Attribute Value Generalization Reduction Based on Conditional Combination Entropy

Hao Zhang, Ming-Quan Ye*

Department of Computer Science, Wannan Medical College, Wuhu, 241002, China

Chang-Rong Wu

School of mathematics & Computer Science, Anhui Normal University, Wuhu, 241001, China
*Corresponding author Email: ymq@wnmc.edu.cn

Abstract
Attribute reduction of decision tables as a major topic in the rough set theory plays its significant role in machine learning and data mining. So far more studies have investigated flat data tables and eliminated the redundant attributes with no effort in reducing the classification performance. However, data with attribute value taxonomies (AVTs) are usually found in real-world applications. In this study, the concept of generalization reduction based on FSG and rough set theory is proposed for finding the optimum generalization levels without losing essential classification information. The conditional combination entropy is adopted to measure the classification ability of hierarchical decision tables, which develops a heuristic algorithm GRCCE to compute generalization reduction from training data with AVTs. The proposed algorithm not only determines an optimal generalization level but also induces the most abstract hierarchical decision table while keeping the same classification ability on the raw decision table. An example is explored to illustrate the GRCCE algorithm is effective to conduct attribute reduction of the decision tables.
Key words ROUGH SET THEORY, GENERALIZATION REDUCTION, COMBINATION ENTROPY, ATTRIBUTE VALUE TAXONOMIES, HIERARCHICAL DECISION TABLE

1. Introduction

Rough set theory presented by Pawlak [1, 2] is a meaningful mathematical tool while dealing with vague, uncertain and imprecise information. Attribute reduction is an important research topic in the rough sets [3-13], which without reducing the classification performance can eliminate the redundant condition attributes effectively. Nowadays, rough set theory has been successfully applied to many field such as data mining [6, 7], knowledge acquisition [8, 9], pattern recognition [10, 11]. Since attribute reduction offers opportunities to discover useful knowledge from training examples, researchers have focused on attribute reduction of static decision tables, i.e., in a certain decision table, the objects and attributes remain constant [3-6]. However, the decision table evolves with time, namely, attribute sets, object sets, or attribute value sets may change. Practical applications face often dynamic decision tables which have been studied intensively in the past few years [6, 12, 14], but the existing methods mainly focus on reducing the number of attributes at single-level of abstraction.

In practice, the attribute values changes according to various reasons, for instance, on the one hand the revise of the occurring errors and on the other hand the variation in the concept level of attribute value taxonomies [13]. In real-world applications, decision tables often face attribute values usually represented by AVTs [13-25]. In every attribute value taxonomy, the root is the “ANY” value of an attribute, leaf nodes are attribute values in the given data table, and the representation of internal nodes is generalized attribute values of their child nodes. With a parent value in the AVT of an attribute, a generalization replaces some values. There exist two generalization models for the classification task, i.e. full-domain generalization (FDG) [20-25] and full-subtree generalization (FSG) [14-18]. All values in an attribute are generalized to the same level of the AVT in the FDG. In the FSG, at a non-leaf node, either all child values or none are generalized, and a generalized attribute has values that form a “cut” through its AVT. Recently, rough set theory has been used in decision rule mining for data with AVTs [23-25]. From different levels by combining the hierarchical structure of multidimensional data model and the techniques of rough set theory, Feng et al. [23] proposed approaches to improvement of the quality and efficiency of decision rules mining. Hong et al. [24] presented the algorithms, which under fuzzy rough sets model could mine cross-level rules. In addition, Hong et al. [25] suggested to combine the fuzzy-set theory and the rough set theory, with hierarchical and quantitative attributes, to cope with generating a set of cross-level maximally general possible and fuzzy certain rules from training data. But these investigations still lack the adaptability in solving data at multiple levels of abstraction. Especially, they adopted the FDG model on the AVTs, which caused the largest distortion of the data, on all paths of a taxonomy tree, for the same granularity level requirement.

The FSG model on the AVTs offers a way, at different levels of granularity for generating accurate, compact, and comprehensible classifiers, to make the organization of data [14-18]. In this paper, with the FSG model on the AVTs, we only investigate decision tables. In other words, how to find an optimal generalization level for maintaining the classification ability of the raw decision table when the condition attribute values evolve from a relatively low concept level to higher concept level while both the attribute set and object set remain constant. Based on the integration of AVTs and FSG, a hierarchical decision table is induced by generalization of a given flat table to multiple decision tables with different levels of abstraction. Hierarchical decision tables of different condition attributes at different levels can be organized as a generalization lattice, in which each node represents a hierarchical decision table. It is possible to find an optimal node which shares the same classification ability on the raw decision table in the generalization lattice. Attribute reduction for a single-level decision table in rough set theory focuses on eliminating redundant attributes while retaining the classification ability of the original condition attributes. In the spirit of attribute reduction, an innovative concept of Generalization reduction based on both rough set theory and FSG model on AVTs is proposed. Generalization reduction for a hierarchical decision tables focuses on generalizing condition attribute values to an optimal concept level while retaining the classification ability of the original condition attribute values. Therefore, Hierarchical decision tables by generalization reduction with the same classification ability on the raw decision table can induce the most abstract hierarchical decision table.

