Genomics 101 (2013) 38–48

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Interval-valued analysis for discriminative gene selection and tissue sample classification using microarray data

Yunsong Qi *, Xibei Yang

School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China

ARTICLE INFO

Article history:
Received 10 April 2012
Accepted 10 September 2012
Available online 20 September 2012

Keywords:
Microarray
Gene selection
Classification
Rough sets
Interval-valued decision table

ABSTRACT

An important application of gene expression data is to classify samples in a variety of diagnostic fields. However, high dimensionality and a small number of noisy samples pose significant challenges to existing classification methods. Focused on the problems of overfitting and sensitivity to noise of the dataset in the classification of microarray data, we propose an interval-valued analysis method based on a rough set technique to select discriminative genes and to use these genes to classify tissue samples of microarray data. We first select a small subset of genes based on interval-valued rough set by considering the preference-ordered domains of the gene expression data, and then classify test samples into certain classes with a term of similar degree. Experiments show that the proposed method is able to reach high prediction accuracies with a small number of selected genes and its performance is robust to noise.

1. Introduction

Microarray technology allows simultaneous measurement of the expression levels of thousands of genes within a biological tissue sample. An important application of gene expression is to classify samples according to their gene expression profiles, such as the diagnosis or the classification of different types or subtypes of cancer [1,2]. Different classification methods from statistical and machine learning have been applied to the classification of cancer. However, high dimensionality and a small number of noisy samples pose great challenges to the existing methods. The main approach to this problem has been based on using the existing algorithms to analyze gene expression data. For example, support vector machines (SVM) [3], neural networks (NN) [4], logistic regression (LR) [5] and *k*-nearest neighbor (*k*-NN) [6] have all been utilized. Most of these classifiers involve complex models containing numerous genes. This has limited the interpretability of the classifiers and this lack of interpretability hampers the acceptance of diagnostic tools. Classification models based on numerous genes can also be more difficult to transfer to other assay platforms, which may be more suitable for clinical application. Several authors have suggested that simple models could perform well in areas such as microarray based cancer prediction [7–9,2]. Investigations have indicated that classifiers could be developed to contain a few genes that provide classification accuracy comparable to that achieved by models that are more complex. Moreover, some more complex algorithms based on numerous genes for classification often overfit the data [10–12].

Prior to classification, a variety of gene selection strategies have been used. The aim of gene selection is to select a small subset of genes from a larger pool. Gene selection methods are classified into three types: (1) filter methods, (2) wrapper methods, and (3) embedded methods. Filter methods evaluate a subset of genes by looking at the intrinsic characteristics of data with respect to class labels, while wrapper methods evaluate the goodness of a gene subset by the accuracy of its learning or classification. Embedded methods are generally referred to as algorithms, where gene selection is embedded in the construction of the classifier. In the gene selection process, an optimal feature subset is always relative to a certain criterion. Every criterion measures the discriminating ability of a gene or a subset of genes to distinguish different class labels. To measure the gene–class relevance, different statistical and theoretical measures such as the *t*-test, entropy and mutual information are typically used [13–15], and different metrics including the Euclidean distance and correlation coefficient [16,17] are employed to calculate the gene–gene redundancy. However, as the *t*-test, Euclidean distance, and the correlation coefficient depend on the actual gene expression values of the microarray data, they are very sensitive to noise or outliers within the dataset [18,19].

Rough set theory is a new paradigm to address uncertainty, vagueness, and incompleteness [20]. It has been applied to a number of methods, including the fuzzy rule extraction, reasoning with uncertainty, fuzzy modeling, feature selection and microarray data analysis [21,22,6,23–25,14]. Rough set theory was initially developed for a finite universe of discourse in which the knowledge base is a partition, obtained by any equivalence relationship defined on the universe of discourse. In rough set theory, the data are organized in a table, known as a decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the dataset, a class

* Corresponding author. Tel.: +86 511 84409018; fax: +86 511 84404905.
E-mail address: qys@ujs.edu.cn (Y. Qi).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.09.004

label indicates the class to which each row belongs. The class label is termed a decision attribute; the remaining attributes are termed condition attributes. Rough set theory distinguishes itself from other machine learning and pattern recognition methods through three notions of indiscernibility, approximation, and reduction of attributes (introduced in Sections 2.2 and 3.4). The first defines a relationship stating that two objects are only equivalent under a selection of attributes. The second gives the ability to define an unknown set of boundaries through the analysis of how that set relates to the objects in the universe. The third allows for the reduction of irrelevant information, thus saving valuable resources. These three important concepts give rough set theory an advantage over other classical methods as it does not need any preliminary or additional information about the data: for example, probability in statistics or grade of membership or the value of possibility in fuzzy set theory, all require further information. The characteristics of the microarray data — small sample size and very large dimensionality, create new challenges in obtaining preliminary information.

In practice, discretization is a common preprocess before rough set based mining on gene expression data, which transforms continuous gene expression levels to categorical item sets [26,27]. If a particular gene's expression level is higher than the discretization threshold, the gene is considered as expressed, otherwise it is considered unexpressed. Obviously, a lot of information is lost in the above transformation of the dataset with the noise, which is especially inherent in the microarray data [28]. Previous research has shown that handling uncertainty in such applications by the representation as interval data leads to accurate learning algorithms [29,30].

In this study, we propose an interval-valued analysis method to select discriminative genes, and to use these genes to classify tissue samples of microarray data. We first select a small subset of genes based on interval-valued rough set by considering the preference-ordered domains of the gene expression data, and then classifying a test sample into a certain class with a term of similar degree.

To summarize the process:

- The interval-valued decision table of the microarray is generated. In the decision table, each row corresponds to a class of tissue samples, and each column (condition attribute) corresponds to a gene's expression value over all classes of samples. To generate the decision table, the decision attribute is the average gene expression value of a class, and the condition attribute is the value of the 1st quartile and the 3rd quartile of the gene expression value within a class.
- In the gene selection step, our objective is to determine the reducts that discern between objects belonging to different classes. The reduct, from rough set theory, corresponds to a minimal subset of discriminative genes. The ordered process of this algorithm is described in Section 4.1.
- The tissue sample classification is based on the selected genes. The proposed interval-valued classification method classifies a sample into a class with the maximum similar degrees.

To facilitate our discussion, we first present the basic notions in Section 2. Section 3 presents the discernibility approach to compute reducts from the compared dominance relationships. In Section 4, we describe our gene selection and tissue sample classification method. In Section 5, we apply our approach to the analysis of real microarray data. In this section, we also discuss RNA-sequencing data, the data from next-generation sequencing technologies, and analysis using the proposed method. Finally, Section 6, summarizes our approach and presents our conclusions.

2. Preliminaries

2.1. Microarray dataset

A microarray dataset is a gene expression matrix, in which each column represents a gene and each row represents a sample (or experiment) with a class label. Let $G=\{g_1,\cdots,g_n\}$ be a set of genes and $U=\{s_1,\cdots,s_m\}$ be a set of samples. The corresponding gene expression matrix can be represented as $X=(x_{i,j})_{m\times n}$, where $x_{i,j}$ is the expression level of gene g_j in sample s_i, and usually $n\gg m$. Here m is the number of samples, and n is the number of genes. The matrix X is composed of m row vectors $s_i\in R^n, i=1,2,\cdots,m$. Each vector s_i in the gene expression matrix may be regarded as a point in n-dimensional space, and each of the n columns consists of an m-element expression vector for a single gene.

A microarray dataset can be regarded as a decision table $\mathbb{S}=<U,AT\cup d,V,f>$, where U denotes the set of samples, AT denotes the set of the condition attributes (genes), d denotes the decision attribute (class label), V is the domain of $AT\cup d$, and $x_{i,j}=f(s_i,g_j)$.

2.2. Rough set

An information system is a 4-tuple, where $\mathbb{S}=<U,A,V,f>$. U is a non-empty and finite set of objects, called as universe; A is a non-empty and finite set of attributes, such that $\forall a\in A: U\to V_a$, where V_a is the domain of attribute a; V is regarded as the domain of all attributes such that $V=V_A=\cup_{a\in A}V_a$; $f(x,a)$ is the value that x holds on $a(\forall x\in U, a\in A)$.

A decision table is an information system $\mathbb{S}=<U,AT\cup d,V,f>$, where $d\notin AT$. d is a complete attribute called a decision, and AT is the condition attribute set.

For an information system $\mathbb{S}$, it is possible to describe relationships between objects through their attribute values. With respect to a subset of attributes such that $A\ AT$, an indiscernibility relationship $IND(A)$ [31] may be defined as:

$$IND(A)=\left\{(x,y)\in U^2:\forall a\in A, f(x,a)=f(y,a)\right\}.$$

$IND(A)$ is an equivalence relationship because it is reflexive, symmetrical and transitive. With the relationship $IND(A)$, two objects are considered to be indiscernible if, and only if, they have the same value on each $a\in A$.

Based on the indiscernibility relationship $IND(A)$, it is possible to derive the lower and upper approximations of an arbitrary subset X of U, which are defined as [31]:

$$\underline{A}(X)=\left\{x\in U:[x]_A\subset X\right\}\quad\text{and}\quad\overline{A}(X)=\left\{x\in U:[x]_A\cap X\neq\phi\right\}$$

respectively, where $[x]_A=\{y\in U:(x,y)\in IND(A)\}$ is the A-equivalence class containing x. The pair $\left[\underline{A}(X),\overline{A}(X)\right]$ is referred to as the Pawlak rough set of X with respect to the subset of attributes A.

2.3. Inclusion degree

A partial order on a set X has a binary relationship $\preceq$ with the following properties: $x\preceq x$ (reflexive), $x\preceq y$ and $y\preceq x$ imply $x=y$ (antisymmetric), $x\preceq y$ and $y\preceq z$ imply $x\preceq z$ (transitive).

Definition 1. [32,33] Let $(X,\preceq)$ be a partially ordered set. If for any $x,y\in X$, there is a real number $\mathcal{I}(y/x)$ with the following properties: (1) $0\le\mathcal{I}(y/x)\le 1$; (2) $x\preceq y$ implies $\mathcal{I}(y/x)=1$; (3) $x\preceq y\preceq z$ implies $\mathcal{I}(x/z)\le\mathcal{I}(x/y)$; then $\mathcal{I}$ is called an inclusion degree on X.

For an information system $\mathbb{S}$, U is the universe, the collection of all normal fuzzy subsets of U is denoted by $\mathscr{F}_0(U)$. Let $F_1,F_2\in\mathscr{F}_0(U)$, if $\mu_{F_1}(x)\le\mu_{F_2}(x)$ for all $x\in U$, then $F_1\ F_2$. It is well known that $(\mathscr{F}_0(U),\subseteq)$ is a partially ordered set.

