



# Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification

Jianhua Dai\*, Qing Xu

College of Computer Science, Zhejiang University, Hangzhou 310027, China

## ARTICLE INFO

### Article history:

Received 22 January 2012

Received in revised form 26 June 2012

Accepted 23 July 2012

Available online 16 August 2012

### Keywords:

Attribute selection  
Mutual information  
Fuzzy rough sets  
Gain ratio  
Tumor classification

## ABSTRACT

Tumor classification based on gene expression levels is important for tumor diagnosis. Since tumor data in gene expression contain thousands of attributes, attribute selection for tumor data in gene expression becomes a key point for tumor classification. Inspired by the concept of gain ratio in decision tree theory, an attribute selection method based on fuzzy gain ratio under the framework of fuzzy rough set theory is proposed. The approach is compared to several other approaches on three real world tumor data sets in gene expression. Results show that the proposed method is effective. This work may supply an optional strategy for dealing with tumor data in gene expression or other applications.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

In clinical applications, physicians usually need to make feasible judgments on the therapies to patients by predicting therapy response, prognosis, metastatic phenotype, etc. before carrying out practical cancer treatments. Tumor is identified as systematic biology diseases of which the mechanism development is not completely known yet. Since tumor treatment of patients of later stage cancers is often not therapeutically effective, medical experts agree that early diagnosis of tumor is of great benefit to the successful therapies of tumor. Tumors are currently diagnosed by histology and immunohistochemistry based on their morphology and protein expression, respectively. However, poorly differentiated tumors can be difficult to diagnose by routine histopathology. In addition, the histological appearance of a tumor cannot reveal the underlying genetic aberrations or biological processes that contribute to the malignant process [27]. In recent years, gene expression profiles based molecular diagnosis of tumor has attracted great interests for the goal of realizing precise and early tumor diagnosis. The curse of dimensionality caused by high dimensionality and small sample size of gene expression data sets seriously challenges the tumor classification. There are thousands of genes in the gene expression data sets, but only a few of them are useful for classification. Therefore, the method for selecting important genes is the key issue for tumor diagnosis.

Rough set theory [36–39], introduced by Pawlak, is a useful mathematic tool for dealing with vague and uncertain information. It has attracted the attention of many researchers who have studied its theories and its applications during the last decades [5,4,7,11–13,26,33,34,44,52,56,58]. Rough set theory can achieve a subset of all attributes which preserves the discernible ability of original features, by using the data only with no additional information. Therefore, it has been playing an important role in attribute selection (also called attribute reduction or feature selection) [18,24,29,31,40,45,48,49,57,60].

However, classical rough set theory can only deal with nominal attribute values. Since continuous attribute values are more common in real world, the crisp rough set theory encounters a challenge. Therefore, it is desirable to develop techniques to handle this case. This can be achieved by using fuzzy rough sets [3,14,30,35,50,51,53,55]. Fuzzy-rough sets encapsulate the related but distinct concepts of vagueness (for fuzzy sets [59]) and indiscernibility (for rough sets), both of which occur as a result of uncertainty in knowledge [15].

Attribute selection of fuzzy rough set theory has been a popular topic in recent years. Jensen et al. [23,24] and Shen et al. [44] generalized the dependency function defined in classical rough set based on positive region into the fuzzy case and presented a rough-fuzzy attribute selection algorithm. Hu et al. [19] extended the Shannon entropy to measure the information quantity by fuzzy equivalence classes in fuzzy sets and reduce hybrid data sets. Actually, Hu et al. used the difference between the entropies before and after an attribute was selected as a criterion for attribute selection which can be called gain similar to Quinlan [42]. In decision tree

\* Corresponding Author.

E-mail address: [jhdai@cs.zju.edu.cn](mailto:jhdai@cs.zju.edu.cn) (J. Dai).

theory, gain has a tendency to prefer the attribute whose partition is more refined. It not only increases the size of the decision tree but also increases the error rate for unseen samples. As an improved version of gain, the gain ratio has been used as the attribute selection criterion in such well-known algorithms as GID3 [6], GID3\* [16], C4 [43], C4.5 [41], and so on. We could expect that gain ratio less prefers a finer partition. Inspired by these researches in decision tree learning, we introduce gain ratio into fuzzy rough theory and propose an attribute selection algorithm based on gain ratio in fuzzy rough set theory. The proposed approach is used in tumor classification problem. Experiments on tumor data sets in gene expression are conducted to evaluate the reduction and classification results compared to crisp rough set methods and fuzzy rough model based on gain.

This paper is organized as follows. Some basic notations of rough set theory and information measures are reviewed in Section 2. An attribute selection approach based on fuzzy information gain ratio is proposed in Section 3. In Section 4, experiments and comparisons on real gene expression tumor data have been conducted. Finally, Section 5 concludes the paper.

## 2. Preliminaries

In this section, we will review some basic notations in rough set theory and information measures which can be found in [2,20,28,36,40].

### 2.1. Information entropy in rough set theory

The concept of indiscernibility is central to rough set theory. Let  $IS = \langle U, A, V, f \rangle$  be an information system (also called information table), where  $U$  is a nonempty set of finite objects (usually called the universe);  $A$  is a nonempty finite set of attributes (or features);  $V$  is the union of attribute domains,  $V = \bigcup_{a \in A} V_a$ , where  $V_a$  is the value set of attribute  $a$ , called the domain of  $a$ ;  $f: U \times A \rightarrow V$  is an information function which assigns particular values from domains of attribute to objects such as  $\forall a \in A, x \in U, f(a, x) \in V_a$ , where  $f(a, x)$  denotes the value of attribute  $a$  for object  $x$ . With any  $B \subseteq A$  there is an associated indiscernibility relation  $IND(B)$ :

$$IND(B) = \{(x, y) | \forall a \in B, f(a, x) = f(a, y)\}. \tag{1}$$

Obviously,  $IND(B)$  is an equivalence relation, which is reflexive, symmetric and transitive. The family of all equivalence classes of  $IND(B)$  will be denoted by  $U/IND(B)$ , or simply  $U/B$ ; an equivalence class of  $IND(B)$  containing  $x$  will be denoted by  $[x]_B$ .

