision system. We consider alternative
ways of defining fuzzy decision reducts, grouped along two directions: the first direction works with an extension of the well-known positive region, while the second one is based on an extension of the discernibility function from classical rough set analysis.

The remainder of this paper is organized as follows: we first recall preliminaries of rough sets, fuzzy sets and their hybridization in Section 2. In Section 3, we propose a general definition of a fuzzy decision reduct, and then develop a number of concrete instances of it. In Section 4, experiments are conducted to evaluate the effectiveness of these alternatives. Finally, in Section 5 we conclude.

2 Preliminaries

2.1 Rough Set Theory

**Definitions** In rough set analysis, data is represented as an information system $(X, \mathcal{A})$, where $X = \{x_1, \ldots, x_n\}$ and $\mathcal{A} = \{a_1, \ldots, a_m\}$ are finite, non-empty sets of objects and attributes, respectively. Each $a$ in $\mathcal{A}$ corresponds to an $X \rightarrow V_a$ mapping, in which $V_a$ is the value set of $a$ over $X$. For every subset $B$ of $\mathcal{A}$, the $B$-indiscernibility relation $R_B$ is defined as $R_B = \{(x, y) \in X^2 \text{ and } (\forall a \in B)(a(x) = a(y))\}$. Clearly, $R_B$ is an equivalence relation. Its equivalence classes $[x]_{R_B}$ can be used to approximate concepts, i.e., subsets of the universe $X$.

Given $A \subseteq X$, its lower and upper approximation w.r.t. $R_B$ are defined by $R_B \downarrow A = \{x \in X|[x]_{R_B} \subseteq A\}$ and $R_B \uparrow A = \{x \in X|[x]_{R_B} \cap A \neq \emptyset\}$.

A decision system $(X, \mathcal{A} \cup \{d\})$ is a special kind of information system, used in the context of classification, in which $d (d \notin \mathcal{A})$ is a designated attribute called decision. Based on the values $v_k$ that $d$ assumes (drawn from the finite $v_a$ set $V_d$), $X$ is partitioned into a number of decision classes $X_k$. Given $B \subseteq \mathcal{A}$, the $B$-positive region $POS_B = \bigcup_{v_k \in V_d} R_B \downarrow X_k$ contains the objects for which the values of $B$ allow to predict the decision class unequivocally. The predictive ability w.r.t. $d$ of the attributes in $B$ is then measured by $\gamma_B = \frac{|POS_B|}{|X|}$ (degree of dependency of $d$ on $B$). A subset $B$ of $\mathcal{A}$ is called a decision reduct if $POS_B = POS_A$, i.e., $B$ preserves the decision making power of $\mathcal{A}$, and if it cannot be further reduced, i.e., there exists no proper subset $B'$ of $B$ such that $POS_{B'} = POS_A$.

**Example 1.** Consider the decision system in Table 1.a). There are two decision classes: $X_0$ contains all $x$ for which $d(x) = 0$, while $X_1$ contains those with $d(x) = 1$. If we want to apply the standard rough set analysis approach, we first have to discretize the system; a possible discretization is given in Table 1.b). Then we can calculate the positive region. For example, for $B = \{a_4, a_5\}$, $POS_B = \{x_1, x_5, x_6, x_7\}$. Also, $POS_A = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$.