Recently, the application of Shannon’s entropy and its generalizations has been wide, and they measure uncertainty in rough set theory [26, 27]. Qian and Liang [27] presented, for measuring the uncertainty of decision tables, the combination entropy and utilized its conditional entropy to select a feature subset. This reduction method can obtain an attribute subset retaining the same number of pairs of the elements recognized each other as the original decision table. The conditional combination entropy in a flat decision table can be applied to indicate the condition attribute’s classification potential. In decision tables, it also can be applied to construct the corresponding significance measures of attributes. In this paper, with different generalization levels, the conditional combination entropy is applied to measure the ability of classification in hierarchical decision tables, and attempt to make the combination of both FSG model and rough set theory on AVTs for achieving generalization reduction.

The rest of this paper is organized as follows. In Section 2, we introduce some basic concepts of rough set theory, conditional combination entropy, and FSG model. In Section 3, based on FSG model and rough set theory, we develop some novel concept for data with AVTs. In Section 4, we propose a generalization reduction algorithm, namely GRCCE, and employ an example to illustrate the GRCCE method. Section 5 is the summarization of the conclusions and presentation of future research topics.

2. Preliminaries

In this section, we introduce several concepts of rough set theory, conditional combination entropy, and FSG model.

2.1. Rough Set Theory

Several concepts of rough set theory are introduced as follows.

Definition 1 (*Decision table*). A decision table (or decision information system) is defined as an ordered tuple S=<U, A, V, f>, where U={x_1, x_2,…, x_n} is treated as a finite set of objects; A=C∪D is a finite set of attributes, while C is believed to be a set of condition attributes, and D refers to a set of decision attributes, C∩D=Ø; V= $\cup_{\alpha\in A} V_a$ where V_a is the domain of attribute *a*; f: U×A→V is an information function associating a unique value of every attribute with each object which belonging to U, such that for any x∈U and a∈A, f(x, a)∈V_a

Definition 2 (*Indiscernibility relation*). Given a decision information system S=<U, C∪D V, f> and an attribute set P ⊆ (C∪D) P determines an indiscernibility relation IND(P) on U as follows: IND(P)={(x, y)∈ U×U:a∈P, f(x, a)=f(y, a)} (1)

By U/IND(P) (or U/P), the equivalence relation IND(P) partitions the set U into disjoint subsets denoted while an equivalence class is named by an element of IND(P).

For every object x∈U , let $[x]_P$ denotes the equivalence class of relation IND(P) that involves element x, be known as the equivalence class of x according to relation IND(P).

Definition 3 (Conditional combination entropy). Given a decision information system S=<U, C∪D, V, f>, U/C={X_1,X_2,…,X_m}, and U/D={Y_1,Y_2,…,Y_n}, the conditional combination entropy of C with respect to D is defined as

$$CCE(D\mid C)=\sum_{i=1}^{m}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}}-\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{C^2_{|X_i\cap Y_j|}}{C^2_{|U|}}\right) \quad (2)$$

where $C^2_{|X_i|}=\frac{|X_i|\times(|X_i|-1)}{2}$, $\frac{|X_i|}{|U|}$ means the probability of an equivalence X_i within the universe U, and $\frac{C^2_{|X_i|}}{C^2_{|U|}}$ denotes the probability of pairs of the elements. Within the whole number of pairs of the elements on the universe U, these elements are not distinguishable each other.

By Definition 3, the classification capability of all condition attributes with regard to the decision attribute can be denoted by CCE(D|C).

2.2. FSG Model

Some definitions of FSG model are introduced as follows.

Definition 4 (*Attribute value taxonomy*). The attribute value taxonomy AVT(a) for attribute a equals to a tree-structured concept hierarchy, which is a partially ordered set (V_a, ≺), while V_a, as a finite set, enumerates every attribute value in a. Besides this, ≺, as the partial order, specifies a relationship among attribute values in V_a. Collectively, AVTs(P)={AVT(a_1), AVT(a_2), …, AVT(a_m)}, associated with P={a_1,a_2,…,a_m}, is the set ordered of attribute value taxonomies.

In AVT(a), Node(a) is the set of all nodes, Root(a) is the set of the root node and Leaf(a) is the set of leaf nodes. It corresponds to actual attribute values of attribute a which shows its appearance in training examples. Internal nodes (i.e. Node(a)−Leaf(a)) denote generalized attribute values of their child nodes, and each arc of the tree reflects its correspondence to a relationship over attribute values. Let Child(v, a) represent the set of a node's all children with correspondence to value v in the AVT(a).

Intuitively, with the FSG model, a "cut" through the AVTs represents a valid generalization. In order to recognize distinct types of cuts, the "cut" through one attribute AVT is called a local cut, and the "cut" through the attribute set AVTs is called a global cut. A local cut γ for AVT(a) is illustrated in the following part.

Definition 5 (*Cut*). A local cut γ, as a subset of elements in Nodes(a), satisfies the following conditions:

(1) Either p∈γ or p is a descendant of an element q∈γ for any leaf nodes p∈ Leaf(a);.

(2) p is neither a descendant nor an ancestor of q for any two nodes p, q∈γ.