Let  $X$  be a subset of  $U$ , the lower approximation and the upper approximation are defined respectively as follows.

$$\underline{B}(X) = \bigcup_{x \in U} \{[x]_B | [x]_B \subseteq X\} \tag{2}$$

$$\overline{B}(X) = \bigcup_{x \in U} \{[x]_B | [x]_B \cap X \neq \emptyset\} \tag{3}$$

where  $\underline{B}(X)$  and  $\overline{B}(X)$  are called  $B$ -lower approximation and  $B$ -upper approximation with respect to  $B$ , respectively. The order pair  $\langle \underline{B}(X), \overline{B}(X) \rangle$  is called a rough set of  $X$  with respect to the equivalence relation  $IND(B)$ . Equivalently, they can be also defined as:

$$\underline{B}(X) = \{x \in U | [x]_B \subseteq X\} \tag{4}$$

$$\overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\} \tag{5}$$

Based on the lower approximation and upper approximation, the boundary region can be defined as:

$$BN_B(X) = \overline{B}(X) - \underline{B}(X) \tag{6}$$

The lower approximation of a set  $X$  with respect to  $IND(B)$  is the set of all objects, which certainly belongs to  $X$  with respect to  $IND(B)$ . The upper approximation of a set  $X$  with respect to  $IND(B)$  is the set of all objects, which possibly belongs to  $X$  with respect to  $IND(B)$ . The boundary region of a set  $X$  with respect to  $IND(B)$  is the set of all objects, which belongs with certainty neither to  $X$  nor to  $X^c$  with respect to  $IND(B)$ , where  $X^c$  denotes the complement of  $X$  in  $U$ .

An equivalence relation induces a partition of the universe. The partition of  $U$ , generated by  $IND(P)$  is denoted as  $U/IND(P)$ . Suppose  $P$  is a subset of  $A$ , and  $\mathbf{X}$  is the partition of the universe induced respectively by  $P$ , where

$$\mathbf{X} = U/IND(P) = \{X_1, X_2, \dots, X_n\}. \tag{7}$$

then probability distributions of  $X$  is defined as:

$$[\mathbf{X}; p] = \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ p(X_1) & p(X_2) & \dots & p(X_n) \end{bmatrix} \tag{8}$$

where  $p(X_i) = \frac{|X_i|}{|U|}$ ,  $i = 1, 2, \dots, n$ . The  $|\cdot|$  denotes the cardinality of a set.

**Definition 2.1.** [20,28] Given an information system  $IS = \langle U, A, V, f \rangle$ ,  $P \subseteq A$ ,  $U/P = \{X_1, X_2, \dots, X_n\}$ . The Shannon's entropy  $H(P)$  of  $P$  is defined by:

$$H(P) = - \sum_{i=1}^n p(X_i) \log p(X_i) = - \sum_{i=1}^n \frac{|X_i|}{|U|} \log \frac{|X_i|}{|U|}. \tag{9}$$

**Definition 2.2.** [20,28] Given an information system  $IS = \langle U, A, V, f \rangle$ ,  $P, Q \subseteq A$ ,  $U/P = \{X_1, X_2, \dots, X_n\}$  and  $U/Q = \{Y_1, Y_2, \dots, Y_m\}$ . The joint entropy of  $P$  and  $Q$  is defined as:

$$\begin{aligned} H(PQ) &= H(P \cup Q) = - \sum_{i=1}^n \sum_{j=1}^m p(X_i Y_j) \log p(X_i Y_j) \\ &= - \sum_{i=1}^n \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|U|} \log \frac{|X_i \cap Y_j|}{|U|}. \end{aligned} \tag{10}$$

where  $p(X_i Y_j) = \frac{|X_i \cap Y_j|}{|U|}$ .

**Definition 2.3.** [20,28] Let  $DS = \langle U, C \cup D, V, f \rangle$  be a decision system, where  $C$  is the condition attribute set and  $D$  is the decision attribute,  $B \subseteq C$ ,  $U/B = \{X_1, X_2, \dots, X_n\}$  and  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ . The conditional entropy of  $D$  conditioned to  $B$  is defined as

$$H(D|B) = - \sum_{i=1}^n p(X_i) \sum_{j=1}^m p(Y_j|X_i) \log p(Y_j|X_i). \tag{11}$$

where  $p(Y_j|X_i) = \frac{|X_i \cap Y_j|}{|X_i|}$ . Then, we can write the conditional entropy as

$$\begin{aligned} H(D|B) &= - \sum_{i=1}^n \frac{|X_i|}{|U|} \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|X_i|} \log \frac{|X_i \cap Y_j|}{|X_i|} \\ &= - \sum_{i=1}^n \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|U|} \log \frac{|X_i \cap Y_j|}{|X_i|}. \end{aligned} \tag{12}$$

**Definition 2.4.** [28,32] Let  $DS = \langle U, C \cup D, V, f \rangle$  be a decision system, where  $C$  is the condition attribute set and  $D$  is the decision attribute,  $B \subseteq C$ ,  $U/B = \{X_1, X_2, \dots, X_n\}$  and  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ . The mutual information of  $B$  and  $D$  is defined as

$$I(B; D) = H(D) - H(D|B). \tag{13}$$

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات