Granular Space-Based Feature Selection and Its Applications

Lin Sun
International WIC Institute, Beijing University of Technology, Beijing 100124, P. R. China
College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, P. R. China
Engineering and Technology Research Center for Computational Intelligence and Data Mining of Universities of Henan Province, Xinxiang 453007, P. R. China
Email: linsunok@gmail.com

Jiucheng Xu, Yuwen Hu, and Lina Du
College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, P. R. China
Engineering and Technology Research Center for Computational Intelligence and Data Mining of Universities of Henan Province, Xinxiang 453007, P. R. China
Email: xjc@htu.cn

***Abstract*—Feature selection is viewed as an important preprocessing step for pattern recognition, machine learning and data mining. Considering a consistency measure introduced in rough sets, the problem of feature selection aims to retain the discriminatory power of original features. Many heuristic feature selection algorithms have been proposed, however, these methods are computationally time-consuming. This paper introduces granular space, positive granular space and negative granular space based on granular computing in simplified decision systems, and then new feature significance measure is proposed. Meanwhile, their important propositions and properties are derived. Furthermore, by virtue of radix sorting and Hash techniques, the object granules as basic processing elements are employed to investigate feature selection, and then a heuristic algorithm with low computational complexity is explored. Numerical simulation experiments show that the proposed approach is indeed efficient, and therefore of practical value to many real-world problems.**

***Index Terms*—granular computing, rough set theory, feature selection, granular space, positive granular space, negative granular space**

I. INTRODUCTION

As the capability of acquiring and storing information increases, feature selection can be viewed as one of the most fundamental problems in the fields of pattern recognition, machine learning and data mining [1]. Generally speaking, the information is usually gathered for multiple learning and mining tasks. Thus, the main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features [2]. Hence, it is useful to select parts of features to the learning algorithm in practical applications. Moreover, learning with a subset of features, rather than the whole features, will reduce the cost of acquiring and storing features, speed up learning and recognition [3, 4].

From the philosophical and theoretical points of views, it has been argued that information granulation is essential to human problem solving, and then has very significant impact on the design and implementation of intelligent system [5, 6]. Granular computing is thus a basic issue in knowledge representation and data mining. The root of granular computing comes from the concept of information granularity presented by Zadeh in the context of fuzzy set theory [7]. Granulation involves partitioning of an object into granules, with a granule being a clump of elements drawn together by indistinguishability, equivalence, similarity, proximity or functionality. An information granule formalizes the concept of finite precision representation of objects in real-life situations, and reducts represent the core of an information system (both in terms of objects and features) in a granular universe. It may be noted that cases also represent the informative and irreducible part of a problem. Hence, rough set theory is a natural choice for case selection in domains which are data rich, contain uncertainties, and allow tolerance for imprecision [8]. Thus, one can gain a better understanding of granular computing within the rough set framework [9]. The process of constructing information granules is called information granulation. It granulates a universe of discourse into a family of disjoint or overlapping granules. Then, different views can be linked together, and a hierarchy of granulations can be established. Hence, one of the main directions of granular computing is to deal with the construction, interpretation, and representation of granules.

In recent years, there are theoretical and practical reasons for the study of granular computing. Granular computing is regarded as a collective term referring to theories, methodologies, techniques, and tools for the analysis of information granules encountered in problem solving [10-12]. It explores the composition of parts, their interrelationships, and connections to the whole, and it focuses on problem solving, based on the commonsense

doi:10.4304/jsw.8.4.817-826

concepts of granule, granulated view, granularity, and hierarchy. Rough set models enable us to precisely define and analyze many notions of granular computing. For example, Yao [13] showed that the notion of levels played a fundamental role in many branches of sciences, a survey on many different interpretations and uses suggested that the concept of integrative levels of granularity might serve as a basis of granular computing, and the triarchic theory of granular computing was centered on granular structures. Zhang and Zhang [14] developed a quotient space theory of problem solving based on hierarchical description and representation of a problem. Liu et al. [15] also proposed some available models and methods of granular computing. Lin and Louie [16] presented a fast association rule algorithm based on granular computing. But in their work, generating different levels of association rules were not considered. Furthermore, how to store bit maps was not very clear. Chen et al. [17] presented a novel model called the information granulation-based data mining approach, which imitated the human ability to process information and acquired knowledge from information granules rather than from numerical data. To get a true hybrid framework or take operational decisions from data, Apolloni and Bassis [18] extended the algorithmic inference approach to the granular computing paradigm. Wu et al. [19] employed discernibility matrices and Boolean functions to determine granular consistent sets and granular reducts in contexts, proposed an approach to knowledge reduction in consistent formal decision contexts. Qiu et al. [20] discussed the information granules and application of granular computing in mining association rules from a relational database table or an information table.

Up to now, many types of feature selection in rough set theory have been proposed in the analysis of information systems and decision systems, each of them aimed at some basic requirements. Various approaches have also been developed to perform feature selection and obtain optimal true, certain, and possible decision rules in decision systems. For example, a distribution reduct is a subset of the feature set that preserved the degree to which every object belonged to each decision class [21]. A maximum distribution reduct preserves all maximum decision rules, but the degree of confidence in each uncertain decision rule may not be equal to the original one [22]. Wu et al. [23] proposed the concepts of β lower distribution reduct and β upper distribution reduct. They preserved all decision classes at some level of classification and eliminated the drawback of β-reduct, which might change decision results of some objects. Lately, Hu et al. [24] proposed methods based on positive regions, the δ neighborhood rough set model and the k-nearest-neighbor rough set model. Both are effective, and have the advantage of being able to deal with mixed features. However, their time complexities are no less than $O(|A|^2|U|^2)$, where $|A|$ and $|U|$ respectively denote the number of condition features and objects. These methods are still inefficient, and thus unsuitable for the reduction of voluminous data. Wang et al. [25] introduced two novel heuristic reduction algorithms with the time complexity $O(|A||U|^2) + O(|U|^3)$ and $O(|A|^2|U|) + O(|A||U|^3)$ respectively. Based on the mutual information, Miao and Hu [26] constructed a heuristic algorithm costing the time complexity $O(|A||U|^2) + O(|U|^3)$. Hence, the disadvantage of these methods is much space-time cost. Based on the indiscernibility relation and positive region, Liu et al. [27] proposed a complete reduction algorithm with time complexity $O(|A|^2|U|\log|U|)$ and space complexity $O(|A||U|)$. Xu et al. [28] designed a new and relatively reasonable reduction algorithm, whose worst time complexity was cut down to $\mathrm{Max}(O(|A||U|), O(|A|^2|U/A|))$. At present, the best idea of many algorithms for computing U/A is based on radix sorting, and its complexity is cut down to $O(|A||U|)$. Liu et al. [29] presented a hash-based algorithm to calculate positive region, and its time complexity decreased to $O(|U|)$. Then based on the characteristics of inconsistency, a new feature measure was introduced, and a reduction algorithm with twice-hash was presented, whose time complexity was $O(|A|^2|U/A|)$. Therefore, proposing an efficient and effective approach to feature selection for both consistent and inconsistent decision systems is very desirable. In this paper, we aim at creating such a solution to solve this problem.