---

5 In this paper, we assume that decisions are always qualitative (discrete-valued).

6 This is a sample taken from the Pima Indians Diabetes dataset, available at http://www.ics.uci.edu/~mlearn/MLRepository.html
Finding Decision Reducts Decision reducts are used to synthesize minimal decision rules, which result from overlaying the reducts over the decision system and reading off the values. Below we recall a well-known approach to generate all reducts of a decision system based on its decision-relative discernibility matrix and function [7]. The decision-relative discernibility matrix of \((X, A ∪ \{d\})\) is the \(n \times n\) matrix \(O\), defined by, for \(i, j \in \{1, ..., n\}\), \(O_{ij} = 0\) if \(d(x_i) = d(x_j)\) and \(O_{ij} = \{a ∈ A | a(x_i) ≠ a(x_j)\}\) otherwise. On the other hand, the discernibility function of \((X, A ∪ \{d\})\) is the \(\{0, 1\}^n → \{0, 1\}\) mapping \(f\), defined by \(f(a^*_1, ..., a^*_m) = ∨\{O_{ij} | 1 ≤ j < i ≤ n\}\), in which \(O^*_ij = \{a^* | a ∈ O_{ij}\}\). The boolean variables \(a^*_1, ..., a^*_m\) correspond to the attributes from \(A\), and we denote \(A^* = \{a^*_1, ..., a^*_m\}\). If \(B ⊆ A\), then the valuation function \(V_B\) corresponding to \(B\) is defined by \(V_B(a^*) = 1\) if \(a ∈ B\). This valuation can be extended to arbitrary formulas, such that \(V_B(f(a^*_1, ..., a^*_m)) = f(V_B(a^*_1), ..., V_B(a^*_m))\). This expresses whether the attributes in \(B\) preserve the discernibility of \((X, A ∪ \{d\})\) (when its value is 1) or not (when it is 0). The discernibility function can be reduced to its disjunctive normal form, that is \(f(a^*_1, ..., a^*_m) = ∨\{A^*_1 \land ... \land A^*_p\}\), in which \(p ≥ 1\), and for all \(i \in \{1, ..., p\}\) it holds that \(A^*_i ⊆ A^*\), and \(A^*_i ∉ A^*_j\) for \(i ≠ j\). If we define \(a^* ∈ A^*_i\), then it can be shown that \(A_i^*, ..., A_p^*\) constitute exactly all decision reducts of \((X, A ∪ \{d\})\).

Example 2. The reduced discernibility function of the decision system in Table 1.b) is given by \(f(a^*_1, ..., a^*_8) = a^*_2 ∨ (a^*_1 ∧ a^*_7) ∨ (a^*_4 ∧ a^*_7) ∨ (a^*_5 ∧ a^*_7) ∨ (a^*_6 ∧ a^*_7)\). Hence, the decision reducts are \(\{a^*_1, a^*_7\}, \{a^*_5, a^*_7\}, \{a^*_6, a^*_7\}\) and \(\{a^*_7, a^*_8\}\).

### 2.2 Fuzzy Set Theory

Recall that a fuzzy set [10] in \(X\) is an \(X → [0, 1]\) mapping, while a fuzzy relation in \(X\) is a fuzzy set in \(X × X\). For all \(y \in X\), the \(R\)-foreset of \(y\) is the fuzzy set \(R_y\) defined by \(R_y(x) = R(x, y)\) for all \(x \in X\). If \(R\) is reflexive and symmetric, i.e., \(R(x, x) = 1\) and \(R(x, y) = R(y, x)\) hold for all \(x, y \in X\), then \(R\) is called a fuzzy tolerance relation. For fuzzy sets \(A\) and \(B\) in \(X\), \(A ⊆ B ⇔ (∀x ∈ X)(A(x) ≤ B(x))\). If \(X\) is finite, the cardinality of \(A\) equals \(|A| = ∑_{x ∈ X} A(x)\).

Fuzzy logic connectives play an important role in the development of fuzzy rough set theory. We therefore recall some important definitions. A triangu-
lar norm (t-norm for short) $\mathcal{T}$ is any increasing, commutative and associative $[0,1]^2 \to [0,1]$ mapping satisfying $\mathcal{T}(1, x) = x$, for all $x$ in $[0,1]$. In this paper, we consider $\mathcal{T}_M$ and $\mathcal{T}_L$, defined by $\mathcal{T}_M(x, y) = \min(x, y)$ and $\mathcal{T}_L(x, y) = \max(0, x + y - 1)$ for $x, y$ in $[0,1]$. An implicator is any $[0,1]^2 \to [0,1]$-mapping $\mathcal{I}$ satisfying $\mathcal{I}(0, 0) = 1$, $\mathcal{I}(1, x) = x$, for all $x$ in $[0,1]$. Moreover, we require $\mathcal{I}$ to be decreasing in its first, and increasing in its second component. In this paper, we consider $\mathcal{I}_M$ and $\mathcal{I}_L$, defined by, for $x, y$ in $[0,1]$, $\mathcal{I}_M(x, y) = 1$ if $x \leq y$ and $\mathcal{I}_M(x, y) = y$ otherwise, and $\mathcal{I}_L(x, y) = \min(1,1 - x + y)$.