For attribute a, we denote Ω_a as the set of all valid local cuts in AVT(a); as the Cartesian product of the cuts, we denote $\Omega_A=\Omega_1\times\Omega_2\times_{...}\times\Omega_m$ through the individual AVTs(P) for attribute set P={a_1,a_2,…,a_m}. For a given global cut Γ, if every cut γ_i∈Γ is selected to be equal to the respective attribute's primitive values a_i, i.e., γ_i=Leaf(a_i), then Γ_{leaf} denote the global cut Γ. Given a global cut Γ, if every cut γ_i∈ Γ is selected to traverse the root of every AVT, and matches to the respective attribute' most generalization values, i.e., γ_i=Root(a_i), then Γ_{root} represents the global cut Γ.

Definition 6 (*Cut refinement and generalization*). In AVT(a), for a local cut γ of attribute a, v∈ γ, a refinement of γ means: γ'=γ∪Child(v,a)−{v}, denoted by γ'≼γ. Correspondingly, for a global cut Γ, γ∈Γ, γ'≼γ, a refinement of Γ is defined as: Γ'=Γ∪ {γ'}−{γ}, denoted by Γ'≼Γ. Conversely, γ is a generalization of the local cut γ' and the same as Γ to the global cut Γ', respectively.

Example 1. Consider a flat decision table with a condition attribute set C={CP, STS} and a decision attribute set D={HD} in Fig. 1. Illustration in Fig. 2

shows a global cut refinement process based on the AVTs(C). The local cut γ_2={Any_STS} in the attribute of STS has been refined to γ_2'={upsloping, nonupsloping} by replacing Any_STS with its two children, upsloping and nonupsloping. Therefore, the global cut Γ_2={{asymptomatic, chest discomfort, angina}, {upsloping, nonupsloping}} is refined by the global cut Γ_1={{asymptomatic, chest discomfort, angina},{Any_STS}}.

U	Chest Pain (CP)	ST Slope (STS)	Heart Disease (HD)
x_1	chest discomfort	flat	absence
x_2	chest discomfort	downsloping	absence
x_3	asymptomatic	flat	presence
x_4	asymptomatic	downsloping	presence
x_5	typical angina	flat	presence
x_6	typical angina	downsloping	presence
x_7	typical angina	upsloping	absence
x_8	atypical angina	flat	presence
x_9	atypical angina	downsloping	presence
x_{10}	atypical angina	upsloping	absence

Figure 1. A falt decision table S

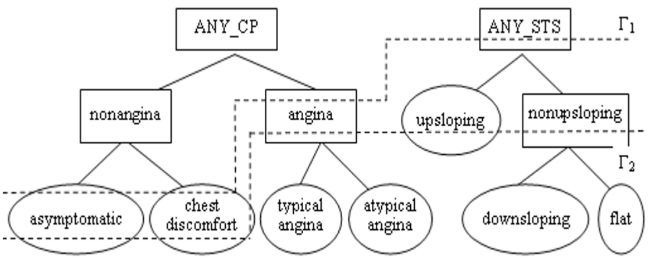


Figure 2. A global cut refinement based on AVTs(C)

3. Generalization reduction for hierarchical decision tables

In this section, AVTs are introduced into a flat decision table, and some concepts for data with AVTs such as hierarchical decision table and generalization reduction are proposed based on rough set theory and FSG model.

Most researches on rough sets pay attention to a flat decision table. A flat decision table is usually given in real-world applications, from which the inherent hierarchical characteristics of data cannot be reflected effectively. A concept hierarchy is introduced into a decision table, and many decision tables with different levels of abstraction will be obtained. Given a global cut Γ, a hierarchical decision table S_Γ can be produced from S which generalized from all attribute values uniquely in S to their taxonomical ancestors in Γ, meanwhile, other attribute values in S are not changed. The hierarchical decision table is defined in the following part [15-17].

Definition 7 (*Hierarchical decision table*) Given a decision table S=<U,C∪D, V, f> with AVTs(C), with respect to AVTs(C), a hierarchical decision table S_Γ on S is determined by a global cut $\Gamma=\{\gamma_1, \gamma_2,\dots, \gamma_m\}$ through AVTs(C):