The remainder of this paper is structured as follows. In Section II, we recall the basic concepts and results related to decision systems, and design a simplified decision system. In Section III, some concepts and operations of information granule are investigated, and some new concepts, such as granular space, positive granular space, negative granular space and feature significance measure, are proposed. Then some important properties and propositions are provided. Furthermore, by using radix sorting and Hash techniques, the object granules as basic processing elements are employed to investigate feature selection, and then a granular space-based heuristic feature selection algorithm with low computational complexity is designed. In Section IV, some numerical experiments from an example given and UCI datasets indicate that the proposed algorithm is efficient and of practical value in engineering. Finally, the conclusions and future work are described in Section V.

II. PRELIMINARIES

In this section, we briefly review some notions and results related to information systems and decision systems. Detailed description of concepts can be found in [4, 9, 10].

A triple (U, A, F) is called an information system (IS), where $U = \{x_1, x_2, ..., x_n\}$ is a non-empty finite set of objects, called the universe of discourse; $A = \{a_1, a_2, ..., a_m\}$ is a non-empty finite set of features; $F = \{f_a \mid \forall a \in A\}$ is a set of functions between U and A, where f_a: $U \rightarrow V_a$, for any $a \in A$, is called an information function, and the set V_a is called the value domain of the feature a.

For each subset of features $B \subseteq A$, the non-empty set determines an indiscernibility relation on U as follows: $R_B = \{(x_i, x_j) \in U \times U \mid f_a(x_i) = f_a(x_j), \forall a \in B\}$. R_B is an equivalence relation on U, and it forms a partition of U, denoted by $U/R_B = \{[x_i]_B \mid x_i \in U\}$, where $[x_i]_B = \{x_j \in U \mid (x_i, x_j) \in R_B\} = \{x_j \mid f_a(x_i) = f_a(x_j), \forall a \in B\}$ is called an equivalence class of x_i with respect to B.

The quintuple (U, A, F, D, G) is called a decision system (DS), where (U, A, F) is an information system; A is a condition feature set; D is a decision feature set with $A \cap D = \emptyset$; $G = \{g_d \mid \forall d \in D\}$, where g_d: $U \rightarrow V_d$, for any $d \in D$; and V_d is the domain of the decision feature d. So, in the DS, it should be noted that $R_D = \{(x_i, x_j) \in U \times U \mid g_d(x_i) = g_d(x_j), \forall d \in D\}$. Thus, it also determines a partition $U/R_D = \{[x]_D \mid x \in U\}$ on U.

Now we define a partial order on all partition sets of U. Let U/P and U/Q be two partitions of a finite set U, and then we define the partition U/P is finer than the partition U/Q (or the partition U/Q is coarser than the partition U/P), denoted by $P \preceq Q$ (or $Q \succeq P$), between partitions by $P \preceq Q \Leftrightarrow \forall P_i \in U/P, \exists Q_j \in U/Q \rightarrow P_i \subseteq Q_j$. If $P \preceq Q$ and $P \succeq Q$, then we say that $P = Q$. If $P \preceq Q$ and $P \neq Q$, then we say that P is strictly finer than Q (or Q is strictly coarser than P) and write $P \prec Q$ (or $Q \succ P$).

Proposition 1. Let $DS = (U, A, F, D, G)$ be a decision system. For any $P, Q \subseteq A \cup D$, if $Q \subseteq P$, then $P \preceq Q$.

Proof. The proof is similar to Proposition 1 in [10].

Let $DS = (U, A, F, D, G)$ be a decision system, and $B \subseteq A$. For any $x_i, x_j \in U$, x_i and x_j conflict with each other from B to D if and only if $f_a(x_i) = f_a(x_j)$ for any $a \in B$, and then $g_d(x_i) \neq g_d(x_j)$ for any $d \in D$. An instance $x \in U$ is a consistent instance in the DS if and only if there does not exist an instance $y \in U$, which conflicts with x from B to D. Hence, we have the conclusion that the DS is consistent if and only if each instance $x \in U$ is a consistent instance.

Proposition 2. Let $DS = (U, A, F, D, G)$ be a decision system. If there exists $R_A \subseteq R_D$, then the DS is consistent. Otherwise, it is inconsistent.

Proof. It is straightforward.

Let $X \subseteq U$, then the P-lower approximation and the P-upper approximation of set X can be denoted by $\underline{P}X = \{x \in U \mid [x]_P \subseteq X\}$ and $\overline{P}X = \{x \in U \mid [x]_P \cap X \neq \phi\}$. If P, $Q \subseteq A$ are two equivalence relations on U, then the positive and negative regions can be denoted by $POS_P(Q) = \bigcup_{X \in U/R_Q} \underline{P}X$ and $NEG_P(Q) = U - \bigcup_{X \in U/R_Q} \overline{P}X$.

Thus, in a decision system $DS = (U, A, F, D, G)$, the positive region of the partition U/R_D with respect to P, denoted by $POS_P(D)$, is the set of all objects of U that can be certainly classified to blocks of the partition U/R_D by means of P. For any $P \subseteq A$, to make $p \in P$, and p in P is unnecessary for D if $POS_P(D) = POS_{P-\{p\}}(D)$. Otherwise p is necessary. Then P is independent relative to D if every element in P is necessary for D.

Proposition 3. Let $DS = (U, A, F, D, G)$ be a decision system, and $POS_A(D) = \{x \in U \mid x$ is consistent instance$\}$. If the DS is consistent, we have $POS_A(D) = U$.

Proof. It is straightforward.

Proposition 4. Let $DS = (U, A, F, D, G)$ be a decision system with $P \subseteq A$. We then have $POS_P(D) \subseteq POS_A(D)$.

Proof. It is straightforward.

Definition 1. Let $DS = (U, A, F, D, G)$ be a decision system, and $U/(A \cup D) = \{[U'_1]_{A \cup D}, [U'_2]_{A \cup D}, ..., [U'_n]_{A \cup D}\}$, where $U = \{U_1, U_2, \ldots, U_m\}$, $n \leq m$, and $U'_i \in U$, then $U' = \{U'_1, U'_2, \ldots, U'_n\}$, F': $U' \times (A \cup D) \rightarrow V'$ is called a new information function. Then it is said that the 6-tuple (U', A, F', D, G, V') is a simplified decision system (SDS).