Definition 2. [34] Suppose that $(\mathscr{F}_0(U),\subseteq)$ is a partially ordered set, then $\mathcal{I}$ is an inclusion degree on $\mathscr{F}_0(U)$, if the following conditions hold: (1) $0\le\mathcal{I}(F_2/F_1)\le 1$; (2) $F_1\subseteq F_2\Rightarrow\mathcal{I}(F_2/F_1)=1$; (3) $F_1\ F_2\ F_3\Rightarrow\mathcal{I}(F_1/F_3)\le\mathcal{I}(F_1/F_2)$, where $F_1,F_2,F_3\in\mathscr{F}_0(U)$.

Proposition 1. [34] *If $\mathcal{I}_1, \mathcal{I}_2$ are defined as:*

(1) $\mathcal{I}_1(F_2/F_1) = \min(\mu_{F_1}(x) \cap \mu_{F_2}(x) : x \in U, \mu_{F_1}(x) = 1)$;
(2) $\mathcal{I}_2(F_2/F_1) = \max(\mu_{F_1}(x) \cap \mu_{F_2}(x) : x \in U)$;

where $F_1, F_2 \in \mathscr{F}_0(U)$, then $\mathcal{I}_1, \mathcal{I}_2$ are inclusion degrees on $(\mathscr{F}_0(U), \subseteq)$.

2.4. Fuzzy dominance-based rough set

The Dominance-based Rough Set Approach (DRSA) is a new improvement of Pawlak's rough set model aimed to deal with information systems with preference-ordered domains of the attributes. Greco et al. further generalized DRSA into a fuzzy environment and then proposed the fuzzy dominance-based rough set [35]. In their generalized approach, the target is a fuzzy set instead of the decision tables.

In the decision table $\mathbb{S}$, if $AT = \{a_1, \cdots, a_m\}$ is the set of condition attributes, and d is the decision attribute, then we consider a universe of discourse U and $m+1$ fuzzy sets [35], denoted by $\tilde{a}_1, \cdots, \tilde{a}_m$ and $\tilde{d}$, are defined on U by means of the membership functions

$$\mu_{\tilde{a}_i} : U \to [0,1], i \in \{1, \cdots, m\} \quad \text{and} \quad \mu_{\tilde{d}} : U \to [0,1].$$

$\mu_{\tilde{a}_i}$ and $\mu_{\tilde{d}}$ represent the values of the object x with respect to the condition attribute a_i and decision attribute d, respectively. Suppose that we want to approximate the knowledge contained in decision d using attributes about $\{\tilde{a}_1, \cdots, \tilde{a}_m\}$. Then, given the information on $\tilde{a}_1, \cdots, \tilde{a}_m$, the lower approximation of the fuzzy set $\tilde{d}$ is a fuzzy set $\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)$, whose membership function for each $x \in U$, denoted by $\mu\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right]$, is defined as follows [35]:

$$\mu\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right] = \min_{Z \in D_{AT}^{\uparrow}(x)} \{\mu_{\tilde{d}}(z)\} \tag{1}$$

for each $x \in U$, $D_{AT}^{\uparrow}(x)$ is a non-empty set defined by

$$D_{AT}^{\uparrow}(x) = \left\{y \in U : \mu_{\tilde{a}_i}(y) \geq \mu_{\tilde{a}_i}(x) \text{ for each } a_i \in AT\right\}. \tag{2}$$

$D_{AT}^{\uparrow}(x)$ is the set of objects dominating x in terms of the set of condition attributes.

The lower approximation $\mu\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right]$ can be interpreted as follows: In the universe U, the following implication holds: If $\mu_{\tilde{a}_1}(y) \geq \mu_{\tilde{a}_1}(x)$ and $\mu_{\tilde{a}_2}(y) \geq \mu_{\tilde{a}_2}(x)$ and $\cdots$ and $\mu_{\tilde{a}_m}(y) \geq \mu_{\tilde{a}_m}(x)$, then $\mu_{\tilde{d}}(y) \geq \mu\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right]$.

Similarly, given the information on $\{\tilde{a}_1, \cdots, \tilde{a}_m\}$, the upper approximation of $\tilde{d}$ is a fuzzy set $\overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)$, whose membership function for each $x \in U$ is defined as follows [35]:

$$\mu\left[\overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right] = \max_{Z \in D_{AT}^{\downarrow}(x)} \left\{\mu_{\tilde{d}}(z)\right\} \tag{3}$$

for each $x \in U$, $D_{AT}^{\downarrow}(x)$ is a non-empty set defined by

$$D_{AT}^{\downarrow}(x) = \left\{y \in U : \mu_{\tilde{a}_i}(y) \leq \mu_{\tilde{a}_i}(x) \text{ for each } a_i \in AT\right\}. \tag{4}$$

$D_{AT}^{\downarrow}(x)$ is the set of objects dominated by x in terms of the set of condition attributes.

The upper approximation $\mu\left[\overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right]$ can be interpreted as follows: In the universe U, the following implication holds: If $\mu_{\tilde{a}_1}(y) \leq \mu_{\tilde{a}_1}(x)$ and $\mu_{\tilde{a}_2}(y) \leq \mu_{\tilde{a}_2}(x)$ and $\cdots$ and $\mu_{\tilde{a}_m}(y) \leq \mu_{\tilde{a}_m}(x)$, then $\mu_{\tilde{d}}(y) \leq \mu\left[\overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), x\right]$.

$\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), \overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)\right]$ is referred to as a rough set of the fuzzy set $\tilde{d}$ by using attributes about the $\{\tilde{a}_1, \cdots, \tilde{a}_m\}$. More details about the properties of $\left[\underline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), \overline{App}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)\right]$, can be found in Ref. [35].

3. Fuzzy rough set in interval-valued decision table

3.1. Pairwise comparison in interval-valued decision table

Example 1. To demonstrate the interval-valued decision tables, we considered the data in Table 1, which describes a small training set with interval-valued samples.

Table 1 is a summary of the evaluations of cars. This table details six cars, evaluated by means of five attributes: a_1: Mileage; a_2: Power; a_3: Compression-ratio; a_4: Max-speed; d: Global evaluation. The universe of discourse is $U = \{x_1, \cdots, x_6\}$, $AT = \{a_1, a_2, a_3, a_4\}$ and is the set of condition attributes and d is the decision attribute. The global evaluation indicates that the higher the value of a car holds on decision d, the better the car.

Since the similarity measure [36,37] of two interval-valued sets is one of the important topics in interval-valued theory, in the interval-valued decision table $\mathbb{S}$, let us denote a function $\mu: U \times U \to [0,1]$ such that $\mu_{\tilde{a}_i}(y,x) \in [0,1]$, and it is used to express the degree that an object y is similar to x on a condition attribute $a_i \in AT$. Thus, there are three possibilities that need to be considered:

- $\mu_{\tilde{a}_i}(y,x) = 0$, i.e., y is completely not similar to x on attribute a_i;
- $0 < \mu_{\tilde{a}_i}(y,x) < 1$, i.e., y is partially similar to x with respect to attribute a_i in degree of $\mu_{\tilde{a}_i}(y,x)$;
- $\mu_{\tilde{a}_i}(y,x) = 1$, i.e., y is completely similar to x on attribute a_i.

The similarity degree discussed here is not necessarily symmetrical, that is, $\mu_{\tilde{a}_i}(y,x) = \mu_{\tilde{a}_i}(x,y)$ does not generally hold. It depends on the choice of similarity measurements. By considering the similarity degree of two objects, we consider a universe U and $m+1$ fuzzy set in the interval-valued decision table $\mathbb{S}$, that is, $\tilde{a}_1, \cdots, \tilde{a}_m$ and $\tilde{d}$, defined on U by means of membership functions $\mu_{\tilde{a}_i} : U \to [0,1], i \in \{1, \cdots, m\}$ and $\mu_{\tilde{d}} : U \to [0,1]$.

Example 2. Using Table 1, we will use the method that was proposed by Leung [38] to compute the similarity degree of two interval-valued values. For each $a_i \in AT$, suppose that $\mu_{\tilde{a}_i}(x) = \left[\mu_{\tilde{a}_i}{}^{-}(x), \mu_{\tilde{a}_i}^{+}(x)\right]$ where $\mu_{\tilde{a}_i}{}^{-}(x)$ and $\mu_{\tilde{a}_i}{}^{+}(x)$ represent the lower and upper limitations of the interval-valued data $\mu_{\tilde{a}_i}(x)$ respectively, then for $\forall x, y \in U$, the degree that y is similar to x is defined as:

$$\mu_{\tilde{a}_i}(y,x) = \begin{cases} 0 : \mu_{\tilde{a}_i}(y) \cap \mu_{\tilde{a}_i}(x) = \varnothing \\ \min\left\{\dfrac{\min\left\{\mu_{\tilde{a}_i}{}^{+}(y) - \mu_{\tilde{a}_i}^{-}(x), \mu_{\tilde{a}_i}^{+}(x) - \mu_{\tilde{a}_i}^{-}(y)\right\}}{1\mu_{\tilde{a}_i}{}^{+}(y) - \mu_{\tilde{a}_i}{}^{-}(y)}, 1\right\} : \mu_{\tilde{a}_i}(y) \cap \mu_{\tilde{a}_i}(x) \neq \varnothing. \end{cases}$$

Table 1
Car evaluations.

U	a_1	a_2	a_3	a_4	d
x_1	[0.85,0.95]	[0.84,0.90]	[0.70,0.80]	[0.70,0.85]	0.80
x_2	[0.65,0.80]	[0.70,0.85]	[0.70,0.75]	[0.65,0.80]	0.75
x_3	[0.70,0.80]	[0.70,0.80]	[0.65,0.72]	[0.30,0.50]	0.60
x_4	[0.60,0.75]	[0.75,0.85]	[0.82,0.90]	[0.70,0.80]	0.70
x_5	[0.50,0.69]	[0.50,0.65]	[0.55,0.60]	[0.40,0.60]	0.65
x_6	[0.30,0.61]	[0.60,0.71]	[0.20,0.50]	[0.30,0.50]	0.55

The similarity degree for each pair of objects described in Table 1 is displayed in Table 2. For example, $\mu_{\tilde{a}_i}(x_1,x_2)=\left\{\frac{0}{a_1},\frac{0.17}{a_2},\frac{0.5}{a_3},\frac{0.67}{a_4}\right\}$ means that x_1 is similar to x_2 on a_1 in degree of 0, on a_2 in degree of 0.17, on a_3 in degree of 0.5 and on a_4 in degree of 0.67.