$$S_\Gamma=\langle U,C\cup D,V_\Gamma, f_\Gamma\rangle \quad (3)$$

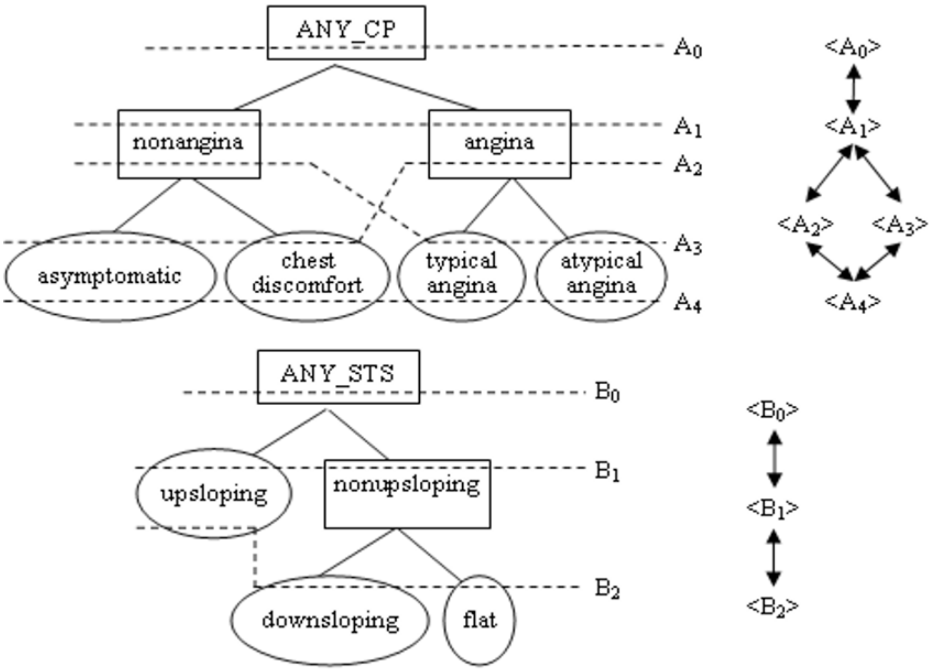
where for any $a\in D$, $V_a=\gamma_i$ for any $a=a_i\in C$; f_Γ, $V_\Gamma=\cup_{a\in C\cup D}V_a$, V_a is unchanged: $U\times C\cup D\rightarrow V_\Gamma$ is, as an information function, such that for any $a\in C\cup D$ and $x\in U$, $f_\Gamma(x,a)\in V_a$.

For a hierarchical decision table S_Γ, the U/C, U/(C∪D), and CCE(D|C) are denoted by U/C_Γ, $U/(C_\Gamma\cup D)$, and $CCE(D|C_\Gamma)$, respectively.

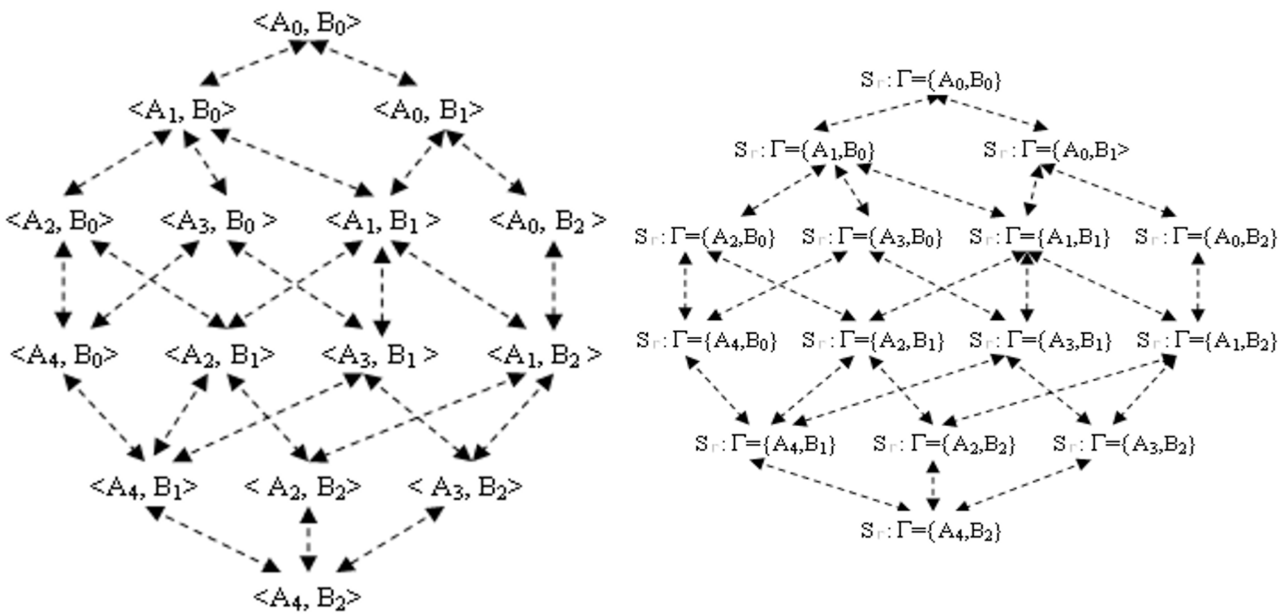
When some of its condition attribute values are generalized to a higher level of abstraction for a decision table, it will change its domain correspondingly and information function. Thus, it will change the decision table. The global cut Γ_{leaf} can determine the original decision table S which denoted by $S_{\Gamma_{leaf}}$. Different hierarchical decision tables will be determined by attribute values with abstraction in different levels.

Definition 8 With AVTs(C), given a decision table S=<U,C∪ D, V, f>, and two global cuts Γ_1, $\Gamma_2\in\Omega_C$, if $\Gamma_1\preccurlyeq\Gamma_2$, then hierarchical decision table S_{Γ_1} is finer than S_{Γ_2}, or S_{Γ_2} is coarser than S_{Γ_1}, denoted by $S_{\Gamma_1}\preccurlyeq S_{\Gamma_2}$.