It follows from Definition 1 that if the DS is consistent, then U' is a simplified consistent objects set. Otherwise, U' includes a simplified inconsistent objects set. Thus, by virtue of this technology of simplicity, lots of redundancy information is deleted, and then the space complexity of decision systems is decreased. Therefore, a simplified decision system introduced is necessary.

As we know, much attention has been paid to reduction in inconsistent decision systems. For example, possible rules and possible reducts have been proposed as a means to deal with inconsistence in inconsistent decision system [22]. However, Xu [28] proposed the recursive calculating method that ignored the inconsistent objects in inconsistent decision systems.

Definition 2. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, then the positive region of the partition U'/R_D with respect to A, i.e., $POS_A(D) = [U'_{i_1}]_A \cup [U'_{i_2}]_A \cup ... \cup [U'_{i_t}]_A$, where $U'_{i_s} \in U'$, $|[U'_{i_s}]_A / D| = 1$, $s = 1, 2, \ldots, t$. Thus, there exists $U'_{POS} = \{U'_{i_1}, U'_{i_2}, ..., U'_{i_t}\}$, and we have $U'_{NEG} = U' - U'_{POS}$.

Thus, form Definitions 1 and 2, it is easy to obtain the following proposition.

Proposition 5. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system. Then $U'_{POS} \subseteq POS_A(D)$ holds.

III. Granular Space-Based Feature Selection Approach

A. Information Granule Operations and Granule Space

In this subsection, we investigate some corresponding concepts, operations and properties of information granule and granular space.

It is known that the connections between logic connectives and set-theoretic operations have been stated in [17, 18, 23]. Under those formulations, we can discuss granules in terms of intensions in a logic setting and in terms of extension in a set-theoretic setting. Thus, a definable granule is represented by a pair $Gr = (\varphi, m(\varphi))$.

Definition 3. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system. An information granule is defined as the tuple $Gr = (\varphi, m(\varphi))$, where φ refers to the intension of information granule, and $m(\varphi)$ represents the extension of information granule.

Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system with $B = \{a_1, a_2, \ldots, a_k\} \subseteq A$. Suppose that $V_{a_i} = \{V_{a_{i,1}}, V_{a_{i,2}}, ..., V_{a_{i,k}}\}$ is the domain of feature a_i, and each $V_{a_{i,j}}$ may be viewed as a concept. Then, there must exist $\varphi = \{I_1, I_2, \ldots, I_k\}$ such that $I_i \in V_{a_i}$ is a set of feature values corresponding to B. Then, the intension of an information granule can be denoted by $\varphi = \{I_1, I_2, \ldots, I_k\}$, and the extension can be denoted by $m(\varphi) = \{u \in U' \mid f(u, a_1) = I_1 \wedge f(u, a_2) = I_2 \wedge \ldots \wedge f(u, a_k) = I_k, a_i \in B, i \in \{1, 2, \ldots, k\}\}$. Here, $m(\varphi)$ describes the internal structure of the information granule. The collection of the extensions of all granules is denoted by GK, thus, we define that the map $U' \to GK$ is to construct the granulation of the SDS, and the map gs: $Gr \to U'$ is to construct the objects set. Hence, there must exist $gs(Gr) = m(\varphi)$.

Definition 4. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and $Gr = (\varphi, m(\varphi))$ be an information granule. Then its size can be defined as the cardinality of the extension of the Gr, namely, $|m(\varphi)|$. Intuitively, the size may be interpreted as the degree of abstraction or concreteness.

Definition 5. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and $Gr = (\varphi, m(\varphi))$ be an information granule. If $\varphi = \{V_{a_{i,j}}\}$, where $V_{a_{i,j}} \in V_{a_i} = \{V_{a_{i,1}}, V_{a_{i,2}}, \ldots, V_{a_{i,k}}\}$ is a categorical value, and V_{a_i} is the domain of feature $a_i \in A \bigcup D$, then Gr is called an elementary information granule of feature a_i, or an elementary granule for short. Namely, $m(\varphi) = \{u \in U' \mid f(u, a_i) = V_{a_{i,j}}, a_i \in A \bigcup D\}$.

Definition 6. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and $I = \{I_1, I_2, \ldots, I_k\}$ be a k-itemset, where $I_i \in V_{a_i}$ $(i = 1, 2, \ldots, k)$ is a feature value of feature a_i, and $B = \{a_1, a_2, \ldots, a_k\} \subseteq A$. Then, information granule $Gr = (I, m(I))$ is called a k-itemset granule, where $m(I) = \{u \in U' \mid f(u, a_1) = I_1 \wedge f(u, a_2) = I_2 \wedge \ldots \wedge f(u, a_k) = I_k, a_i \in B, i \in \{1, 2, \ldots, k\}\}$.

It should be ensured that a 1-itemset granule is an elementary granule satisfying the given conditions, and then we can obtain the following properties.

Property 1. If $I = \{I_1, I_2, \ldots, I_k\}$ is a k-itemset, then $m(I) = m(\{I_1\}) \bigcap m(\{I_2\}) \bigcap \ldots \bigcap m(\{I_k\})$.

Property 2. If $I \subseteq V'$, $J \subseteq V'$, and $I \subseteq J$, then $m(J) \subseteq m(I)$.

Definition 7. Given a k_1-itemset granule $(I, m(I))$ and a k_2-itemset granule $(J, m(J))$, if $I \subseteq J \subseteq V'$, then $m(J) \subseteq m(I)$, i.e., the intension of $(J, m(J))$ is more concrete than that of $(I, m(I))$, denoted by $m(I) \prec m(J)$. Then, all these granules lead to a hierarchical structure by using the $\prec$ order, called a multi-dimensional granular hierarchy.

Hence, through the discussions above, we can use $m(\varphi)$ to denote the all object sets which satisfy the formula φ in rough logic. In the case of formulation granule, $Gr = (\varphi, m(\varphi))$ is also the compound granules, thus, a set composed of all elementary granules, which are separated from a simplified decision system, is regarded as a granular space, simply denoted by GrS. Meanwhile, there also exists a map between the granules in GrS and U', denoted by gs: $GrS \to U'$, such that $gs(Gr) = m(\varphi)$ for any $Gr \in GrS$. In other words, there must exist $Gr \in GrS$ such that $u \in gs(Gr)$ for any $u \in U'$.

Definition 8. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. The objects set, separated from the positive region $POS_A(D)$ in the SDS, is regarded as a positive granular space, denoted by $GrSP$. The objects set, separated from the negative region $NEG_A(D)$ in the SDS, is regarded as a negative granular space, denoted by $GrSN$.

Definition 9. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space, $P \subseteq A$. For any $Gr_i \in GrSP$, if there exist $Gr_j \in GrSP$ and $Gr_i \neq Gr_j$ such that the intension of Gr_i and Gr_j has the equal sets of feature values, but they have different sets of decision values, then the granule Gr_i is called the conflict granule of $GrSP$ with respect to P. On the other hand, there exist $Gr_j \in GrSN$ and $Gr_i \neq Gr_j$ such that the intension of Gr_i and Gr_j has the equal sets of feature values, then the granule Gr_i is called the conflict granule (contradictory granule) of $GrSP$ with respect to P. Otherwise, Gr_i is called the non-conflict granule of $GrSP$ with respect to P.

B. Feature Significance Measure

Proposition 6. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. Suppose that $U'_{GrSP} = \bigcup \{gs(Gr) \mid Gr \in GrSP\}$, $U'_{GrSN} = \bigcup \{gs(Gr) \mid Gr \in GrSN\}$, then there must exist $U'_{GrSP} = U'_{POS}$ and $U'_{GrSN} = U'_{NEG}$.

Proof. It can be achieved by Definitions 2 and 8.

Definition 10. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. Suppose that $U'_{GrSP} = gs(GrS)$ or $U'_{GrSN} = \emptyset$, then the SDS is consistent. Otherwise it is inconsistent.

Proposition 7. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. Then, the SDS is consistent if and only if for any $Gr \in GrS$, there must exist $x \in gs(Gr) \wedge y \in gs(Gr) \Rightarrow f_D(x) = f_D(y)$ for any $x, y \in U'$.

Proof. It can be achieved by Proposition 6 and Definition 10.

Through the analyses of reduction algorithms in [27, 28, 30], based on positive region from the algebraic point of view, we know that feature dependency-based approaches select the next feature to add into or remove from the reduction through considering the changes of positive regions, and finally only one reduct is found. Meanwhile, the key points of these approaches focus on that how to design high efficiency measures of computing the positive region, and how to acquire the minimum reducts based on positive region. In the following, we analyze the process of reduction in a decision system. In a decision system $DS = (U, A, F, D, G)$, $P \subseteq A$, and $U/P = \{P_1, P_2, \ldots, P_n\}$, if $P_i \subseteq POS_P(D)$ for any $P_i \in U/P$, one has that $U = U - P_i$, and $POS_A(D) = POS_A(D) - P_i$. Thus, it helps to reduce the quantity of computation, time complexity and space complexity of search. In this case, if $P_i \not\subset POS_P(D)$, we then use $U/(P \bigcup \{a\}) = \bigcup \{X/\{a\} \mid X \in U/P\}$ to form a new partition of U again, and if $P_i \subseteq U - POS_C(D)$, then P_i is deleted from U, because it contributes nothing to computing the positive region of the *DS*. Therefore, we have the conclusion that the methods above only consider the information, which can distinguish one class from the remaining classes, and then omit that of distinguishing one class from another class. To address this problem, we investigate some definitions and methods to judge a granular space-based reduct in a simplified decision system.

Let $DS = (U, A, F, D, G)$ be a decision system with any $P \subseteq A$. Then, a positive region-based reduct of the *DS* is presented in [28] as follows: if $POS_P(D) = POS_A(D)$ and P is independent relative to D, then P is a reduct of the *DS*.

Definition 11. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. For any $P \subseteq A$, then the positive region of the partition U'/R_D with respect to P in GrS is defined as

$$POS'_P(D) = \bigcup\{X \mid X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}\}. \quad (1)$$

From the definition of positive region-based reduct and Definition 11, it is easy to obtain the following proposition.

Proposition 8. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. For any $P \subseteq A$, if there does not exist any conflict granule with respect to P in $GrSP$, then P is a relative reduct of A with respect to D.

Proposition 9. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. For any $P \subseteq A$, and $p \in A - P$, we obtain a granular partition of U' for any $Gr \in GrS$, denoted by

$$gs(Gr)/R_{P\cup\{p\}} = \bigcup\{X/R_{\{p\}} \mid X \in gs(Gr)/R_P\}. \quad (2)$$

Proof. It can be achieved under Property 2 in [27].

Definition 12. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GrS be its granular space. For any $P \subseteq A$, and $p \in A - P$, the feature significance measure of p in P is defined as

$$SIG_P(p) = \frac{|U'_{P\cup\{p\}} - U'_P|}{|U'|}, \quad (3)$$

where $U'_P = \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}} X\} \cup \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSN}} X\}$.

Noticing that if $P = A$, then we have that $SIG_P(\varnothing) = 0$.

From Definition 12, it can be seen that if the feature significance measure is used as a heuristic function, then the reduction is a kind of algebraic reduct from the granule concept, and the granular space GrS must be constructed. Thus, to reduce the quantity of computation and space complexity of search, the granular space GrS is separated into $GrSP$ and $GrSN$. From the algebraic point of view in [30], this idea can compensate for these current disadvantages of the classical reduction algorithms, based on positive region and information entropy. This shows that feature selection has already been studied from the algebra viewpoint and information viewpoint of rough set theory, respectively. However, the concepts of feature selection, based on these two different viewpoints, are not equivalent to each other. In [30], the relationship between these conceptions from the two viewpoints was rather an inclusion than an equivalence, due to the fact that the rough set theory discussed from the information point of view restricted features and decision systems more specifically than it did when considered from the algebra point of view. Hence, the identity of the two viewpoints will hold in consistent decision systems only. That is, the algebra viewpoint and information viewpoint are equivalent in a consistent decision system, while different in an inconsistent one. Thus it is concluded that the algebraic point of view proposed based on granular computing can not only include more information than that based on positive region, but also compensate for the limitations of the algebra viewpoint and information viewpoint. The results are significant for the design and development of methods for feature selection.

In what follows, to reduce space complexity of search, and construct an efficient reduction algorithm, we further derive some important properties of feature significance measure in simplified decision systems.

Proposition 10. Let $SDS = (U', A, F', D, G, V')$ be a simplified decision system, and GS be its granular space, $P \subseteq A$. Then, the simplified expression of the feature significance measure of $p \in A - P$ in P is as follows

$$SIG_P(p) = \frac{|U'_{P\cup\{p\}} - U'_P|}{|U'|} = \frac{\left|\bigcup_{X \in (U' - GrS_P)/R_P} \{\{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\} \cup \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\}\right|}{|U'|}.$$

Proof. Form Definition 12, we can obtained that

$$\begin{aligned}
SIG_P(p) &= \frac{|U'_{P\cup\{p\}} - U'_P|}{|U'|} \\
&= \frac{1}{|U'|}\Big|\bigcup_{X \in gs(GrS)/R_P} \{\{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\} \cup \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\} \\
&\quad - \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}} X\} \cup \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSN}} X\}\Big| \\
&= \frac{1}{|U'|}\Big|\bigcup_{X \in gs(GrS)/R_P} \{\{\bigcup_{X \subseteq U'_{GrSP}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\}\} \cup \\
&\quad \{\bigcup_{X \not\subseteq U'_{GrSP}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\}\}\} \cup \{\{\bigcup_{X \subseteq U'_{GrSN}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\} \\
&\quad \cup \{\bigcup_{X \not\subseteq U'_{GrSN}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\}\} - \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}} X\} \\
&\quad \cup \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSN}} X\}\Big| \\
&= \frac{1}{|U'|}\Big|\bigcup_{X \in gs(GrS)/R_P} \{\{\bigcup_{X \subseteq U'_{GrSP}} X\} \cup \{\bigcup_{X \not\subseteq U'_{GrSP}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\}\} \\
&\quad \cup \{\bigcup_{X \subseteq U'_{GrSN}} X\} \cup \{\bigcup_{X \not\subseteq U'_{GrSN}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\}\} \\
&\quad - \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}} X\} \cup \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSN}} X\}\Big| \\
&= \frac{1}{|U'|}\Big|\{\bigcup_{X \in gs(GrS)/R_P} \{\{\bigcup_{X \subseteq U'_{GrSP}} X\} \cup \{\bigcup_{X \subseteq U'_{GrSN}} X\}\}\} \cup \\
&\quad \bigcup_{X \in gs(GrS)/R_P} \{\{\bigcup_{X \not\subseteq U'_{GrSP}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSP}} Y\}\} \\
&\quad \cup \{\bigcup_{X \not\subseteq U'_{GrSN}} \{\bigcup_{Y \in X/R_{\{p\}} \wedge Y \subseteq U'_{GrSN}} Y\}\}\} - \{\bigcup_{X \in gs(GrS)/R_P \wedge X \subseteq U'_{GrSP}} X\}
\end{aligned}$$

$$
\begin{aligned}
&\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\subseteq U'_{GrSN}} X\}| \\
&=\frac{1}{|U'|}|\{\{\bigcup_{X\in gs(GrS)/R_P\wedge X\subseteq U'_{GrSP}} X\}\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\subseteq U'_{GrSN}} X\}\} \\
&\quad\cup\{\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\}\cup \\
&\quad\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}\}- \\
&\quad\{\bigcup_{X\in gs(GrS)/R_P\wedge X\subseteq U'_{GrSP}} X\}\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\subseteq U'_{GrSN}} X\}| \\
&=\frac{1}{|U'|}|\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\}\cup \\
&\quad\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}| \\
&=\frac{1}{|U'|}|\{\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\} \\
&\quad\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\}\} \\
&\quad\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}| \\
&=\frac{1}{|U'|}|\{\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\}\} \\
&\quad\cup\{\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSN}}\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}| \\
&=\frac{1}{|U'|}|\bigcup_{X\in gs(GrS)/R_P\wedge X\not\subset U'_{GrSP}\wedge X\not\subset U'_{GrSN}}\{\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\cup \\
&\quad\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}| \\
&=\frac{|\bigcup_{X\in(U'-GrS_P)/R_P}\{\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSP}} Y\}\cup\{\bigcup_{Y\in X/R_{\{p\}}\wedge Y\subseteq U'_{GrSN}} Y\}\}|}{|U'|}.
\end{aligned}
$$

It can be observed that since $X\in gs(S)/P\wedge X\not\subset U'_{GrSP}\wedge X\not\subset U'_{GrSN}$ in the above formula, there must exist $X\notin U'_p$ such that $X\in(U'-U'_p)/P$. Hence, the theorem holds.

Obviously, it shows from Proposition 10 that through a series of formulary transform, and then it finally only need compute some partitions and sets. So, we can make full use of some efficient measures to compute partitions and sets for the feature significance measure. Therefore, this computation of $SIG_P(p)$ is more direct and efficient, and it can reduce the quantity of computation, time complexity, and space complexity of search.

It is well known that a reduct can be found by deleting current redundant features from the condition feature set. However, the reduct obtained depends on the order that the features are considered. Hence, in a simplified decision system $SDS = (U', A, F', D, G, V')$, it is important to construct an appropriate input sequence for the features in A. We can use the concept of feature significance measure to solve this problem. That is, we obtain the input sequence by sorting the features in descending order by the feature significance measure. Thus, the feature significance measure, $SIG_P(p)$ of $p\in A-P$ in P, is usually defined in terms of dependency degree. Meanwhile, $SIG_P(p)$ is called a heuristic function from the algebraic point of view in the *SDS*, and then we assume that $P = \emptyset$, adding features bottom-up on P. Thus, if a new corresponding granule $Gr = (\varphi, m(\varphi))$ separated from the *SDS*, is a non-conflict granule with respect to A, and contained by *GrSP*, or a conflict granule with respect to A, is contained by *GrSN*. Then, we have $SIG_P(p) \neq 0$. That is to say, when we continually add any feature p to P given, if the radix conflict granule of *GrSP* with respect to $P\cup\{p\}$ is not changed, then $SIG_P(p) \neq 0$. Hence, $SIG_P(p)$ describes a decrease in the dependency degree of P. Thus, Proposition 10 and the properties above provide an effective way to determine whether or not a feature should be contained in a granular space-based reduct.

C. Feature Selection Algorithm

It is known that introducing heuristic search is to solve the problem of NP-hard, and is also greatly effective and feasible in acquiring the minimal or optimal reducts. However, the shortcoming of classical methods is its high space-time cost and incompleteness. In order to improve computational efficiency, the complicated problem of selecting feature subset can be decomposed into several relatively independent sub-problems. In the following, we derive sub-algorithm for feature selection, and then present a complete and efficient algorithm for calculating reducts in a simplified decision system. What's more, to improve efficiency, we must employ some effective measures for computing partitions, positive region, and separating the decision system to acquire its granular space. Then, based on twice-hash in [29], the time complexity of computing partitions, positive region and negative region object sets based on Hash has been decreased to $O(|U|)$. Thus, we can make full use of the feasible measures above, and construct the processes of acquiring positive and negative granular spaces of decision systems, which consists of the following steps.

Algorithm 1

Input: A decision system $DS = (U, A, F, D, G)$, where $A = \{a_1, a_2, \ldots, a_n\}$, and $D=\{d\}$

Output: *GrSP*, *GrSN*.