Given that $(y,x),(w,z)\in(U\times U)^2$, the pair of objects (y,x) dominate (w,z) with respect to the set of condition attributes AT if y is similar to x, at least as strong as w is similar to z with respect to each $a_i\in AT$. Precisely, "at least as strong as" means the degree of y being similar to x is equal to or higher than the degree of w similar to z. Conversely, given $(y,x),(w,z)\in U\times U$, the pair of objects (y,x) is to be dominated by (w,z) with respect to the set of condition attributes AT if y is similar to x at most as strong as w is not similar to z with respect to each $a_i\in AT$. Similarly, "at most as strong as" means the degree that y is similar to x is equal or lower than the degree of w is similar to z.

From the discussion above, by comparing the similarity degrees of different pairs of objects, the dominance relationship can be defined as follows:

Definition 3. Given an interval-valued decision table $\mathbb{S}$, the dominance relationship in terms of the set of condition attributes AT can be defined as:

$$\mathscr{R}_{AT}=\left\{((y,x),(w,z))\in(U\times U)^2:\forall a_i\in AT,\mu_{\tilde{a}_i}(y,x)\geq\mu_{\tilde{a}_i}(w,z)\right\}.$$

Unlike the dominance relationship proposed by Greco [39], the dominance relationship presented here is based on the comparison of different pairs of objects. Thus, we call $\mathscr{R}_{AT}$ a pairwise compared dominance relationship.

Though the idea of pairwise comparison has been used to form dominance relationship by Greco in Ref. [39], our pairwise compared dominance relationship is different from Greco's. Greco's dominance relationship is based on the ordinal properties of preferred degrees of pairs of objects while our approach is based on the ordinal properties of similarity degrees of pairs of objects.

Proposition 2. *Given an interval-valued decision table $\mathbb{S}$, if $A\subseteq AT$, then we have $\mathscr{R}_{AT}\subseteq\mathscr{R}_A$.*

Proposition 2 is consistent to the property in the traditional rough set, that is to say, the more attributes we have, the finer binary relationship we obtained.

3.2. Fuzzy rough approximations

Suppose that we want to approximate the knowledge contained in d by using the comparison of pairs of objects, given the information on $\{\tilde{a}_1,\cdots,\tilde{a}_m\}$, the lower approximation of $\tilde{d}$ is a fuzzy set $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$, whose membership function for each $(y,x)\in U\times U$, denoted by $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$, is defined as follows:

$$\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]=\min_{Z\in D^{\uparrow}_{AT}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}\tag{5}$$

where $D^{\uparrow}_{AT}(y,x)$ is a non-empty set defined by:

$$D^{\uparrow}_{AT}(y,x)=\{w\in U:((w,x),(y,x))\in\mathscr{R}_{AT}\}.\tag{6}$$

$D^{\uparrow}_{AT}(y,x)$ is set of objects dominating y in terms of the similarity degrees of x. The formulation of $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$ is concordant with the syntax of the decision rules induced by means of DRSA in a pairwise comparison of objects. Thus, the lower approximation membership $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$ is concordant with the decision rules of the type: *If w is similar to x in degree at least $\mu_{\tilde{a}_1}(y,x)$ on attribute a_1 and $\cdots$ and w is similar to x in degree at least $\mu_{\tilde{a}_m}(y,x)$ on attribute a_m, then $\mu_{\tilde{d}}(w)\geq\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$.*

Given the information on $\{\tilde{a}_1,\cdots,\tilde{a}_m\}$, the upper approximation of $\tilde{d}$ is a fuzzy set $\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$, whose membership function for each $(y,x)\in U\times U$, denoted by $\mu\left[\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$, is defined as follows:

$$\mu\left[\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]=\min_{Z\in D^{\downarrow}_{AT}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}\tag{7}$$

where $D^{\downarrow}_{AT}(y,x)$ is a non-empty set defined by:

$$D^{\downarrow}_{AT}(y,x)=\{w\in U:((y,x),(w,x))\in\mathscr{R}_{AT}\}.\tag{8}$$

$D^{\downarrow}_{AT}(y,x)$ is a set of objects dominating y in terms of the similarity degrees of x. The upper approximation $\mu\left[\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$ is concordant with decision rules of the type: *If w is similar to x in degree at most $\mu_{\tilde{a}_1}(y,x)$ on attribute a_1 and $\cdots$ and w is similar to x in degree at most $\mu_{\tilde{a}_m}(y,x)$ on attribute a_m, then $\mu_{\tilde{d}}(w)\leq\mu\left[\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]$.*

$\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)\right]$ is referred to as a pair of rough set of fuzzy knowledge contained in decision d in terms of the similarity degrees comparison of pairs of objects.

Table 2
Similarity degrees of different cars in Table 1.

$\mu_{\tilde{a}_i}(y,x)$	a_1	a_2	a_3	a_4	a_5	a_6
x_1	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.17}{a_2},\frac{0.5}{a_3},\frac{0.67}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.17}{a_2},\frac{0}{a_3},\frac{0.67}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$
x_2	$\left\{\frac{0}{a_1},\frac{0.07}{a_2},\frac{1}{a_3},\frac{0.67}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.67}{a_1},\frac{0.67}{a_2},\frac{0.4}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0.67}{a_1},\frac{0.67}{a_2},\frac{0}{a_3},\frac{0.67}{a_4}\right\}$	$\left\{\frac{0.27}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.07}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$
x_3	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0.29}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{0.29}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.5}{a_1},\frac{0.5}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0.5}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.1}{a_2},\frac{0}{a_3},\frac{1}{a_4}\right\}$
x_4	$\left\{\frac{0}{a_1},\frac{0.1}{a_2},\frac{0}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.67}{a_1},\frac{1}{a_2},\frac{0}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.33}{a_1},\frac{0.5}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.6}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0.07}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$
x_5	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0.21}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0.5}{a_4}\right\}$	$\left\{\frac{0.47}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.58}{a_1},\frac{0.34}{a_2},\frac{0}{a_3},\frac{0.5}{a_4}\right\}$
x_6	$\left\{\frac{0}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.09}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0}{a_1},\frac{0.09}{a_2},\frac{0}{a_3},\frac{1}{a_4}\right\}$	$\left\{\frac{0.03}{a_1},\frac{0}{a_2},\frac{0}{a_3},\frac{0}{a_4}\right\}$	$\left\{\frac{0.35}{a_1},\frac{0.45}{a_2},\frac{0}{a_3},\frac{0.5}{a_4}\right\}$	$\left\{\frac{1}{a_1},\frac{1}{a_2},\frac{1}{a_3},\frac{1}{a_4}\right\}$

3.3. Some properties

Proposition 3. *Given an interval-valued decision table* $\mathbb{S}$*, if* $A \subseteq AT$*, for each* $(y,x) \in U \times U$*, we have:*

$$D_A^{\uparrow}(y,x) = \cup\left\{D_{AT}^{\uparrow}(w,x) : w \in D_A^{\uparrow}(y,x)\right\} \tag{9}$$

$$D_A^{\downarrow}(y,x) = \cup\left\{D_{AT}^{\downarrow}(w,x) : w \in D_A^{\downarrow}(y,x)\right\}. \tag{10}$$

Proposition 3 holds due to the transitive of the used pairwise comparison dominance relationship. Generally, if the used binary relation is reflexive and transitive, then such properties hold no matter what kind of rough approximation is selected.

Proposition 4. *Given an interval-valued decision table* $\mathbb{S}$*, for each* $(y,x) \in U \times U$*, we have:*

$$\mu\left[\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (y,x)\right] = \mathcal{I}_1\left(D_{AT}^{\uparrow}(y,x)\right) \tag{11}$$

$$\mu\left[\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (y,x)\right] = \mathcal{I}_2\left(D_{AT}^{\downarrow}(y,x)\right). \tag{12}$$

Proposition 4 shows that the proposed rough approximations are equivalent to the inclusion degree presented in Ref. [34]. It presents a relationship between the rough set and included degree.

Proposition 5. *Given an interval-valued decision table* $\mathbb{S}$*, the following properties are satisfied:*

(1) *Let us denote by* $\tilde{d} \times \tilde{d}$ *the Cartesian product on fuzzy set* $\tilde{d}$*, that is, for each* $(y,x) \in U \times U$*,* $U_{\tilde{d}} \times \tilde{d}(y,x) = \min\left(\mu_{\tilde{d}}(y), \mu_{\tilde{d}}(x)\right)$*, then*

$$\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right) \tilde{d} \times \tilde{d} \, \overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right). \tag{13}$$

(2) *For any negation* $N(\cdot)$*, being a strictly decreasing function* $N:[0,1] \to [0,1]$ *such that* $N(1) = 0$ *and* $N(0) = 1$*,*

$$\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}^C\right) = \left(\overline{App}_{\sigma}\left(\tilde{a}^C{}_1, \cdots, \tilde{a}^C{}_m, \tilde{d}\right)\right)^C \tag{14}$$

$$\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}^C\right) = \left(\underline{App}_{\sigma}\left(\tilde{a}^C{}_1, \cdots, \tilde{a}^C{}_m \tilde{d}\right)\right)^C \tag{15}$$

$$\left(\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)\right)^C = \overline{App}_{\sigma}\left(\tilde{a}^C{}_1, \cdots, \tilde{a}^C{}_m, \ \tilde{d}^C\right) \tag{16}$$

$$\left(\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)\right)^C = \underline{App}_{\sigma}\left(\tilde{a}^C{}_1, \cdots, \tilde{a}^C{}_m, \tilde{d}^C\right) \tag{17}$$

where for a given fuzzy set W, the fuzzy set W^C *is the complement of W, defined by* $\mu_{W^C}(x) = N(\mu_W(x))$*.*

(3) *For each set of condition attributes such that* $\left\{\tilde{a}_1, \cdots, \tilde{a}_l, \tilde{d}\right\} \subseteq \left\{\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right\}$*,*

$$\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_l, \tilde{d}\right) \subseteq \underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right) \tag{18}$$

$$\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_l, \tilde{d}\right) \supseteq \overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right) \tag{19}$$

(4) *If* $\mu_{\tilde{a}_i}(y,x) \geq \mu_{\tilde{a}_i}(w,x)$ *for each* $a_i \in AT$*, where* $(y,x),(w,x) \in U \times U$*, then*

$$\mu\left[\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (y,x)\right] \geq \mu\left[\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (w,x)\right] \tag{20}$$

$$\mu\left[\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (y,x)\right] \geq \mu\left[\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right), (w,x)\right]. \tag{21}$$

Results (1), (2), (3) and (4) of Proposition 5 can be regarded as fuzzy counterparts of results, which is well known in the classical rough set theory. More precisely, (1) shows that the fuzzy set $\tilde{d} \times \tilde{d}$ includes its lower approximation and is included in its upper approximation; (2) represents complementarily properties of the proposed fuzzy rough approximations; (3) expresses monotonicity of the proposed fuzzy rough set in terms of the monotonous varieties of condition attributes; and (4) says that the lower and upper approximations are monotonic with respect to the monotonicity of similarity degree of pairs of objects.