Example 2. Consider the flat table in Fig. 1 and the AVTs(C) in Fig. 2. In Fig. 3, all the valid local cuts for each of the two condition attributes in fig.2 is denoted as a lattice of local cuts. The global cuts for the two condition attributes can be denoted as a lattice of global cuts, as in Fig. 4(a), and the corresponding lattice of hierarchical decision tables is shown in Fig. 4(b). In Fig. 4(a), the arrows illustrate the possible cut generalization (or refinement) paths that can be taken through the lattice, each node denotes a global cut, the top node $\langle A_0,B_0\rangle$ denotes the most generalization cut, i.e. Γ_{root}, and the bottom node $\langle A_4,B_2\rangle$ denotes the most refinement cut, i.e. Γ_{leaf}. A series of connected paths from the bottom node to the top node represents a cut generalization strategy. Conversely, a series of connected paths from the top node to the bottom node represents a cut refinement strategy.



Figrure 3. All valid local cuts and a lattice of local cuts for each attribute of {CP,STS}



(a) A lattice of global cuts for {CP, STS} (b) A lattice of hierarchical decision tables

Figure 4. A lattice of global cuts and corresponding lattice of hierarchical decision tables

In rough sets, attribute reduction for a flat decision table focuses on eliminating redundant attributes while retaining the classification ability of the original condition attributes. Given a flat table and a global cut, a hierarchical decision table is induced with respect to different condition attributes at various levels of abstraction. With the same classification ability, it is important to find the most abstract hierarchical decision table for the raw decision table, and no other hierarchical decision table is more abstract than it. In this paper, the conditional combination entropy is used to measure hierarchical decision tables' classification ability with various levels of generalization. In the spirit of the attribute reduction, an innovative concept of generalization reduction based on both rough set theory and FSG model on AVTs is defined as follows.

Definition 9 (*Generalization reduction*) With AVTs(C), given a decision table $S=<U,C\cup D,V,f>$, a global cut $\Gamma\in\Omega_C$, C_Γ is a generalization reduction of C referring to D if

(1) $CCE(D|C_\Gamma)=CCE(D|C)$;

(2) if $\Gamma \preccurlyeq \Gamma'$ for any $\Gamma' \in \Omega_C$, then $CCE(D|C_{\Gamma'}) \neq CCE(D|C)$.

Generalization reduction for decision tables with AVTs is focused on generalizing condition attribute values to an optimal concept level while retaining the classification ability of the original condition attribute values.

From Definition 9, we can obtain two properties about the lattice of hierarchical decision tables. First, if a node N is found to be a generalization reduction table, then all nodes below N on the same generalization strategies that pass through N are also have the same ability in classification just as every condition attribute with respect to the decision attribute on the raw table S. Second, if a node N is found not to be generalization reduction table, then all nodes above N on the same generalization strategies that pass through N are also not generalization reduction table. One of these nodes in Fig. 4(a) or corresponding node in Fig. 4(b) is the optimal solution and the objective of a generalization reduction algorithm is to find it efficiently.

Considering attribute a with its AVT(a), the number of every valid local cut is computed recursively by

$$\#cuts(Root(a))=1+\Pi_{r\in Child(Root(a),\, a)}\#cuts(r) \quad (4)$$

where Child(Root(a),a) is the set of Root(a)'s all child nodes.

Obviously, the number of cuts makes exponential growth with that of internal nodes on the AVT of an attribute. Therefore, the general method is to apply a heuristic search, which contains two important aspects: evaluation of a candidate node and search strategy through the lattice of global cuts. We take the conditional combination entropy to evaluate the node while the search strategy and generalization reduction method will be discussed in the next Section.

4. Generalization reduction algorithm based on conditional combination entropy

We take the conditional combination entropy to construct the corresponding significance measures of attribute values refinement, and propose a heuristic algorithm (GRCCE) to obtain generalization reduction from data with AVTs in this section. We also adopt an example to demonstrate the proposed method.

4.1. Attribute values refinement significance

Given AVTs(C) and a decision table $S=\langle U,C\cup D,V,f\rangle$, the goal is to find an optimal cut Γ such that S_Γ is the most abstract hierarchical decision table with the same classification ability of C with regard to D on the raw decision table S. Suppose the search of S_Γ is done by iteratively specializing a general value into its child values. Generalization reduction C_Γ is refined by beginning from the most abstract level basing on the most abstract value of each condition attribute and refining the global cut successively while using a criterion devised for accuracy in classification. In order to measure quantitatively the benefits of a refinement, we develop a utility refinement significance measure as follows.

Definition 10 (*Refinement Significance*) With AVTs(C), given a decision table $S=\langle U,C\cup D,V,f\rangle$, a global cut $\Gamma\in \Omega_C$, $a\in C$, and $\gamma_a \in \Gamma$. If $v\in\gamma_a$ and $v\notin$ Leaf(a), then the significance of the refinement (i.e. $v\rightarrow$Child(v, a)) is defined as

$RS(v, \gamma_a, \Gamma)=CCE(|D|C_\Gamma)-CCE(|D|C_{\Gamma'})$,

where $\Gamma'=\Gamma\cup \{\gamma_a'\}-\{\gamma_a\}$, $\gamma_a'=\{\gamma_a\}\cup Child(v, a)-\{v\}$, Γ' denotes a refinement of a global cut Γ by replacing v with its Child(v, a), $CCE(D|C_\Gamma)$ denotes the conditional combination entropy of C with respect to D before the refinement, and $CCE(D|C_{\Gamma'})$ is the conditional combination entropy of C with regard to D after the refinement.