(1) Calculate the partition $U/(A\cup D)$ to obtain U', i.e., $SDS = (U', A, F', D, G, V')$

(2) Initialize $GrSP = \emptyset$, $GrSN = \emptyset$, $m = |U'/A|$, and $n = |A|$

(3) For $i = 1$ to m begin

(3.1) Calculate the partition $[U'_i]_A/D$, where $[U'_i]_A \in U'/A$

(3.2) $k = |[U'_i]_A/D|$

(3.3) For $j = 1$ to $|[U'_i]_A|$ begin

(3.3.1) Fetch $x_j \in [U'_i]_A \in U'/A$

(3.3.2) Calculate its corresponding granule $Gr = (\varphi, m(\varphi))$, where $\varphi = \{I_1, I_2, \ldots, I_n\}$

(3.3.3) $m(\varphi) = \{u\in U' \mid f(u, a_1) = I_1 \wedge f(u, a_2) = I_2 \wedge \ldots \wedge f(u, a_n) = I_n, a_t\in A, t \leq n$

(3.3.4) If $k = 1$ then $GrSP = GrSP\cup\{Gr\}$, i.e., Gr is called the non-conflict granule with respect to A

else $GrSN = GrSN \bigcup \{Gr\}$, i.e., Gr is called the conflict granule of with respect to A

(4) Return $GrSP$, $GrSN$

Evidently, the time complexity of Algorithm 1 is determined by step (3). When $|A|$ is the maximum value for the circle times, the time complexity of Algorithm 1 is approximate to $O(|U'/A||U'|)$, while that of step (1) is $O(|U|)$. Hence, in the DS, the time complexity of Algorithm 1 is obviously no more than $O(|U| + |U'/A||U'|)$.

Using Algorithm 1, we construct the processes of granular space-based feature selection algorithm (also called GSFSA), which consists of the following steps: firstly, detaching objects from granular space step by step; then obtaining the minimum relative reducts through adding features bottom-up in terms of dependency degree of feature significance measure; finally, checking the completeness of reducts through gradually deleting feature from the end to the beginning. Obviously, we construct the input sequence by sorting the features in order of descending the feature significance measure. The relative reduct can be found by repeatedly adding the head node with the maximum $SIG_P(p)$ as follows.

Algorithm 2 GSFSA

Input: $SDS = (U', A, F', D, G, V')$

Output: Set P of selected features

(1) Calculate $GrSP$ and $GrSN$ with Algorithm 1

(2) Initialize $P = \varnothing$, $G_P = \varnothing$, and $G_N = \varnothing$

(3) Calculate U'_P, $(U' - U'_P)/P$, $(U' - U'_P)/(P \bigcup \{p\})$ with $p \in A - P$

(4) Calculate $G_P = \bigcup_{X \in (U'-U'_P)/P} \{ \bigcup_{Y \in X/\{a\} \wedge Y \subseteq U'_{GS_P}} Y\}$, and $G_N = \bigcup_{X \in (U'-U'_P)/P} \{ \bigcup_{Y \in X/\{a\} \wedge Y \subseteq U'_{GS_N}} Y\}$

(5) Calculate $SIG_P(p) = \dfrac{|G_P \bigcup G_N|}{|U'|}$ for any $p \in A - P$

(6) Select a feature a with the maximum of $SIG_P(p)$

(7) If this feature is not only
 then select that with the maximum of $|U'/(P \bigcup \{p\})|$
 else select the front

(8) $P = P \bigcup \{p\}$

(9) $U' = U' - G_P - G_N$, i.e., $U' = U' - U'_P$, to reduce the search space

(10) If $U' \neq \varnothing$ then turn to (3)

(11) $s = |P|$

(12) For $i = 1$ to s begin

 (12.1) Fetch $p_i \in P$ from the end to the beginning of P

 (12.2) If $|POS_A(D)| = |POS_{P-\{p_i\}}(D)|$, i.e., $SIG_P(p) = 0$, then $P = P - \{p_i\}$

(13) Return P

By using this algorithm GSFSA, the time complexity of feature selection from a decision system is polynomial. At step (1), the time complexity is $O(|U| + |U'/A||U'|)$, and that of step (3) through step (5) is $O(|A - P||U' - U'_P|)$. In fact, the most time-consuming task is determined by step (3) through step (10), thus, the time complexity is approximate to $O(|A||U'|) + O((|A| - 1)|U' - U'_{P_1}|) + O((|A| - 2)|U' - U'_{P_2}|) + \ldots + O(|A - P_k||U' - U'_{P_k}|)$, where P_k is a reduct of the SDS. Furthermore, the aim of step (12) is to ensure the completeness of GSFSA, therefore, its time complexity is approximate to $O(|A||U'|)$. Obviously, the total time complexity of GSFSA is $O(|U| + |U'/A||U'|) + O(|A - P||U' - U'_P|) + O(|A||U'|) + O((|A| - 1)|U' - U'_{P_1}|) + O((|A| - 2)|U' - U'_{P_2}|) + \ldots + O(|A - P_k||U' - U'_{P_k}|) + O(|A||U'|)$. Therefore, when $|A|$ is the maximum value for the circle times, the time complexity of GSFSA is approximate to $O(|A||U'|) + O((|A| - 1)|U'|) + O((|A| - 2)|U'|) + \ldots + O(|U'|) = O(|A|^2|U'|)$, that is, $O(|A|^2|U/A|)$, and its worst space complexity is $O(|A||U|)$. Thus, this means that the algorithm for feature selection requires less computation and memory.

IV. Applications and Experimental Results

In this section, we shall demonstrate the performances of our feature selection algorithm given in Subsection C above, and the objective is to evaluate the algorithm in terms of number of selected features and running time on selected features. These experiments are performed on a personal computer with Windows XP, Intel(R) Core(TM) Quad CPU 3.1 GHz, and 4 GB memory. Here, we first introduce an example of decision system to illustrate the basic method proposed.

Example 1. We adopt the decision system (U, A, F, D, G) shown in Table I from [28], where $U = \{x_1, x_2, ..., x_{15}\}$, $A = \{a, b, c, d\}$.

TABLE I.
A Decision Information System

U	a	b	c	d	D
x_1	1	1	1	1	0
x_2	2	2	2	1	1
x_3	1	1	1	1	0
x_4	2	3	2	3	0
x_5	2	2	2	1	1
x_6	3	1	2	1	0
x_7	1	2	3	2	2
x_8	2	3	1	2	3
x_9	3	1	2	1	1
x_{10}	1	2	3	2	2
x_{11}	3	1	2	1	1
x_{12}	2	3	1	2	3
x_{13}	4	3	4	2	1
x_{14}	1	2	3	2	3
x_{15}	4	3	4	2	2

TABLE II.
A Simplified Decision Information System

U'	a	b	c	d	D
x_1	1	1	1	1	0
x_2	2	2	2	1	1
x_4	2	3	2	3	0
x_6	3	1	2	1	0
x_7	1	2	3	2	2
x_8	2	3	1	2	3
x_9	3	1	2	1	1
x_{13}	4	3	4	2	1
x_{14}	1	2	3	2	3
x_{15}	4	3	4	2	2

Then, we obtain the simplified decision system showed in Table II. Using the feature significance measure proposed, we can compute the significance and conditional significance of single feature and select the features with maximal significance and conditional significance one by one, then we have the minimum relative reduct $\{a, d\}$.