3.4. Attribute reduction

3.4.1. Attribute reduction of pairwise compared dominance relationship

Definition 4. Given an interval-valued decision table $\mathbb{S}$, $A \subseteq AT$, then A is referred to as a reduct of AT in terms of a pairwise compared dominance relationship if the following two conditions hold:

(1) $\mathscr{R}_{AT} = \mathscr{R}_A$;
(2) $\mathscr{R}_{AT} \neq \mathcal{R}_B$ for each $B \subset A$.

A reduct of AT is actually a minimal subset of condition attributes which preserves the pairwise compared dominance relation $\mathscr{R}_{AT}$.

$\forall (y,x),(w,z) \in (U \times U)^2$, let us denote by $\mathscr{R}_{AT}((y,x),(w,z)) = \{a_i \in AT : ((y,x),(w,z)) \notin \mathscr{R}_{a_i}\} = \left\{a_i \in AT : \mu_{a_i}(y,x) < \mu_{a_i}(w,z)\right\}$, $\mathscr{D}_{AT}((y,x),(w,z))$ is referred to as the discernibility attribute sets of pairs (y,x) and (w,z), $\mathscr{D}_{AT} = \left\{\mathscr{D}_{AT}((y,x),(w,z)) : (y,x),(w,z) \in (U \times U)^2\right\}$ is referred to as the discernibility matrix of $\mathbb{S}$.

Theorem 1. *Given an interval-valued decision table* $\mathbb{S}$*,* $A \subseteq AT$*, for each* $\mathscr{D}_{AT}((y,x),(w,z)) \neq \varnothing$*, we have* $\mathscr{R}_{AT} = \mathscr{R}_A \Leftrightarrow A \cap \mathscr{D}_{AT}((y,x),(w,z)) \neq \varnothing$*.*

3.4.2. Attribute reduction of fuzzy rough set

The approach of attribute reduction discussed above is with respect to the pairwise compared dominance relationship. In other words, no decision attribute is considered. In the following steps, we will present the practical approach to attribute reductions about fuzzy rough sets.

Definition 5. Given an interval-valued decision table $\mathbb{S}$, $A \subseteq AT = \{a_1, \cdots, a_m\}$,

(1) A is referred to as a reduct of lower approximation $\underline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)$ if and only if for each $(y,x) \in U \times U$,
(a) $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) = \mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right)$,
(b) $\mathcal{I}_1\left(d/D_B^{\uparrow}(y,x)\right) \neq \mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right)$ for $\forall B \subset A$;
(2) A is referred to as a reduct of upper approximation $\overline{App}_{\sigma}\left(\tilde{a}_1, \cdots, \tilde{a}_m, \tilde{d}\right)$ if and only if for each $(y,x) \in U \times U$,
(a) $\mathcal{I}_2\left(d/D_{AT}^{\downarrow}(y,x)\right) = \mathcal{I}_2\left(d/D_A^{\downarrow}(y,x)\right)$,
(b) $\mathcal{I}_2\left(d/D_B^{\downarrow}(y,x)\right) \neq \mathcal{I}_2\left(d/D_{AT}^{\downarrow}(y,x)\right)$ for $\forall B \subset A$;

Since Proposition 4 shows that the lower and upper approximations are equal to two different inclusion degrees, we define the reducts of lower and upper approximations based on the inclusion degrees. Obviously, the reducts of lower and upper approximations are minimal subsets of attributes, which preserve the lower and upper approximation memberships for each pair $(y,x) \in U \times U$.

Theorem 2. (Judgment Theorem.) *Given an interval-valued decision table* $\mathbb{S}$*,* $A \subseteq AT$*, then for each* $(y,x) \in U \times U$*,*

(1) $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) = \mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right) \Leftrightarrow$ *if* $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) > \mathcal{I}_1(d/D_{AT}^{\uparrow}(w,x))$ *then* $D_A^{\uparrow}(w,x) D_A^{\uparrow}(y,x)$;

(2) $\mathcal{I}_2\left(d/D^{\downarrow}_{AT}(y,x)\right) = \mathcal{I}_2\left(d/D^{\downarrow}_{A}(y,x)\right) \Leftrightarrow if\ \mathcal{I}_2(d/\ D^{\downarrow}_{AT}(w,x)) > \mathcal{I}_2(d/\ D^{\downarrow}_{AT}(y,x))\ then\ D^{\downarrow}_{A}(w,x) D^{\downarrow}_{A}(y,x)$;

The above theorem provides approaches to judge whether a subset of condition attributes is persevering the lower (upper) approximate membership value for each $(y,x) \in U \times U$. We can further obtain practical approaches to compute lower and upper approximate reducts in the decision table. We first give the following notions.

Definition 6. Given an interval-valued decision table $\mathbb{S}$, denoted by

$$\mathscr{D}^{L}_{AT} = \left\{\{(w,x),(y,x)\} : \mathcal{I}_1\left(d/D^{\uparrow}_{AT}(y,x)\right) > \mathcal{I}_1\left(d/D^{\uparrow}_{AT}(w,x)\right)\right\};$$
$$\mathscr{D}^{U}_{AT} = \left\{\{(w,x),(y,x)\} : \mathcal{I}_2\left(d/D^{\downarrow}_{AT}(w,x)\right) > \mathcal{I}_2\left(d/D^{\downarrow}_{AT}(y,x)\right)\right\};$$

where,

$$\mathscr{D}^{L}_{AT}\{(w,x),(y,x)\} = \begin{cases} \left\{a_i \in AT : \left\{((w,x),(y,x)) \notin \mathscr{R}_{a_i}\right\} & : \{(w,x),(y,x)\} \in \mathscr{D}^{L}_{AT} \\ AT & : \{(w,x),(y,x)\} \notin \mathscr{D}^{L}_{AT} \end{cases}$$

$$\mathscr{D}^{U}_{AT}\{(w,x),(y,x)\} = \begin{cases} \left\{a_i \in AT : \left\{((y,x),(w,x)) \notin \mathscr{R}_{a_i}\right\} & : \{(w,x),(y,x)\} \in \mathscr{D}^{U}_{AT} \\ AT & : \{(w,x),(y,x)\} \notin \mathscr{D}^{U}_{AT} \end{cases}.$$

$\mathscr{D}^{L}_{AT}\{(w,x),(y,x)\}$ and $\mathscr{D}^{U}_{AT}\{(w,x),(y,x)\}$ are referred to as lower and upper approximate discernibility attribute sets, respectively, $\mathscr{D}^{L}_{AT}$ and $\mathscr{D}^{U}_{AT}$ are referred to as lower and upper approximate discernibility matrices, respectively.

Standard approaches to finding reducts are based on the discernibility matrix. Usually there are many reducts in an information system. The intersection of all reducts is called the CORE. In a discernibility matrix, every entry represents a set of attributes discerning two objects. If an entry consists of only one attribute, then it has a higher significance and the unique attribute must be a member of the CORE. Also, shorter entry is more significant than the longer one. If the times of appearance of an attribute are more than that of the others in the same entry, then this attribute may contribute more classification power to the reduct.

According to the above declaration, we assigned a weight $W(a_i)$ to each attribute a_i. The value of weight $W(a_i)$ for each a_i, which is set to zero initially, is calculated sequentially throughout the whole matrix using the following formula when a new entry C_t is met in the discernibility matrix:

$$W(a_i) = W(a_i) + k_{C_t}|A|/|C_t|, a_i \in C_t \tag{22}$$

where $|A|$ is the cardinality of attribute set A of the information system, $|C_t|$ is the cardinality of the new entry C_t, k_{C_t} is the number of the same entry C_t in the merged matrix.

The heuristic method is based on the fact that if the dataset is consistent, then the intersection of a reduct and an entry in the discernibility matrix cannot be empty; otherwise, the involved two objects would be indiscernible with respect to the reduct according to the definition of the reduct in which the reduct possesses discernible capability for all objects.

Based on the above, we can search reducts with a discernibility matrix.

4. Gene selection and tissue sample classification

As mentioned in Section 2.1, a microarray dataset can be regarded as a decision table. Reducts, from the rough set theory, correspond to a minimal subset of discriminative genes. Our objective is to determine the reducts that can discern between objects belonging to different classes. Tissue sample classification is based on the decision table from the microarray dataset. To generate the decision table, the decision attribute d takes its value from the average of the corresponding condition attribute (gene expression value) of a class, and the condition attribute takes its value from the 1st quartile and the 3rd quartile of the gene expression value within a class.

4.1. Gene selection

Based on the introduced concept, we present an interval-valued reduct (IVR) method to select the genes. The ordered process of this algorithm is:

Algorithm IVR.

1. Initialize the parameters of the algorithm: the designated output reduct $Red = \varnothing$, weight values $W(a_i) = 0, i = 1, \cdots, n$.
2. Compute the lower and upper approximate discernibility matrices D^{L}_{AT} or D^{U}_{AT};
3. Form a new discernibility matrix (D^{L}_{AT} or D^{U}_{AT}), merge all the same entries in the discernibility matrix, record their frequencies and sort all entries in the matrix according to their length (the number of attributes involved in each entry) in descending order; if two entries have the same length, the entry with more frequency is preferred.
4. Use formula (22) to compute the weight value of each attribute in the entry.
5. Calculate the intersection $InSet$ between the reduct Red and an entry C_t: $InSet = Red \cap C_t, t = 1,2,\cdots$, when $InSet = \varnothing$ is obtained go to the next step.
6. An attribute a_i with the maximal weight value is chosen and added to Red.
7. It will go back to the intersection calculation and repeat the process if there is an entry left in the discernibility matrix, otherwise the resulting output Red is the optimal reduct.

In this paper, if a reduct is calculated from $\mathscr{D}^{L}_{AT}(\mathscr{D}^{U}_{AT})$, it is referred to as the lower(upper) approximate reduct $Red^{L}(Red^{U})$.