For the current global cut $\Gamma$and the given attribute value v, if Γ' is obtained by replacing v with its children in Γ, the $CCE(D|C_{\Gamma'})$ decreases faster and the $maxsize(CCE(D|C_\Gamma)-CCE(D|C_{\Gamma'}))$ is larger than by the replacement of any other attribute value, which increases faster the classification ability.

If $RS(v, \gamma_a, \Gamma)=0$, we say $v\rightarrow$Child(v, a) is superfluous, which means the refinement is useless in Γ.

Example 3. Within the consideration of the decision table in Fig. 1 and the AVTs(C) in Fig. 2. Two global cuts are fulfilled as follows:

Γ_1={{ANY_CP},{ANY_STS}}, and

Γ_2={{nonangina, angina}, {ANY_STS}}.

The refinement step from Γ_1 down to Γ_2 is by specializing ANY_CP into nonangina and angina. Let γ_1={ANY_CP}.

The calculation of the RS of the refinement {ANY_CP}→{nonangina, angina} is as follows.

$U/D=\{\{x_1,x_2,x_7,x_{10}\},\{x_3,x_4,x_5,x_6,x_8,x_9\}\}$,

$U/C_{\Gamma_1}=\{U\}$, and

$U/C_{\Gamma_2}=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6,x_7,x_8,x_9,x_{10}\}\}$.

Therefore, $CCE(D|C_{\Gamma_1})=0.7476$, and $CCE(D|C_{\Gamma_2})=0.1867$.

Computing the RS of {ANY_CP}→{nonangina, angina}, we have that RS(ANY_CP, γ_1, Γ_1)=0.56.

4.2. KRCCE Algorithm

In this paper, a new generalization reduction algorithm based on Conditional Combination Entropy (GRCCE) is proposed, and the pseudo-code is shown in Table 1. In the pseudo-code, only the top most value for each condition attribute the global cut Γ contains. The valid refinements make up the set of candidates to be fulfilled next. In each iteration, firstly, for each valid v (i.e not a leaf node) in γ and $\gamma\in\Gamma$, we compute the significance measure RS(v,γ,Γ) (Lines5-9); Secondly, we explore the candidate of the highest RS(va,γa,Γ), denoted va, and refine Γ by replacing va with its child values (Lines 10-11). Finally, we update SΓ with Γ and compute CCE(D|CΓ) (Line 12). The algorithm terminates when CCE(D|CΓ)=CCE(D|C), under which circumstance, the reduction table SΓ is returned together with the generalization reduction CΓ.

Table 1. The pseudo-code of GRCCE

Algorithm 1. GRCCE
Input: S=(U,C∪D,V,f), AVTs(C), and C={$a_1,a_2,\ldots,a_m$}};
Output: the generalization reduction C_Γ, and the reduction hierarchical table S_Γ.
1. According to the raw table S, compute CCE(D|C);
2. Γ={$\gamma_1,\gamma_2,\ldots,\gamma_m$}, where γ_i of a_i is the top most value;
3. Construct the hierarchical decision table S_Γ, and compute CCE(D|C_Γ);
4. While CCE(D|C_Γ)≠CCE(D|C) do
5. For any $\gamma_i\in\Gamma$ do
6. For any $v\in\gamma_i$ and v∉Leaf(a_i) do
7. Compute RS(v, γ_i, Γ);
8. Endfor
9. Endfor
10. RS(v_a,γ_a,Γ)=max{RS(v, γ_i, Γ), $v\in\cup\gamma_i$};
11. Γ=Γ∪Child(v_a, a)−{v_a};
12. Update S_Γ with Γ, and compute CCE(D|C_Γ);
13. Endwhile;
14. Return C_Γ, and S_Γ.

4.3. An Example

An example serves to illustrate the GRCCE algorithm bellow.

Example 4. Consider the decision table in Fig. 1 and the AVTs(C) in Fig. 2. We illustrate the generalization reduction process using GRCCE algorithm in Fig. 5.

To begin with, we have CCE(D|C)=0, and U/D={{x_1,x_2,x_7,x_{10}},{x_3,x_4,x_5,x_6,x_8,x_9}}.

The top most hierarchical decision table contains one row on C with Γ={A_0,B_0}, where A_0=Root(CP)={ANY_CP}, and B_0=Root(STS)= ={ANY_STS}.

From here, we can obtain that CCE(D|C_Γ)=0.7467.

Note that CCE(D|C_Γ)≠CCE(D|C), and then we calculate the RS of the three candidate refinements:

{ANY_CP}→{nonangina, angina}, and
{ANY_STS}→{upsloping, nonupsloping}.
From here, we can obtain that
RS(ANY_CP, A_0, Γ)=0.56, and
RS(ANY_Age, B_0, Γ)=0.4534.