### 2.3 Fuzzy Rough Set Theory

Research on hybridizing fuzzy sets and rough sets has focused mainly on fuzzifying the formulas for lower and upper approximation. In this process, the set $A$ is generalized to a fuzzy set in $X$, allowing that objects can belong to a concept to varying degrees. Also, rather than assessing objects’ indiscernibility, we may measure their closeness, represented by a fuzzy tolerance relation $R$. For the lower and upper approximation of $A$ by means of a fuzzy tolerance relation $R$, we adopt the definitions proposed in [6]: given an implicator $\mathcal{I}$ and a t-norm $\mathcal{T}$, $R \uparrow A$ and $R \downarrow A$ are defined by, for all $y$ in $X$, $\mathcal{I}(R \uparrow A)(y) = \inf_{x \in X} \mathcal{I}(R(x, y), A(x))$ and $\mathcal{I}(R \downarrow A)(y) = \sup_{x \in X} \mathcal{I}(R(x, y), A(x))$.

In this paper, given a quantitative attribute $a$ with range $l(a)$, we compute the approximate equality between two objects w.r.t. $a$, by the parametrized relation $R_a$, defined by, for $x$ and $y$ in $X$, $R_a(x, y) = \max\left(0, \min\left(1, \beta - \alpha \frac{|a(x) - a(y)|}{l(a)}\right)\right)$. The parameters $\alpha$ and $\beta$ ($\alpha \geq \beta \geq 1$) determine the granularity of $R_a$.

Discernibility, or distance, of two objects $x$ and $y$ w.r.t. $a$ can be computed as the complement of their closeness: $1 - R_a(x, y)$. Assuming that for a qualitative (i.e., nominal) attribute $a$, the classical way of discerning objects is used, i.e., $R_a(x, y) = 1$ if $a(x) = a(y)$ and $R_a(x, y) = 0$ otherwise, we can define, for any subset $B$ of $A$, the fuzzy $B$-indiscernibility relation by $R_B(x, y) = \min_{a \in B} R_a(x, y)$.

It can easily be seen that $R_B$ is a fuzzy tolerance relation, and also that if only qualitative attributes (possibly stemming from discretization) are used, then the traditional concept of $B$-indiscernibility relation is recovered.

**Example 3.** For the non-dicretized decision system in Table 1a), assume that $\alpha = 5$ and $\beta = 1.2$ are used in $R_a$ for each attribute $a$, and that the attributes’ ranges are determined by the minimal and maximal occurring values in the decision system. It can be verified e.g. that $R_{a_1}(x_2, x_3) = 0.575$, $R_{a_2}(x_2, x_3) = 0$, $R_{a_4}(x_3, x_6) = 1$, and also that $R_{\{a_3, a_4\}}(x_3, x_4) = \min(0.95, 0.88) = 0.88$.

### 3 Fuzzy-Rough Attribute Reduction

In this section, we extend the framework for rough set analysis described in Section 2.1 using concepts of fuzzy set theory, to deal with quantitative attributes more appropriately. In order to do so, we introduce a number of increasing,
[0,1]-valued measures to evaluate subsets of \(A\) w.r.t. their ability to maintain discernibility relative to the decision attribute. Once such a measure, say \(M\), is obtained, we can associate a notion of fuzzy decision reduct with it.

**Definition 1. (Fuzzy \(M\)-decision reduct)** Let \(M\) be a monotonic \(P(A) \rightarrow [0,1]\) mapping, \(B \subseteq A\) and \(0 < \alpha \leq 1\). \(B\) is called a fuzzy \(M\)-decision reduct to degree \(\alpha\) if \(M(B) \geq \alpha\) and for all \(B' \subset B\), \(M(B') < \alpha\).