4.2. Tissue sample classification

Our interval-valued classification (IVC) method for the microarray is based on the union of Red^{L} and Red^{U} from IVR. Suppose that $\mu^{-}_{d}(x_i,c_j)$ and $\mu^{+}_{d}(x_i,c_j)$ represent the lower and upper approximation similar degree between sample x_i and class c_j, $[d_{c_j}{}^{-}, d_{c_j}{}^{+}]$ represents the interval of decision attribute value, let $\mu_d(x_i,c_j) = [\mu^{-}_{d}(x_i,c_j), \mu^{+}_{d}(x_i,c_j)]$, $d_{c_j} = \left[d^{-}_{c_j}, d^{+}_{c_j}\right]$, then classify sample x_i to class c_k:

$$k = \arg\max_{j=1,2,\cdots,5} \left\{\mu\left(\mu_d\left(x_i,c_j\right), d_{c_j}\right)\right\},$$

where,

$$\mu\left(\mu_d\left(x_i,c_j\right), d_{c_j}\right) = min\begin{cases} 0 & : \mu_d\left(x_i,c_j\right) \cap d_{c_j} = \varnothing \\ \left\{\dfrac{min\left\{\mu^{+}_{d}\left(x_i,c_j\right) - d^{-}_{c_j}, d^{+}_{c_j} - \mu^{-}_{d}\left(x_i,c_j\right)\right\}}{\mu^{+}_{d}\left(x_i,c_j\right) - \mu^{-}_{d}\left(x_i,c_j\right)}, 1\right\} & : \mu_d\left(x_i,c_j\right) \cap d_{c_j} \neq \varnothing \end{cases}$$

Table 3
Summary of the ALL/AML dataset.

Dataset	Samples	Genes	Classes
ALL–AML-3	72	7129	3
ALL–AML-4	72	7129	4

5. Experiments

In this section, we present experimental results provided by the IVR and the IVC methods. We evaluate the discriminative performance of our selected gene set on different classifiers. We also compare the performance of our classifying method to a wide range of standard classifiers: Naive Bayes (NB), *k*-NN, Decision Tree (DT) and SVM. We have performed only a limited parameter optimization. For the *k*-NN classifier, three types of *k*-NN classifiers ($k = 1, 3, 5$) were compared to assess the significant performance and we set the parameter k as 3. For the SVM classifier, the linear SVM shows the best performance among the linear, polynomial kernel with exponent 2 and RBF kernel. Since most microarray datasets only have relatively few samples, we chose the leave-one-out cross-validation method for evaluation.

To evaluate our gene selection method IVR, the compared gene sets are selected by the program of Significance Analysis of Microarrays (SAM) [40], a statistical technique for finding significant genes. For the standard classifiers, deciding the number of discriminative genes to select is the first question. A set of experiments are conducted on the dataset by varying the number of genes selected to receive the highest classification accuracy.

Our implementation of the various compared classifiers is based on the Weka environment (http://www.cs.waikato.ac.nz/ml/weka/). The classification accuracy is used as the performance measure. For all the dataset, normalizations are performed so that every observed gene expression has a mean equal to 0 and a variance equal to 1.

5.1. Results on ALL/AML leukemia dataset

To evaluate the performance of our proposed method in practice, we have used a dataset containing gene expression profiles from patients with acute lymphoblastic leukemia (ALL) and acute myeloblastic leukemia (AML) and compared the ALL/AML dataset with the IVR and the IVC methods.

The ALL portion of the dataset is derived from two cell types, B-cells and T-cells, while the AML part is split into two types, bone marrow (BM) samples and peripheral blood (PB). The dataset was studied in [1]. However, due to the bipartition of each component, it can be treated both as a three-class dataset (B-cell, T-cell, and AML) and as a four-class dataset (B-cell, T-cell, AML-BM, and AML-PB). Here the three-class version is referred to as ALL–AML-3 and the four-class version as ALL–AML-4. Table 3 provides a summary of the ALL/AML dataset.

The decision table for ALL–AML-4 dataset is shown in Table 4. The lower and upper approximation reducts are shown in Table 5. Based on these reducts, we can classify a test sample into a certain class. Classification accuracy on ALL–AML datasets are shown in Tables 6 and 7. The experimental results show that our proposed methods have a dominating performance. The reason may be that microarrays contain various technical noises. In Subsection 5.3, we design a set of simulations to examine the robustness of our approach to noise.

Table 5
Genes selected in a reduct set of ALL–AML datasets.

Lower approximation reduct		Upper approximation reduct	
ALL–AML-3	ALL–AML-4	ALL–AML-3	ALL–AML-4
CD33	CD33	CD33	CD33
IL8	IL8	IL8	IL8
ZYX	ZYX	ZYX	ZYX
ASAH1	ASAH1	PSME1	PRSS3
PRSS3	PSME1	PRAME	PRAME
	MAL		MAL

5.2. Results on NCI-60 dataset

To further test the performance of the proposed method, we applied our algorithm to the National Cancer Institute's anti-cancer drug-screen data (NCI-60) of Ross et al. [41] consisting of 61 samples from human cancer cell lines. The NCI-60 dataset spans nine classes and gene expression levels were measured for 10,000 genes. The prediction accuracy of 66.66% is reported in reference [41] using one-versus-the rest SVM with 150 selected genes. To test our algorithm on an external dataset (independent set), 43 samples are used for the training dataset while 18 samples as testing dataset. Based on 150 genes selected by SAM and 12 genes selected by IVR, we report the classification accuracy of all the compared algorithms with Fig. 1. Consistent with the results on leukemia dataset, in this experiment, our proposed method also achieved the highest classification accuracy.

5.3. Simulations for noise sensitivity analysis

High noise is always a challenge in existing gene expression data analysis algorithms. We studied the noise sensitivity property of our method and the standard classifiers on simulated datasets generated by adding artificial noise to real gene expression datasets.

In the experiments, the simulated noise has Gaussian distribution, $N(0, w\delta_j^2)$, where δ_j^2 is the jth gene expression level's variance, and w is the weight of the simulated noise. After adding the generated Gaussian noise, expression level x_j is shifted to $x_j + N(0, w\delta_j^2)$.

Fig. 2 shows the comparison of IVC classifier and standard classifiers against synthetic noise on the ALL–AML datasets. These results show that our IVC method is the most robust method. The IVC method achieves reasonable classification accuracy, even when the data contain a lot of noise. This is because, in our rough set based method, gene expression data are treated as interval-valued data by considering the preference-ordered domains and compared with a term of similar degree, which already takes the noise into consideration. The accuracy of Naive Bayes, *k*-NN, Decision Tree and SVM decreases dramatically with the increase in noise, meaning that all of the methods are sensitive to noise.

5.4. Discussion

From the results shown in Tables 6 and 7, we observe the following:

(1) The accuracy of the classification is highly dependent on the choice of the classification method. For instance, with the gene set selected by the IVR method, the IVC classifier has an

Table 4
Decision table of ALL–AML-4 dataset.

U	G_1	G_2	···	G_{7129}	d	$d_{interval}$
c_1	[0.0332, 0.1220]	[0.0622, 0.1998]	···	[0.0314, 0.1183]	0.7223	[0.0655, 1.0000]
c_2	[0.3246, 0.6484]	[0.0464, 0.1436]	···	[0.0671, 0.1227]	0.6302	[0.0512, 0.9165]
c_3	[0.0783, 0.1712]	[0.0715, 0.1524]	···	[0.0908, 0.1789]	0.5278	[0.0301, 0.7533]
c_4	[0.0734, 0.1548]	[0.0524, 0.1243]	···	[0.1050, 0.1733]	0.4309	[0.0186, 0.5296]

Table 6
The classification accuracy with gene set selected by SAM method.

Classifier	Number of selected genes		Classification accuracy	
	ALL–AML-4	ALL–AML-3	ALL–AML-4	ALL–AML-3
Naive Bayes	80	85	78.33%	82.44%
k-NN	40	50	87.81%	89.73%
Decision Tree	60	55	65.35%	72.78%
SVM	85	100	89.04%	91.41%
IVC	30	30	88.25%	92.17%

accuracy of 97.47% on the ALL–AML-4 dataset, while the accuracy by Decision Tree is 70.92%. To better understand IVC's performance, we rank the average classification accuracy of all the algorithms throughout Tables 6 and 7 as follows:

$$\mathrm{IVC}(94.05\%) > \mathrm{SVM}(92.72\%) > k-\mathrm{NN}(91.03\%) > \mathrm{NB}(86.14\%) > \mathrm{DT}(72.09\%).$$

It is observed that our IVC method gets the best performance. The experimental results on NCI-60 dataset (Fig. 1) are also consistent with this.

(2) The accuracy of the classification is also highly dependent on the selected gene set. When the genes are selected by the SAM method, the SVM classifier has an accuracy of 91.41% on the ALL–AML-3 dataset. On the same dataset with the gene set selected by IVR method, the accuracy of SVM is 97.27%. This significant differential accuracy between the gene selection method of SAM and IVR also occurs with the other classifiers.

(3) Although the number of selected genes with the IVR method is much less than by SAM, the accuracies of all the methods are improved on the two ALL/AML datasets. The classification accuracy in Table 6 is generally smaller for all algorithms compared to the results in Table 7. Yet the algorithms in Table 7 use much less genes for the classification than in Table 6. The main idea of rough set theory is to reduce the redundancy of data by attribute reduction, while preserving the ability of classification. Compared with other approaches to attribute reduction, rough set theory can be used to discover data dependencies and reduce the number of attributes contained in a data set by purely structural methods. The reduced set of attribute preserves the underlying semantics of the features. In theory, a reduct can represent all the discriminative attributes contained in a data. In practice, it is observed that, when the number of selected attributes with purely structural methods is greater than a certain degree, the variation of the classifying accuracy is small [42]. The higher the degree of overlap between the reduct and the selected attributes indicates that the higher the percentage of the selected attributes contained the discriminative attributes in a data.

(4) When the sample sizes decrease (ALL–AML-3 dataset vs. ALL–AML-4 dataset), the performance of the IVC classifier is even more outstanding. For example, compared to the SVM classifier, with the gene set selected by the IVR method, the accuracy of IVC classifier is improved by 4.32% on the ALL–AML-4 dataset, while it is only improved by 1.05% on the ALL–AML-3 dataset. Note that, the number of samples in each class of the ALL–AML-4 dataset is lower than those in the ALL–AML-3 dataset.

Table 7
The classification accuracy with gene set selected by IVR method.

Classifier	Number of selected genes		Classification accuracy	
	ALL–AML-4	ALL–AML-3	ALL–AML-4	ALL–AML-3
Naive Bayes	8	7	89.87%	93.93%
k-NN	8	7	91.22%	95.36%
Decision Tree	8	7	70.92%	79.29%
SVM	8	7	93.15%	97.27%
IVC	8	7	97.47%	98.32%

1. The selected gene set is the union of lower and upper approximation reduct.
2. Note that some genes appear in both lower and upper approximation reduct.