Relating to the RS criterion, the first refinement of ANY_CP will occure according to the highest RS. The result is shown in Fig. 5. Then, we can obtain that

A_1={nonangina, angina}, Γ={A_1, B_0}}, and CCE(D|C_Γ)=0.1867.

Note that CCE(D|C_Γ)≠CCE(D|C), and so we need to further calculate the RS of the three candidate refinements:

{nonangina}→{asymptomatic, chest discomfort},
{angina}→{typical angina, atypical angina}, and
{ANY_STS}→{upsloping, nonupsloping}.
From here, we obtain that
RS(nonangina, A_1, Γ)=0.0445, RS(angina, A_1, Γ)=0.1111, and RS(ANY_STS, B_0, Γ)=0.1423.

Relating to the RS criterion, the further refinement of ANY_STS will happen according to the highest RS. The result is shown in Fig. 5. Then, we obtain that

B_1={upsloping, nonupsloping}, Γ={A_1,B_1}, and CCE(D|C_Γ)=0.0444.

Note that CC(D|C_Γ)≠CCE(D|C), and so we need to further calculate theRS of the three candidate refinements:

{nonangina}→{asymptomatic, chest discomfort},
{angina}→{typical angina, atypical angina}, and
{nonupsloping}→{flat, downsloping}.
From here, we obtain that
RS(nonangina, A_1, Γ)=0.0444,
RS(nangina, A_1, Γ)=0, and
RS(ANY_STS, B_1, Γ)=0.0355.

Relating to the RS criterion, the further refinement of nonangina will occure according to the highest RS. The result is shown in Fig. 5. Then, we obtain that

A_2={asymptomatic, chest discomfort, angina}, Γ={A_2, B_2}, and CCE(D|C_Γ)=0.

Note that CCE(D|C_Γ)=CCE(D|C). Hence, C_Γ is a Generalization reduction, where Γ={{A_2, B_2}}.

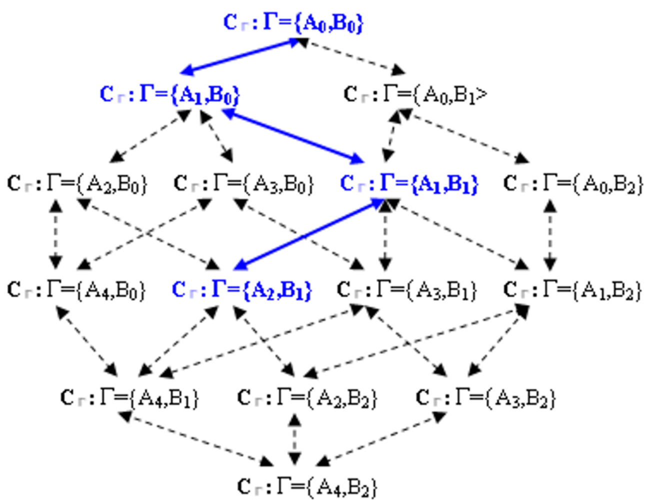


Figure 5. An illustration of generalization reduction process using GRCCE

According to the generalization reduction CΓ, we obtain the reduction table SΓ as shown in Fig.6. From the reduction table, we have four decision rules as follows.

(1) If CP=”chest discomfort” and STS=”nonupsloping”, then HD=”absence”.

(2) If CP=”asymptomatic” and STS=”nonupsloping”, then HD=”presence”.

(3) If CP=”angina” and STS=”nonupsloping”, then HD=”presence”.

(4) If CP=”angina” and STS=”upsloping”, then HD=”absence”.

From the flat decision table in Fig.1, we obtain ten decision rules. In Example 4, from different concept levels, there are only four items of decision rules mined, while from the primitive level, there exist ten items of decision rules mined. We observe that the number of decision rules mined from the generalization reduction table with different degrees of generalization is much fewer than those mined from the raw flat decision table at the primitive level. Furthermore, we find the optimum levels without loosing essential classification information.

U	Chest Pain (CP)	ST Slope (STS)	Heart Disease (HD)
x_1, x_2	chest discomfort	nonupsloping	absence
x_3, x_4	asymptomatic	nonupsloping	presence
x_5, x_6, x_8, x_9	angina	nonupsloping	presence
x_7, x_{10}	angina	upsloping	absence

Figure 6. Reduction table S_Γ with generalization reduction C_Γ

5. Conclusions

Attribute reduction as an important research topic in the rough set theory has been explored intensively for a few ages. We have proposed a concept of ge- neralization reduction which aims to find the optimum levels without losing essential classification information by combining the FSG model on AVTs and rough set theory. We have employed the conditional combination entropy based on rough set theory to measure the classification ability of hierarchical decision tables, and presented a heuristic algorithm GRCCE to compute generalization reduction from training data with AVTs. The proposed algorithm can not only determine an optimal generalization level with various degrees of generalization, but also induce the most abstract hierarchical decision table while keeping the same classification ability on the raw decision table. The results from experiment have illustrated that the proposed algorithm shows its effectiveness to conduct attribute reduction of flat decision tables.

In the future, there are many additional issues to be studied further such as (1) generalization reduction with respect to other ability in classification measures in rough sets, (2) the proposed method extending to hierarchical decision rule mining, and (3) considerations on not only the classification ability, but also coverage and strength.