### 3.1 Fuzzy Positive Region

Using fuzzy \(B\)-indiscernibility relations, we can define, for \(y\) in \(U\), \(\text{POS}_B(y) = (\bigcup_{v \in V_B} R_B \downarrow X_k)(y)\). Hence, \(\text{POS}_B\) is a fuzzy set in \(X\), to which \(y\) belongs to the extent that its \(R_B\)-foreset is included into at least one of the decision classes. However, only the decision class \(y\) belongs to needs to be inspected:

**Proposition 1.** For \(y\) in \(X\), \(\text{POS}_B(y) = (R_B \downarrow X_k^*)(y)\) with \(X_k^*(y) = 1\).

**Example 4.** Let us come back to the decision system in Table 1a). Using the same indiscernibility relations as in Ex. 3, and \(I = I_L\), we can calculate the fuzzy positive region for \(B = \{a_4, a_5\}\). For instance, \(\text{POS}_B(x_3) = 0.42\). The complete result is \(\text{POS}_B = \{(x_1, 1), (x_2, 0.65), (x_3, 0.42), (x_4, 0.42), (x_5, 1), (x_6, 1), (x_7, 1)\}\). Compare this with Ex. 1, where \(\text{POS}_B\) was computed for the discretized system: the fuzzy positive region allows gradual membership values, and hence is able to express that e.g. \(x_2\) is a less problematic object than \(x_3\) and \(x_4\). Finally, it can also be verified that, with these parameters, still \(\text{POS}_A = X\).

Once we have fixed the fuzzy positive region, we can define an increasing \([0,1]\)-valued measure to obtain fuzzy decision reducts. We may extend the degree of dependency, as proposed by Jensen and Shen in [3, 4], or, rather than considering an average, it is also possible to focus on the most problematic element. These alternatives are reflected by the following normalized\(^7\) measures:

\[
\gamma_B = \frac{\text{POS}_B}{\text{POS}_A} \quad \delta_B = \frac{\min_{x \in X} \text{POS}_B(x)}{\min_{x \in X} \text{POS}_A(x)}
\]

**Proposition 2.** If \(B_1 \subseteq B_2 \subseteq A\), then \(\gamma_{B_1} \leq \gamma_{B_2}\) and \(\delta_{B_1} \leq \delta_{B_2}\).

**Example 5.** For \(B\) as in Ex. 4, \(\gamma_B = 5.49/7 = 0.78\) and \(\delta_B = 0.42\). Also, \(B\) is a fuzzy \(\gamma\)-decision reduct to degree 0.77, since for \(B' \subset B\), \(\gamma_{B'} < 0.77\).

### 3.2 Fuzzy Discernibility Function

The closeness relation \(R_B\) can be used to redefine the discernibility function as an \([0,1]^m \rightarrow [0,1]\) mapping, such that, for each combination of conditional attributes, a value between 0 and 1 is obtained indicating how well these attributes maintain discernibility, relative to the decision attribute, between all objects.

\(^7\) In this paper, we assume \(\text{POS}_A(x) > 0\) for every \(x \in X\).
A faithful extension of the decision-relative discernibility matrix, in which \( O_{ij} \) (\( i, j \) in \( \{1, \ldots, n\} \)) is a fuzzy set in \( \mathcal{A} \), is obtained by defining, for any attribute \( a \) in \( \mathcal{A} \), \( O_{ij}(a) = 0 \) if \( d(x_i) = d(x_j) \) and \( O_{ij}(a) = 1 - R_a(x_i, x_j) \) otherwise. Accordingly, we can define \( O_{ij}^* \) as the fuzzy set in \( \mathcal{A}^* \), such that \( O_{ij}^*(a^*) = O_{ij}(a) \). Interpreting the connectives in the crisp discernibility function by the minimum and the maximum, we can then extend it to a \( \{0, 1\}^n \rightarrow [0, 1] \) mapping:

\[
f(a_1^*, \ldots, a_m^*) = \min_{1 \leq i < j \leq n} c_{ij}(a_1^*, \ldots, a_m^*) \quad \text{(1)}
\]

\[
c_{ij}(a_1^*, \ldots, a_m^*) = \begin{cases} 
1 - \frac{\max(O_{ij}^*(a_1)^* \ldots O_{ij}^*(a_m)^*)}{1 - R_{A^*}(x_i, x_j)} & \text{if } O_{ij} = \emptyset \\
1 & \text{otherwise} 
\end{cases} \quad \text{(2)}
\]

Referring again to the valuation \( V_B \) corresponding to a subset \( B \) of \( \mathcal{A} \), \( V_B(f(a_1^*, \ldots, a_m^*)) \) is now a value between 0 and 1 that expresses the degree to which, for all object pairs, different values in attributes of \( B \) correspond to different values of \( d \). Rather than taking a minimum operation in (1), which is rather strict, one can also consider the average over all object pairs:

\[
g(a_1^*, \ldots, a_m^*) = \frac{2 \sum_{1 \leq i < j \leq n} c_{ij}(a_1^*, \ldots, a_m^*)}{n(n - 1)} \quad \text{(3)}
\]

The following two propositions express that \( f \) and \( g \) are monotonic, and that they assume the value 1 when all the attributes are considered.