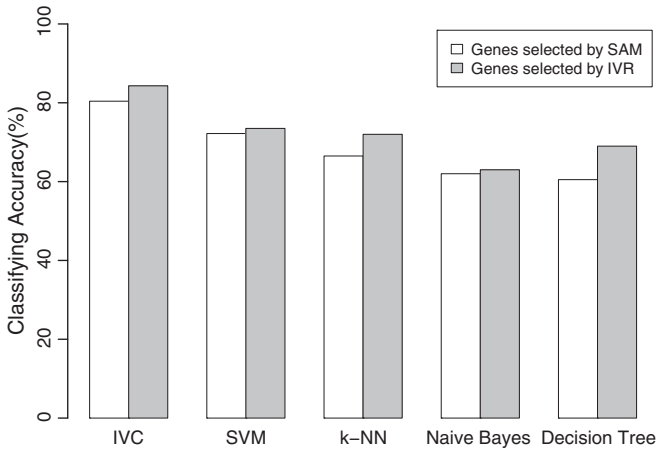

Fig. 1. The comparison of classifying accuracy on NCI-60 dataset with 150 genes selected by SAM and 12 genes selected by IVR.

From the results shown in Tables 6 and 7, we can see that the IVC is shown to be the best method for tissue classification based on gene expression. It achieves better performance than any of the other classifiers. It is conceivable that feature selection raises the accuracy since it can reduce the number of insignificant dimensions, thereby overcoming the curse of dimensionality. This appears to be the case for *k*-NN, Decision Tree and SVM classifier methods. The accuracy of *k*-NN, Decision Tree and SVM is improved on the two datasets with genes selected by the IVR method. The accuracy of the Naive Bayes is also dramatically improved on the experimental datasets. Also, remarkably, with the aid of feature selection, IVC achieves the 98.32% accuracy on the ALL–AML-3 dataset and 97.47% accuracy on the ALL–AML-4 dataset. For the SVM method, it is possible to achieve very high accuracy on most of the microarray datasets [42]. However, the best performance on the experimental datasets does not outperform the IVC method. These two datasets have smaller sample sizes than the other datasets, so one may conclude that multiclass classification based on gene expression can be effectively solved when the sample size is large. Although it has been widely used in text categorization, Naive Bayes reported that it did not appear to perform very well for tissue classification based on gene expression using the standard feature selection method [42]. This is not very surprising, since Naive Bayes is based on the assumption that the features are conditionally independent given the class label, which may not be the case for gene expression data because of co-regulation. For the IVR method, genes were selected as a reduct without redundancy and results in independence between selected feature genes.

The selected gene set for leukemia classification, including PSME1, CD33, IL8, PRAME, ASAH1, PRSS3, MAL and ZYX, that achieve 97.47% and 98.32% classifying accuracy is experimentally proved to be correlated to leukemia of ALL or AML. Specifically, Gene PSME1 inhibits programmed cell death and promotes survival of C-cell chronic lymphocytic leukemia (B-CLL) cells in culture [43]. By delaying apoptosis, PSME1 may extend the life span of the malignant cells. The human differentiation antigen CD33 is a marker of leukemia (ALL) and also a member of the sialic acid-binding immunoglobulin-like lectin (Siglec) family of inhibitory receptors. It is also a therapeutic target for AML [44]. Gene IL8 was found to be up regulated in human

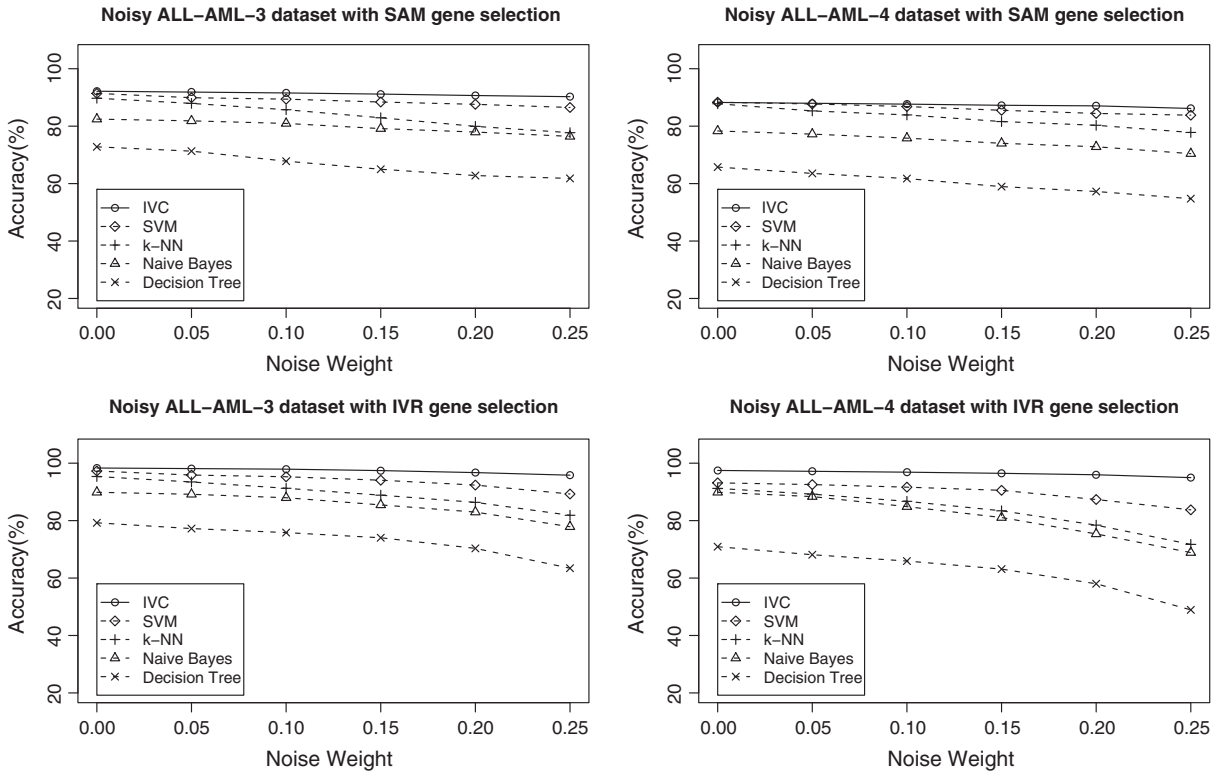

Fig. 2. The comparison of classifying accuracy on the noisy ALL/AML dataset with SAM and IVR gene selection.

T-cell acute lymphoblastic leukemia [45]. Overexpression of the preferentially expressed antigen PRAME of melanoma was found in 62% ($n=31$) of 50 patients with higher rates of overall and disease-free survival from AML. PRAME expression at diagnosis was negatively correlated to the white blood cell count ($P<0.05$), which was significantly higher in patients with t(8;21) and corresponded with those at relapse ($P<0.001$), suggesting that its expression is an indicator of favorable prognosis, and could be a useful tool for monitoring minimal residual disease in childhood AML [46]. The gene ASAH1 is correlated to the survival of cytotoxic lymphocytes [47]. The serine protease family member PRSS3 is a putative tumor suppressor gene due to its loss of expression in acute lymphoblastic leukemia [48] and other tumors. It may be functionally important as a serine protease in tumor development. The MAL gene is related to T-cell ALLs [49]. ZYX is a gene correlated to leukemia of ALL [50]. It is also localized at focal contacts in adherent erythroleukemia cells [51].

From the experimental results above, we conclude that our proposed approach is superior to other methods. This may be due to the following advantages: interval-valued analysis, minimum redundancy of the selected gene subset, and simple classifiers. First, we used interval-valued gene expressions instead of point-valued gene expressions. A major challenge in DNA microarray analysis is to eliminate the effects of noise. Our method is non-parametric and has an advantage over other methods since no assumption about the nature of the noise is required. Second, the selected gene subset is the union of lower and upper approximation reducts which is actually a minimum redundancy–maximum relevance condition attribute subset. Such a feature set covers the data domain better and improves the performance of various classifiers [52]. Third, the proposed IVC method classifies a sample to a class with the maximum similar degrees. The IVC method does not need parameter tuning. The small sample size and high dimensionality of the microarray data constrain the possibility of properly validating the chosen classification model. If a complex model was required to tune many parameters, a large computational effort would be required, compounded with a high risk of overfitting.

Although the proposed method was originally designed for microarray data analysis, it can be applied to the data from the next-generation sequencing technologies. Recently, high-throughput RNA sequencing (RNA-seq) has emerged as a powerful new technology for transcriptome analysis [53]. By mapping millions of RNA-seq reads to individual gene transcripts, it is possible to estimate the overall mRNA abundance and detect DEGs [54]. Currently, most of the methodologies proposed so far rely on parametric assumptions and use Poisson or negative binomial distributions to model feature counts [55,56], following the rationale of the sampling procedure in RNA-seq analysis. However, the subsequent confirmation of distribution assumptions is important as they might not always hold true [57]. Moreover, usually there are very few replicates making the estimation of model parameters difficult. Additionally, parametric approaches tend to be problematic for assessing differential expression in low count features [57]. Based on the rough set theory, the IVR method takes into account the discrete nature of gene expression quantification in RNA-seq. In the context of RNA-seq analysis, we can derive an estimate of gene expression interval-valued level with the number of RNA-seq reads that uniquely mapped to its constitutive exons, i.e. exons are always incorporated into the final transcripts during splicing. Thus, we predict that the IVR method is more robust than the model-based approaches in RNA-seq data analysis.

6. Summary and conclusions

In this paper, we propose a combination method of interval-valued analysis based gene selection and tissue classification of microarray data. We have demonstrated that this approach reduces the number of genes selected and increases the classification accuracy rate. We performed various studies to compare the performance between different types of classifiers including Naive Bayes, *k*-NN, Decision Tree and SVM. The performances of all the methods were improved by the IVR gene selection method. The IVR gene selection method can further improve the performance of the IVC classification method to achieve

the maximum accuracy of 98.32%. Overall, the best classification accuracy rate is achieved by our proposed method with a relatively small gene subset. As our experiments indicate, interval-valued analysis can reduce the infection of the noise data while retaining the information hidden in the data, and the rough set based technique can remove redundant genes while keeping the required genes to as low a number as possible, and thus improve the classifier performances. Though the experimental datasets are related to gene expression data, the method can be applied to other large datasets that require feature selection.

Acknowledgments

This work is supported in part by the Natural Science Foundation of China (No. 61100116), Natural Science Foundation of Jiangsu Province of China (No. BK2011492), Natural Science Foundation of Jiangsu Higher Education Institutions of China (Nos. 10KJB520006, 11KJB520004) and Postdoctoral Science Foundation of China (No. 20100481149).

Appendix A. Proofs of propositions

Proof of Proposition 2. By the Definition of pairwise compared dominance relation, it is trivial to prove this property.