Acknowledgments

This work was sponsored by the Provincial Natural Science Research Key Project of Anhui Colleges under Grant KJ2014A266.

References

1. Z. Pawlak (1982) Rough Sets. *International Journal of Computer and Information Science*, 11(5), 341-356.
2. Z. Pawlak (2007) A. Skowron, Rough sets: some extensions. *Information Sciences*, 177(1), p.p.28-40.
3. N. Parthalain, Q. Shen, R. Jensen (2010) A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE Transactionson Knowledge and Data Engineering*, 22(3), p.p.305-317.
4. L. Sun, J. Xu, Y. Tian (2012) Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. *Knowledge-Based Systems*, 36, p.p.206-216.
5. Y. Yao, Y. Zhao (2008) Attribute reduction in decision-theoretic rough set models. *Information Sciences*, 178(17), p.p.3356-3373.
6. C. Liu, D. Miao, N. Zhang (2012) Graded rough set model based on two universes and its properties. *Knowledge-Based Systems*, 33, p.p.65-72.
7. S. Foithong, O. Pinngern, B. Attachoo (2012) Feature subset selection wrapper based on mutual information and rough sets. *Expert Systems with Applications*, 39(1), p.p.574-584.
8. Y.Y. Guan, H.K. Wang, Y. Wang, F. Yang (2009) Attribute reduction and optimal decision rules acquisition for continuous valued information systems. *Information Sciences*, 179(17), p.p.2974-2984.
9. L. Feng, T.R. Li, D. Ruan, S. Gou (2011) A vague-rough set approach for uncertain knowledge acquisition. *Knowledge-Based Systems*, 24(6), p.p.837-843.
10. Y.H. Qian, J.Y. Liang,W. Pedrycz, C.Y. Dang (2011) An efficient accelerator for attribute reduction from incomplete data in rough set framework. *Pattern Recognition*, 44(8), p.p.1658-1670.

11. M. Salamó, M. López-Sánchez (2011) Rough set based approaches to feature selection for Case-Based Reasoning classifiers. *Pattern Recognition Letters*, 32(2), p.p.280-292.
12. Y. Xu, L. Wang, R. Zhang (2011) A dynamic attribute reduction algorithm based on 0-1 integer programming. *Knowledge-Based Systems*, 24(8), p.p.1341-1347.
13. H. Chen, T.R. Li, D. Ruan (2012) Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining. *Knowledge-Based Systems*, 31, p.p.140-161.
14. J. Zhang, D.K. Kang, A. Silvescu, V. Honavar (2006) Learning accurate and concise naive Bayes classifiers from attribute value taxonomies and data. *Knowledge and Information Systems*, 9(2), p.p.157-179.
15. M.Q. Ye, X.D. Wu, X.G. Hu, D.H. Hu (2014) Knowledge reduction for decision tables with attribute value taxonomies. *Knowledge-Based Systems*, 56(1), p.p.68-78.
16. M.Q. Ye, X.D. Wu, X.G. Hu, D.H. Hu (2013) Anonymizing classification data using rough set theory. *Knowledge-Based Systems*, 43(1), p.p.82-94.
17. M.Q. Ye, X.D. Wu, X.G. Hu, D.H. Hu (2013) Multi-level rough set reduction for decision rule mining. Applied Intelligence, 39(3), p.p.642-658.
18. D.K. Kang, M.J. Kim, Propositionalized attribute taxonomies from data for data-driven construction of concise classifiers. *Expert Systems with Applications*, 38(10), p.p.12739-12746.
19. X.H. Hu, N. Cercone (2001) Discovering maximal generalized decision rules through horizontal and vertical data reduction. *Computational Intelligence*, 17(4), p.p.685-702.
20. J. Han, Y. Fu (1999) Mining multiple-level association rules in large databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(5), p.p.798-805.
21. Y.L. Chen, Y.Y. Wu, R.I. Chang, From data to global generalized knowledge. *Decision Support Systems*, 52(2), p.p.295-307.
22. Y.Y. Wu, Y.L. Chen, R.I. Cha (2011) Mining negative generalized knowledge from relational databases. *Knowledge-Based Systems*, 24(1), p.p.134-145.
23. Q. Feng, D. Miao, Y. Cheng (2010) Hierarchical decision rules mining. *Expert Systems with Applications*, 37(3), p.p.2081-2091.
24. T.P. Hong, C.E. Lin, J.H. Lin, S.L.Wang (2008) Learning cross-level certain and possible rules by rough sets. *Expert Systems with Applications*, 34(3), p.p.1698-1706.
25. T.P. Hong, Y.L. Liou, S.L. Wang (2009) Fuzzy rough sets with hierarchical quantitative attribute. *Expert systems with Applications*, 36(3), p.p.6790-6799.
26. J.H. Dai, W.T. Wang, Q. Xu, H.W (2012) Tian, Uncertainty measurement for interval-valued decision systems based on extended conditional entropy. *Knowledge-Based Systems*, 27, p.p.443-450.
27. Y.H. Qian, J.Y. Liang (2008) Combination entropy and combination granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(2), p.p.179-193.