**Proposition 3.** If \( B_1 \subseteq B_2 \subseteq \mathcal{A} \), then \( V_{B_1}(f(a_1^*, \ldots, a_m^*)) \leq V_{B_2}(f(a_1^*, \ldots, a_m^*)) \) and \( V_{B_1}(g(a_1^*, \ldots, a_m^*)) \leq V_{B_2}(g(a_1^*, \ldots, a_m^*)) \).

**Proposition 4.** \( V_{\mathcal{A}}(f(a_1^*, \ldots, a_m^*)) = V_{\mathcal{A}}(g(a_1^*, \ldots, a_m^*)) = 1 \)

**Example 6.** For \( B \) as in Ex. 4, it can be verified that \( V_B(f(a_1^*, \ldots, a_m^*)) = f(0, 0, 0, 1, 1, 0, 0, 0) = 0.42 \), and that \( V_B(g(a_1^*, \ldots, a_m^*)) = g(0, 0, 0, 1, 1, 0, 0, 0) = 0.96 \). Here, it holds e.g. that \( B \) is a fuzzy \( g \)-decision reduct to degree 0.95.

### 4 Experimental Analysis

In this section, we evaluate the performance of our measures in classification, and compare the results to the approaches from \([3, 4]\); the latter have already been shown to outperform other state-of-the-art feature selection techniques in terms of accuracy. In order to select suitable attribute subsets of a decision system \((\mathcal{X}, \mathcal{A} \cup \{d\})\) according to a given measure \( \mathcal{M} \) and threshold \( \alpha \), we used a heuristic algorithm called ReverseReduct, adapted from \([3]\). ReverseReduct starts off with \( B = \mathcal{A} \), and progressively eliminates attributes from \( B \) as long as \( \mathcal{M}(B) \geq \alpha \); at each step, the attribute yielding the smallest decrease in \( \mathcal{M} \) is omitted. By construction, when the algorithm finishes, \( B \) is a fuzzy \( \mathcal{M} \)-reduct of \((\mathcal{X}, \mathcal{A} \cup \{d\})\) to degree \( \alpha \). After feature selection, the decision system is reduced and classified. In our experiments, we used JRip for classification, implemented
in WEKA [9]. The benchmark datasets come from [4], and also include the full version of the Pima dataset used in our running example.