Next is the second part of our experiments. We compared the performance of GSFSA with two algorithms on one discrete UCI dataset Mushroom (UCI datasets can be downloaded at http://www.ics.uci.edu). In the current experiment, GSFSA is compared with Algorithm 4 in [27] and ReduceBasedSig( ) Algorithm in [28]. The experimental results are summarized in Fig. 1, in which the judged instances with the decrease of selected feature subsets on Mushroom are given. From Fig. 1, we can see that the performances of GSFSA and ReduceBasedSig( ) Algorithm are very close though GSFSA performs a little better than Algorithm 4 in [27].

In what follows, we choose three data sets from the UCI datasets, outlined in Table III, which are used to do the final experiment. Meanwhile, we select the MIBARK Algorithm [26], the ReduceBasedSig( ) Algorithm [28], and Algorithm with twice-hash [29] in comparison with the proposed GSFSA in this article, shortly denoted by Alg_a, Alg_b, Alg_c, and Alg_d, respectively. Their reduction results and running times are compared, thus, we run all four algorithms with ten times for the corresponding experimental data, however, for the precision of analysis, we delete the beginning and the end of experimental results, and take the average of the rest of eight experimental data. What's more, in Table III, m and t respectively denote the number of reducts and running times, expressed in millisecond. Then, it can be seen from Table III that the better performances of the proposed approach can be established and applied to both consistent and inconsistent decision systems.

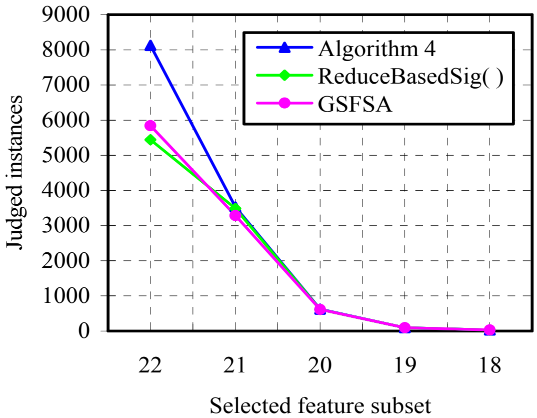

Figure 1. Selected features versus judged instances

TABLE III.
REDUCTS AND RUNNING TIMES OF FEATURE SELECTION

Data sets	Samples	Features	Data type	Alg_a		Alg_b		Alg_c		Alg_d	
				m	*t*	*m*	*t*	*m*	*t*	*m*	*t*
Voting-records	435	16	consistent	10	500	8	400	8	180	8	180
Zoo	101	17	inconsistent	11	350	10	120	10	80	10	80
Mushroom	8124	22	consistent	5	8700	4	6400	4	2450	4	2400

V. CONCLUSIONS AND FUTURE WORK

As an effective approach to feature selection, rough set has been one of the most advanced areas popularizing granular computing, and it has received extensive application in various fields [31-34]. Although a few algorithms for dealing with decision systems have been proposed, the inclusion of irrelevant, redundant and noisy features can result in poor predictive performance and increase computation cost, so that their complexities are always no less than $O(|A|^2|U|^2)$. Thus, they are unsuitable for large data sets, greatly limiting potential applications. Hence, we investigated the components of information granule, granular space, positive granular space and negative granular space in decision systems. Then an efficient algorithm for feature selection, costing the worst time complexity $O(|A|^2|U/A|)$, was proposed in a decision system. Finally, numerical experiments from the UCI datasets illustrated that the proposed algorithm was indeed effective and efficient. The experiment results were consistent with our theoretical analysis. In sum, the proposed approach to feature selection based on granular space is feasible, and outperforms other approaches available to feature selection in decision systems, especially large scale data sets. In the future work, more experimentation and further investigation into this technique may be required, and then we will extend the granular computing method to some more kinds of information systems, such as ordered information systems, incomplete decision systems [35, 36] and so on.

ACKNOWLEDGMENT

We are highly grateful to the anonymous reviewers, referees and Editor-in-Chief for their valuable comments and hard work.

This work was supported by the National Natural Science Foundation of China under Grant (Nos. 60873104, 61040037, 60905027), the Key Project of Science and Technology Department of Henan Province (No. 112102210194), the Science and Technology Research Key Project of Educational Department of Henan Province (No. 12A520027), Beijing Natural Science Foundation (No. 4102007), the CAS/SAFEA International Partnership Program for Creative Research Teams, the Education Fund for Youth Key Teachers of Henan Normal University, and the Postgraduate Science and Technology Program of Beijing University of Technology (No. ykj-2012-6765).

REFERENCES

[1] L. Sun, J. C. Xu, S. Q. Li, X. Z. Cao, and Y. P. Gao, "New Approach for Feature Selection by Using Information Entropy", *Journal of Information and Computational Science*, vol.8, pp.2259-2268, 2011.

[2] L. Sun, J. C. Xu, and Y. Tian, "Feature Selection Using Rough Entropy-Based Uncertainty Measures in Incomplete Decision Systems", *Knowledge-Based Systems*, vol.36, pp.206-216, 2012.

[3] Q. H. Hu, J. F. Liu, and D. R. Yu, "Mixed Feature Selection Based on Granulation and Approximation", *Knowledge-Based System*, vol.21, pp.294-304, 2008.

[4] L. Sun, J. C. Xu, Z. A. Xue, and L. J. Zhang, "Rough Entropy-Based Feature Selection and Its Application", *Journal of Information and Computational Science*, vol. 8, pp.1525-1532, 2011.

[5] L. Sun, J. C. Xu, C. Wang, T. H. Xu, and J. Y. Ren, "Granular Computing-Based Granular Structure Model and Its Application in Knowledge Retrieval", *Information Technology Journal*, vol.11, 1714-1721, 2012.

[6] L. Sun, J. C. Xu, J. Y. Ren, and T. H. Xu. "Granularity Partition-Based Feature Selection and Its Application in Decision Systems", *Journal of Information and Computational Science*, vol.9, pp.3487-3500, 2012.

[7] A. Bargiela and W. Pedrycz, *Granular Computing: An Introduction*. Kluwer Acad. Publishers, 2002.

[8] K. P. Sankar and M. Pabitra. "Case Generation Using Rough Sets with Fuzzy Representation", *IEEE Transactions on Knowledge and Data Engineering*, vol.16, pp.292-300, 2004.

[9] D. Q. Miao, G. Y. Wang, Q. Liu, T. Y. Lin, and Y. Y. Yao, *Granular Computing: Past, Present, and The Future Perspectives*. Beijing: Acad. Press, 2007.

[10] L. Sun, J. C. Xu, Z. A. Xue, and J. Y. Ren, "Decision Degree-Based Decision Tree Technology for Rule Extraction", *Journal of Computers*, vol.7, pp.1769-1779, 2012.