Proof of Proposition 3. For $\forall w\in D_A^{\uparrow}(y,x)$, there must be $\mu_{\tilde{a}_i}(w,x)=\mu_{\tilde{a}_i}(w,x)$ for each $a_i\in AT$, thus $w\in D_{AT}^{\uparrow}(w,x)$, i.e., $D_A^{\uparrow}(y,x)\cup\{D_{AT}^{\uparrow}(w,x):w\in D_A^{\uparrow}(y,x)\}$. It must be proved that $D_A^{\uparrow}(y,x)\supseteq\cup\{D_{AT}^{\uparrow}(w,x):w\in D_A^{\uparrow}(y,x)\}$.

Let us firstly prove that $w\in D_A^{\uparrow}(y,x)\Rightarrow D_A^{\uparrow}(w,x)\subseteq D_A^{\uparrow}(y,x)$. For $\forall z\in D_A^{\uparrow}(w,x)$, we have $\mu_{\tilde{a}_i}(z,x)\geq\mu_{\tilde{a}_i}(w,x)$ for each $a_i\in A$. By $w\in D_A^{\uparrow}(y,x)$ we obtain that $\mu_{\tilde{a}_i}(w,x)\geq\mu_{\tilde{a}_i}(y,x)$ for each $a_i\in A$. It follows that $\mu_{\tilde{a}_i}(z,x)\geq\mu_{\tilde{a}_i}(y,x)$ for each $a_i\in A$, i.e., $z\in D_A^{\uparrow}(y,x)$.

Therefore, for $\forall D_{AT}^{\uparrow}(w,x)\in\cup\{D_{AT}^{\uparrow}(w,x):w\in D_A^{\uparrow}(y,x)$, since $w\in D_A^{\uparrow}(y,x)$, we have $D_{AT}^{\uparrow}(w,x)\subseteq\{D_A^{\uparrow}(w,x)\subseteq D_A^{\uparrow}(,x)\}$ because $A\subseteq AT$, from which we obtain that $\cup\{D_{AT}^{\uparrow}(w,x):w\in D_A^{\uparrow}(y,x)\}\subseteq D_A^{\uparrow}(y,x)$.

From the discussion above, we have $D_A^{\uparrow}(y,x)=\cup\{D_{AT}^{\uparrow}(w,x):w\in D_A^{\uparrow}(y,x)\}$.

Similarly, it is not difficult to prove that $D_A^{\downarrow}(y,x)=\cup\{D_{AT}^{\downarrow}(w,x):w\in D_A^{\downarrow}(y,x)\}$.

Proof of Proposition 4. By definitions of $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$, $\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$, and Proposition 1, it is trivial to prove this property.

Proof of Proposition 5.

1. Obviously, for each $(y,x)\in U\times U$, we have $y\in D_{AT}^{\uparrow}(y,x)$. Since $\mu\left[\underline{App}_\sigma(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}),(y,x)\right]=\min_{z\in D_{AT}^{\uparrow}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}$, we can see that $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]\leq\mu_{\tilde{d}}(y)$.
   On the other hand, since $\mu_{\tilde{a}_i}(x,x)=1$ for each $a_i\in AT$, we have $\mu_{\tilde{a}_i}(x,x)\geq\mu_{\tilde{a}_i}(y,x)\Rightarrow x\in D_{AT}^{\uparrow}(y,x)$. It follows that $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]\leq\mu_{\tilde{d}}(x)$, from which we can conclude that $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]\leq\min\left\{\mu_{\tilde{d}}(y),\mu_{\tilde{d}}(x)\right\}$, i.e., $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)\subseteq\tilde{d}\times\tilde{d}$.
   Similarly, it is not difficult to prove that $\tilde{d}\times\tilde{d}\subseteq\overline{App_\sigma}\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$.
2. For each $(y,x)\in U\times U$, by formula (5), we have $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}^C\right)=\min_{z\in D_{AT}^{\uparrow}(y,x)}\left\{N\left(\mu_{\tilde{d}}(z)\right)\right\}=N\left(\max_{z\in D_{AT}^{\uparrow}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}\right)$.
   Let us associate the negation of the membership functions $\mu_{\tilde{a}_i}(y,x)$, i.e., $N\left(\mu_{\tilde{d}}(y,x)\right)$, we can obtain $D_{AT}^{\uparrow}(y,x)=\{w\in U:((w,x),(y,x))\in\mathcal{R}_{AT}\}=\left\{w\in U:\mu_{\tilde{a}_i}(w,x)\geq\mu_{\tilde{a}_i}(y,x)\right.$ for each $\left.a_i\in AT\right\}=\left\{w\in U:N\left(\mu_{\tilde{a}_i}(w,x)\leq\mu_{\tilde{a}_i}(y,x)\right)\right.$ for each $\left.a_i\in AT\right\}=D_{AT}^{\downarrow C}(y,x)$, where "$C$" in $D_{AT}^{\downarrow C}(y,x)$ denotes that we are considering the negation of the membership functions $\mu_{\tilde{a}_i}(y,x)$.
   From the discussion above, we have

$$N\left(\max_{z\in D_{AT}^{\uparrow}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}\right)=N\left(\max_{z\in D_{AT}^{\downarrow C}(y,x)}\left\{\mu_{\tilde{d}}(z)\right\}\right)=N\left(\overline{App_\sigma}\left(\tilde{a}^C{}_1,\cdots,\tilde{a}^C{}_m,\tilde{d}\right)\right),$$

   from which we can conclude that $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}^C\right)=\left(\overline{App_\sigma}\left(\tilde{a}^C{}_1,\cdots,\tilde{a}^C{}_m,\tilde{d}\right)\right)^C$.
   Similarly, it is not difficult to prove formulas (15), (16) and (17).
3. Suppose that $A=\{\tilde{a}_1,\cdots,\tilde{a}_l\}$, $AT=\{\tilde{a}_1,\cdots,\tilde{a}_m\}$, $A\ AT$, then for each $(y,x)\in U\times U$, we have $D_A^{\uparrow}(y,x)\supseteq D_{AT}^{\uparrow}(y,x)$ and $D_A^{\downarrow}(y,x)\supseteq D_{AT}^{\downarrow}(y,x)$ by Proposition 2. Thus, $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]=\min_{z\in D_{AT}^{\uparrow}(y,x)}\left\{\tilde{d}(z)\right\}\geq\min_{z\in D_A^{\uparrow}(y,x)}\left\{\tilde{d}(z)\right\}=\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_l,\tilde{d}\right),(y,x)\right]$, from which we can conclude that $\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_l,\tilde{d}\right)\subseteq\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right)$.
   Similarly, it is not difficult to prove formula (19).
4. By condition we have $D_{AT}^{\uparrow}(y,x)\ D_{AT}^{\uparrow}(w,x)$ and $D_{AT}^{\downarrow}(y,x)\ D_{AT}^{\downarrow}(w,x)$. Thus, $\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]=\min_{z\in D_{AT}^{\uparrow}(y,x)}\left\{\tilde{d}(z)\right\}\geq\min_{z\in D_{AT}^{\uparrow}(w,x)}\left\{\tilde{d}(z)\right\}=\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(w,x)\right]$.
   Similarly, it is not difficult to prove formula (21).

Appendix B. Proofs of theorems

Proof of Theorem 1. "⇒": Since $\mathcal{R}_{AT}=\mathcal{R}_A$, for each $(y,x)\in U\times U$, if $\mathcal{D}_{AT}((y,x),(w,z))\neq\emptyset$, then there must be $((y,x),(w,z))\notin\mathcal{R}_{AT}\Leftrightarrow((y,x),(w,z))\notin\mathcal{R}_A$. It follows that $a_i\in\mathcal{D}_A((y,x),(w,z))\subseteq\mathcal{D}_{AT}((y,x),(w,z))$ such that $\mu_{a_i}(y,x)<\mu_{a_i}(w,z)$, i.e., $A\cap D_{AT}((y,x),(w,z))\neq\emptyset$.

"⇐": Since $A\subseteq AT$, there must be $((y,x),(w,z))\in\mathcal{R}_{AT}\Rightarrow((y,x),(w,z))\in\mathcal{R}_A$. It must be proved that $((y,x),(w,z))\in\mathcal{R}_A\Rightarrow((y,x),(w,z))\in\mathcal{R}_{AT}$. For each $\mathcal{D}_{AT}((y,x),(w,z))\neq\emptyset$, we have $((y,x),(w,z))\notin\mathcal{R}_{AT}$, since $A\cap\mathcal{D}_{AT}((y,x),(w,z))\neq\emptyset$, then there must be $a_i\in A$ such that $\mu_{a_i}(y,x)<\mu_{a_i}(w,z)$, i.e., $((y,x),(w,z))\notin\mathcal{R}_A$.

Proof of Theorem 2. Suppose that $A=\{a_1,\cdots,a_l\}$ $\{a_1,\cdots,a_m\}=AT$.

(1) "⇒": If $D_A^{\uparrow}(w,x)\ D_{AT}^{\uparrow}(y,x)$, by formula (20) we have

$$\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_l,\tilde{d}\right),(y,x)\right]\leq\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(w,x)\right].$$

By Proposition 4, $\mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right)\leq\mathcal{I}_1\left(d/D_A^{\uparrow}(w,x)\right)$ holds. By assumption, we have $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right)=\mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right)$, $\mathcal{I}_1(d/D_{AT}^{\uparrow}(w,x))=\mathcal{I}_1\left(d/D_A^{\uparrow}(w,x)\right)$, it follows that $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right)\leq\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(w,x)\right)$.

"⇐": Obviously, for each $(y,x)\in U\times U$, we have $A\subseteq AT\Rightarrow D_{AT}^{\uparrow}(y,x)\ D_A^{\uparrow}(y,x)\Rightarrow\ \mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_m,\tilde{d}\right),(y,x)\right]\geq\mu\left[\underline{App}_\sigma\left(\tilde{a}_1,\cdots,\tilde{a}_l,\tilde{d}\right),(y,x)\right]=\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right)\geq\mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right)$.

Consequently, it must be proved that $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right)\leq\mathcal{I}_1\left(d/D_A^{\uparrow}(y,x)\right)$.

Since $w\in D_A^{\uparrow}(w,x)$, $D_A^{\uparrow}(w,x)\subseteq D_A^{\uparrow}(y,x)\Rightarrow w\in D_A^{\uparrow}(y,x)$ holds.

Conversely, by proof of Proposition 2 we have $w\in D_A^{\uparrow}(y,x)\Rightarrow D_A^{\uparrow}(w,x)\ D_A^{\uparrow}(y,x)$, which implies that $w\in D_A^{\uparrow}(y,x)\Leftrightarrow D_A^{\uparrow}(w,x)\subseteq D_A^{\uparrow}(y,x)$.