In a first experiment, we fixed $\alpha$ to 0.9 (ReverseReduct looks for a fuzzy $M$-decision reduct to degree 0.9). Table 2 records the results obtained with $\gamma, \delta$, and the min- and average-based variants of the fuzzy discernibility function; for the measures based on the positive region, we worked with $I = I_L$ and $I = I_M$ as implicators. To compute approximate equality, we used $R_B$ as defined in Section 2.3, with $\alpha = 5$ and $\beta = 1.2$ for the average-based approaches, and $\alpha = 15$ and $\beta = 1$ for the min-based approaches$. The one but last column contains accuracy and size for the unreduced dataset, and the last one records the best accuracy obtained in [4], along with the size of the corresponding attribute set.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\gamma_{I_L}$</th>
<th>$\gamma_{I_M}$</th>
<th>$\delta_{I_L}$</th>
<th>$\delta_{I_M}$</th>
<th>$f$</th>
<th>$g$</th>
<th>Unred. [4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pima</td>
<td>77.0 (7)</td>
<td>76.0 (8)</td>
<td>76.8 (6)</td>
<td>76.0 (8)</td>
<td>77.0 (7)</td>
<td>77.6 (2)</td>
<td>76.0 (8)</td>
</tr>
<tr>
<td>Cleveland</td>
<td>54.2 (9)</td>
<td>54.5 (9)</td>
<td>53.2 (8)</td>
<td>53.9 (9)</td>
<td>53.2 (8)</td>
<td>53.9 (2)</td>
<td>52.2 (13)</td>
</tr>
<tr>
<td>Glass</td>
<td>65.4 (6)</td>
<td>67.8 (6)</td>
<td>63.1 (5)</td>
<td>71.5 (9)</td>
<td>65.9 (8)</td>
<td>55.1 (3)</td>
<td>71.5 (9)</td>
</tr>
<tr>
<td>Heart</td>
<td>80.7 (7)</td>
<td>81.9 (8)</td>
<td>73.7 (8)</td>
<td>73.7 (8)</td>
<td>73.7 (8)</td>
<td>75.2 (2)</td>
<td>77.4 (13)</td>
</tr>
<tr>
<td>Olitos</td>
<td>67.5 (8)</td>
<td>65.0 (12)</td>
<td>68.3 (5)</td>
<td>68.3 (5)</td>
<td>68.3 (5)</td>
<td>64.2 (2)</td>
<td>70.8 (25)</td>
</tr>
<tr>
<td>Water 1</td>
<td>82.8 (11)</td>
<td>81.0 (17)</td>
<td>83.1 (8)</td>
<td>83.1 (8)</td>
<td>83.1 (8)</td>
<td>83.3 (1)</td>
<td>83.9 (38)</td>
</tr>
<tr>
<td>Water 3</td>
<td>83.3 (11)</td>
<td>83.6 (17)</td>
<td>82.8 (7)</td>
<td>82.8 (7)</td>
<td>82.8 (7)</td>
<td>85.9 (2)</td>
<td>82.8 (38)</td>
</tr>
<tr>
<td>Wine</td>
<td>92.7 (6)</td>
<td>87.6 (8)</td>
<td>84.3 (5)</td>
<td>84.3 (5)</td>
<td>84.3 (5)</td>
<td>88.2 (2)</td>
<td>92.7 (13)</td>
</tr>
</tbody>
</table>

The results show that, on the whole, our methods are competitive with those from [4]. Moreover, for three of the datasets, strictly better accuracy results can be obtained with at least one of the new approaches. Also, in many cases shorter attribute subsets are produced. In particular, note that $g$ generates very short subsets that have reasonable, and sometimes even excellent, accuracy.

We also investigated the influence of $\alpha$ on the quality of the fuzzy decision reducts; Fig. 1 plots the results obtained for Pima with the four approaches$^9$ as a function of $\alpha$. All approaches reach their optimum for $\alpha < 1$, which clearly endorses using fuzzy decision reducts. For the average-based measures, $\alpha = 0.9$ seems a good compromise value, while the min-based approaches generally require smaller values$^{10}$. The corresponding reduct size decreases gradually for most approaches, except for $g$ which is sensitive to small changes when $\alpha$ is large.

5 Conclusion

We have introduced a framework for fuzzy-rough set based feature selection, built up around the formal notion of a fuzzy reduct. By expressing that an attribute

\begin{itemize}
  \item \textit{Since the min-based approaches are stricter, they require crisper definitions of approximate equality to perform well.}
  \item \textit{For $\gamma$ and $\delta$, $I_L$ was used as implicator.}
  \item \textit{Incidentally, the best overall accuracy, 78.1%, was obtained for $\delta$ with $\alpha \in [0.4, 0.7]$.}
\end{itemize}

\footnotetext[8]{Since the min-based approaches are stricter, they require crisper definitions of approximate equality to perform well.}

\footnotetext[9]{For $\gamma$ and $\delta$, $I_L$ was used as implicator.}

\footnotetext[10]{Incidentally, the best overall accuracy, 78.1%, was obtained for $\delta$ with $\alpha \in [0.4, 0.7]$.}
Fig. 1. Accuracy results and reduct size for varying values of threshold parameter $\alpha$.

subsets should retain the quality of the full feature set to a certain extent only, we are able to generate shorter attribute subsets, without paying a price in accuracy. For the future, we plan to further investigate the role of the various parameters. We also hope to extend the approach to deal with quantitative decisions.

Acknowledgment

Chris Cornelis’ research is funded by the Research Foundation—Flanders.

References