[11] X. Y. Li, J. X. Sun, G. H. Gao, and J. H. Fu, "Research of Hierarchical Clustering Based on Dynamic Granular Computing", *Journal of Computers*, vol.6, pp.2526-2533, 2011.

[12] L. Sun, J. C. Xu, and Y. P. Song, "Information Quantity-Based Decision Rule Acquisition from Decision Tables", *Journal of Convergence Information Technology*, vol.7, no.2, pp.57-67, 2012.

[13] Y. Y. Yao, "Integrative Levels of Granularity", in Human Centric Information Processing Through Granular Modeling, A. Bargiela and W. Pedrycz, Eds. Springer-Verlag, 2009, pp.31-47.

[14] L. Zhang and B. Zhang, "The Quotient Space Theory of Problem Solving", *Fundamenta Informaticae*, vol.59, pp.287-298, 2004.

[15] Q. Liu, H. Sun, and Y. Wang, "Granulations Based on Semantics of Rough Logical Formulas and Its Reasoning", in Proceedings of RSFDGrC2007, Lecture Notes in Artificial Intelligence 4482, 2007, pp.419-426.

[16] T. Y. Lin and E. Louie, "Finding Association Rules by Granular Computing: Fast Algorithms for Finding Association Rules", in Proceedings of DMRSGC2002, Berlin, German, 2002, pp.23-42.

[17] M. C. Chen, L. S. Chen, C. C. Hsu, and W. R. Zeng, "An Information Granulation Based Data Mining Approach for Classifying Imbalanced Data", *Information Sciences*, vol.178, pp.3214-3227, 2008.

[18] B. Apolloni and S. Bassis, "Algorithmic Inference: from Information Granules to Subtending Functions", *Nonlinear Analysis: Hybrid Systems*, vol.2, no.2, pp.665-683, 2008.

[19] W. Z. Wu, Y. Leung, and J. S. Mi, "Granular Computing and Knowledge Reduction in Formal Contexts", *IEEE Transactions on Knowledge and Data Engineering*, vol.21, pp.1461-1474, 2009.

[20] T. R. Qiu, Q. Liu, and H. K. Huang, "A Granular Computing Approach to Knowledge Discovery in Relational Databases", *Acta Automatica Sinica*, vol. 35, no.8, pp.1071-1079, 2009.

[21] M. Beynon, "Reducts within The Variable Precision Rough Sets Model: A Further Investigation", *European Journal of Operational Research*, vol.134, no.3, pp.592-605, 2001.

[22] M. Kryszkiewicz, "Comparative Study of Alternative Types of Knowledge Reduction in Insistent Systems", *International Journal of Intelligent Systems*, vol.16, pp.105-120, 2001.

[23] W. Z. Wu, M. Zhang, H. Z. Li, and J. S. Mi, "Knowledge Reduction in Random Information Systems via Dempster-Shafer Theory of Evidence", *Information Sciences*, vol.174, pp.143-164, 2005.

[24] Q. H. Hu, D. R. Yu, and Z. X. Xie, "Neighborhood Classifiers", *Expert Systems with Applications*, vol.34, pp.866-876, 2008.

[25] G. Y. Wang, H. Yu, and D. C. Yang, "Decision Table Reduction Based on Conditional Information Entropy", *Chinese Journal of Computers*, vol.25, pp.759-766, 2002.

[26] D. Q. Miao and G. R. Hu, "A Heuristic Algorithm for Reduction of Knowledge", *Chinese Journal of Computer Research and Development*, vol.36, pp.681-684, 1999.

[27] S. H. Liu, Q. J. Sheng, B. Wu, Z. Z. Zhi, and F. Hu, "Research on Efficient Algorithms for Rough Set Methods", *Chinese Journal of Computers*, vol.26, pp.524-529, 2003.

[28] Z. Y. Xu, Z. P. Liu, B. R. Yang, and W. Song, "A Quick Attribute Reduction Algorithm with Complexity of $\max(O(|C||U|),O(|C|^2|U/C|))$", *Chinese Journal of Computers*, vol.29, pp.391-399, 2006.

[29] Y. Liu, R. Xiong, and J. Chu, "Quick Attribute Reduction Algorithm with Hash", *Chinese Journal of Computers*, vol.32, pp.1493-1499, 2009.

[30] G. Y. Wang, J. Zhao, J. J. An, and Y. Wu, "A Comparative Study of Algebra Viewpoint and Information Viewpoint in Attribute Reduction", *Fundamenta Informaticae*, vol.68, no.3, pp.289-301, 2005.

[31] L. Sun, J. C. Xu, and S. Q. Li, "Knowledge Reduction Based on Granular Computing from Decision Information Systems", in Proceedings of RSKT2010, Lecture Notes in Artificial Intelligence 6401, 2010, pp.46-53.
[32] J. C. Xu and L. Sun, "Knowledge Reduction and Its Rough Entropy Representation of Decision Tables in Rough Set", In Proceedings of Grc2007, Silicon Valley, California, 2007, pp.249-252.
[33] J. C. Xu and L. Sun, "A New Knowledge Reduction Algorithm Based on Decision Power in Rough Set", *Transactions on Rough Sets*, vol.12, pp.76-89, 2010.
[34] J. C. Xu and L. Sun, "Research of Knowledge Reduction Based on New Conditional Entropy", in Proceedings of RSKT2009, Lecture Notes in Computer Science 5589, 2009, pp.144-151.
[35] L. Sun, J. C. Xu, S. Q. Li, and Y. P. Gao, "Rough Entropy Extensions for Feature Selection under Incomplete Decision Information Systems", *Advances in Information Sciences and Service Sciences*, vol.3, no.11, pp.264-274, 2011.
[36] L. Sun, J. C. Xu, and L. J. Zhang, "Approaches to Knowledge Reduction of Decision Systems Based on Conditional Rough Entropy", *International Journal of Advancements in Computing Technology*, vol.3, no.9, pp.129-139, 2011.

Lin Sun was born in 1979. He received his B.S. and M.S. degrees in Computer Science and Technology from Henan Normal University, Henan, China, in 2003, and 2007, respectively. Currently, he is a Ph.D. candidate from 2010, in School of Electronic Information & Control Engineering, Beijing University of Technology. His main research interests include granular computing, rough set, data mining, intelligent information processing, web intelligent.

Jiucheng Xu was born in 1964. He received his B.S. degree in Mathematics from Henan Normal University, Henan, China, in 1986, the M.S. and the Ph.D. degrees in Computer Science and Technology from Xi'an Jiaotong University, Shanxi, China, in 1995, and 2004, respectively. Currently, he is a Professor in the College of Computer & Information Engineering, Henan Normal University. His main research interests include granular computing, rough set, data mining, intelligent information processing, image retrieval.