By Proposition 2, we have

$$D_A^{\uparrow}(y,x) = \cup\left\{D_{AT}^{\uparrow}(w,x) : w \in D_A^{\uparrow}(y,x)\right\} = \cup\left\{D_{AT}^{\uparrow}(w,x) : D_A^{\uparrow} \subseteq D_A^{\uparrow}(y,x)\right\},$$

By assumption we have the following:

$$D_A^{\uparrow}(w,x)\ D_A^{\uparrow}(y,x) \Rightarrow \mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) \le \mathcal{I}_1\left(d/D_{AT}^{\uparrow}(w,x)\right).$$

Thus, for $\forall D_{AT}^{\uparrow}(w,x) \in D_A^{\uparrow}(y,x)$, we have $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) \le \mathcal{I}_1(d/D_{AT}^{\uparrow}(w,x))$, it follows that $\mathcal{I}_1\left(d/D_{AT}^{\uparrow}(y,x)\right) \le \mathcal{I}_1\ \left(d/D_A^{\uparrow}(y,x)\right)$.

(2) The proof of (2) is similar to the proof of (1).

References

[1] T.R. Golub, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[2] X. Wang, R. Simon, Microarray-based cancer prediction using single genes, BMC Bioinformatics 12 (1) (2011) 391.
[3] R. Debnath, T. Kurita, An evolutionary approach for gene selection and classification of microarray data based on SVM error-bound theories, Biosystems 100 (1) (2010) 39–46.
[4] N. Eiamkanitchat, N. Theera-Umpon, S. Auephanwiriyakul, Colon tumor microarray classification using neural network with feature selection and rule-based classification, in: Z. Zeng, J. Wang (Eds.), Advances in Neural Network Research and Applications, Vol. 67 of Lecture Notes in Electrical Engineering, Springer, Berlin, 2010, pp. 363–372.
[5] C. Bielza, V. Robles, P. Larrańaga, Regularized logistic regression without a penalty term: an application to cancer classification with microarray data, Expert Syst. Appl. 38 (5) (2011) 5110–5118.
[6] W. Wu, E. Xing, C. Myers, I. Mian, M. Bissell, Evaluation of normalization methods for cDNA microarray data by *k*-NN classification, BMC Bioinformatics 6 (1) (2005) 191.
[7] D. Geman, C. d'Avignon, D. Naiman, R. Winslow, Classifying gene expression profiles from pairwise mRNA comparisons, Stat. Appl. Genet. Mol. Biol. 3 (1) (2004) 19.
[8] S. Baker, Simple and flexible classification of gene expression microarrays via Swirls and Ripples, BMC Bioinformatics 11 (1) (2010) 452.
[9] X. Wang, O. Gotoh, Accurate molecular classification of cancer using simple rules, BMC Med. Genomics 2 (1) (2009) 64.
[10] A. Statnikov, C. Aliferis, I. Tsamardinos, D. Hardin, S. Levy, A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis, Bioinformatics 21 (5) (2005) 631–643.
[11] E. Pranckeviciene, R. Somorjai, On classification models of gene expression microarrays: the simpler the better, in: Neural Networks, 2006. IJCNN'06. International Joint Conference on, IEEE, 2006, pp. 3572–3579.
[12] M. Schumacher, H. Binder, T. Gerds, Assessment of survival prediction models based on microarray data, Bioinformatics 23 (14) (2007) 1768–1774.
[13] S. Zhang, J. Cao, A close examination of double filtering with fold change and *t*-test in microarray analysis, BMC Bioinformatics 10 (1) (2009) 402.
[14] Y. Wang, H. Yan, Entropy based sub-dimensional evaluation and selection method for DNA microarray data classification, Bioinformation 3 (3) (2008) 124.
[15] D. Ucar, F. Altiparmak, H. Ferhatosmanoglu, S. Parthasarathy, Mutual information based extrinsic similarity for microarray analysis, in: S. Rajasekaran (Ed.), Bioinformatics and Computational Biology, Vol. 5462 of Lecture Notes in Computer Science, Springer, Berlin, 2009, pp. 424–436.
[16] S. Dixon, R. Brereton, Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure, Chemom. Intell. Lab. Syst. 95 (1) (2009) 1–17.
[17] C. Liao, C. Lin, J. Liu, Noninferiority tests based on concordance correlation coefficient for assessment of the agreement for gene expression data from microarray experiments, J. Biopharm. Stat. 17 (2) (2007) 309–327.
[18] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[19] J. Laurie, S. Heyer, Y. Shibu, Exploring expression data: identification and analysis of coexpressed genes, Genome Res. 10 (1) (2000) 1106–1115.
[20] Z. Pawlak, Rough set theory and its applications to data analysis, Cybernet. Syst. 29 (1998) 661–688.
[21] S. Pal, S. Mitra, P. Mitra, Rough-fuzzy mlp: modular evolution, rule generation, and evaluation, IEEE Trans. Knowl. Data Eng. 15 (1) (2003) 14–25.
[22] W.Z. Wu, W.X. Zhang, H.Z. Li, Knowledge acquisition in incomplete fuzzy information systems via the rough set approach, Expert Syst. Appl. 20 (5) (2003) 280–286.
[23] J. Grzymala-Busse, Knowledge acquisition under uncertainty — a rough set approach, J. Intell. Robot. Syst. 1 (1) (1988) 3–16.
[24] Y. Qian, J. Liang, F. Wang, A positive-approximation based accelerated algorithm to feature selection from incomplete decision table, Chin. J. Comput. 34 (3) (2011) 435–442.
[25] Q. Hu, S. An, D. Yu, Soft fuzzy rough sets for robust feature evaluation and selection, Inf. Sci. 180 (22) (2010) 4384–4400.
[26] H. Midelfart, J. Komorowski, K. Norsett, F. Yadetie, A. Sandvik, A. Lægreid, Learning rough set classifiers from gene expressions and clinical data, Fundam. Inform. 53 (2) (2002) 155–183.
[27] M. Banerjee, S. Mitra, H. Banka, Evolutionary rough feature selection in gene expression data, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 37 (4) (2007) 622–632.
[28] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, On the effectiveness of discretization on gene selection of microarray data, in: The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, 2010, pp. 1–8.
[29] L. El Ghaoui, G. Lanckriet, G. Natsoulis, Robust Classification with Interval Data, Computer Science Division, University of California, 2003.
[30] J. Sanz, A. Fernández, H. Bustince, F. Herrera, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning, Inf. Sci. 180 (19) (2010) 3674–3685.
[31] Z. Pawlak, Rough Sets — Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.
[32] Z.B. Xu, J.Y. Liang, C.Y. Dang, K.S. Chin, Inclusion degree: a perspective on measures for rough set data analysis, Inf. Sci. 141 (3–4) (2002) 227–236.
[33] L. Polkowski, A. Skowron, Rough mereology: a new paradigm for approximate reasoning, Int. J. Approx. Reason. 15 (4) (1996) 333–365.
[34] M. Zhang, W. Zhang, Knowledge reductions in information systems with fuzzy decisions, Chin. J. Eng. Math. 20 (2003) 53–58.
[35] S. Greco, M. Inuiguchi, R. Słowiński, Fuzzy rough sets and multiple-premise gradual decision rules, Int. J. Approx. Reason. 41 (2) (2006) 179–211.
[36] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, Fuzzy Set. Syst. 157 (11) (2006) 1477–1484.
[37] C. Zhang, H. Fu, Similarity measures on three kinds of fuzzy sets, Pattern Recognit. Lett. 27 (12) (2006) 1307–1317.
[38] Y. Leung, M.M. Fischer, W.Z. Wu, J.S. Mi, A rough set approach for the discovery of classification rules in interval-valued information systems, Int. J. Approx. Reason. 47 (2) (2008) 233–246.
[39] S. Greco, B. Matarazzo, R. Słowiński, Rough sets theory for multicriteria decision analysis, Eur. J. Oper. Res. 129 (1) (2001) 1–47.
[40] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. U. S. A. 98 (9) (2001) 5116–5121.
[41] D. Ross, U. Scherf, M. Eisen, C. Perou, C. Rees, P. Spellman, V. Iyer, S. Jeffrey, M. Van de Rijn, M. Waltham, et al., Systematic variation in gene expression patterns in human cancer cell lines, Nat. Genet. 24 (3) (2000) 227–235.
[42] T. Li, C. Zhang, M. Ogihara, A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression, Bioinformatics 20 (15) (2004) 2429–2437.
[43] M. Buschle, et al., Interferon gamma inhibits apoptotic cell death in B cell chronic lymphocytic leukemia, J. Exp. Med. 177 (1) (1993) 213–218.
[44] S. Orr, et al., CD33 responses are blocked by SOCS3 through accelerated proteasomal-mediated turnover, Blood 109 (3) (2007) 1061–1068.
[45] M. Scupoli, et al., Bone marrow stromal cells and the upregulation of interleukin-8 production in human T-cell acute lymphoblastic leukemia through the CXCL12/CXCR4 axis and the NF-κb and JNK/AP-1 pathways, Haematologica 93 (4) (2008) 524–532.
[46] D. Steinbach, J. Hermann, S. Viehmann, F. Zintl, B. Gruhn, Clinical implications of PRAME gene expression in childhood acute myeloid leukemia, Cancer Genet. Cytogenet. 133 (2) (2002) 118–123.
[47] M. Shah, et al., Molecular profiling of LGL leukemia reveals role of sphingolipid signaling in survival of cytotoxic lymphocytes, Blood 112 (3) (2008) 770–781.
[48] J. Roman-Gomez, et al., Promoter hypermethylation of cancer-related genes: a strong independent prognostic factor in acute lymphoblastic leukemia, Blood 104 (8) (2004) 2492–2498.
[49] J. Koo, I. Sohn, S. Kim, J. Lee, Structured polychotomous machine diagnosis of multiple cancer types using gene expression, Bioinformatics 22 (8) (2006) 950–958.
[50] I. Wadman, et al., Specific in vivo association between the bHLH and LIM proteins implicated in human t cell leukemia, EMBO J. 13 (20) (1994) 4831–4839.
[51] T. Macalma, et al., Molecular characterization of human zyxin, J. Biol. Chem. 271 (49) (1996) 31470–31478.
[52] H. Peng, C. Ding, Minimum redundancy and maximum relevance feature selection and recent advances in cancer classification, in: Feature Selection for Data Mining, 2005, p. 52.
[53] Z. Wang, M. Gerstein, M. Snyder, RNA-seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (1) (2009) 57–63.
[54] J. Marioni, C. Mason, S. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. 18 (9) (2008) 1509–1517.
[55] M. Robinson, G. Smyth, Moderated statistical tests for assessing differences in tag abundance, Bioinformatics 23 (21) (2007) 2881–2887.
[56] M. Robinson, D. McCarthy, G. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140.
[57] J. Bullard, E. Purdom, K. Hansen, S. Dudoit, Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments, BMC Bioinformatics 11 (1) (2010) 94.