山西大學

Uncertainty and Feature Selection in Rough Set Theory

Jiye Liang

School of Computer and Information Technology,
Shanxi University, Taiyuan, 030006, Shanxi, China
ljy@sxu.edu.cn

Contents

1. Uncertainty in Rough Set Theory

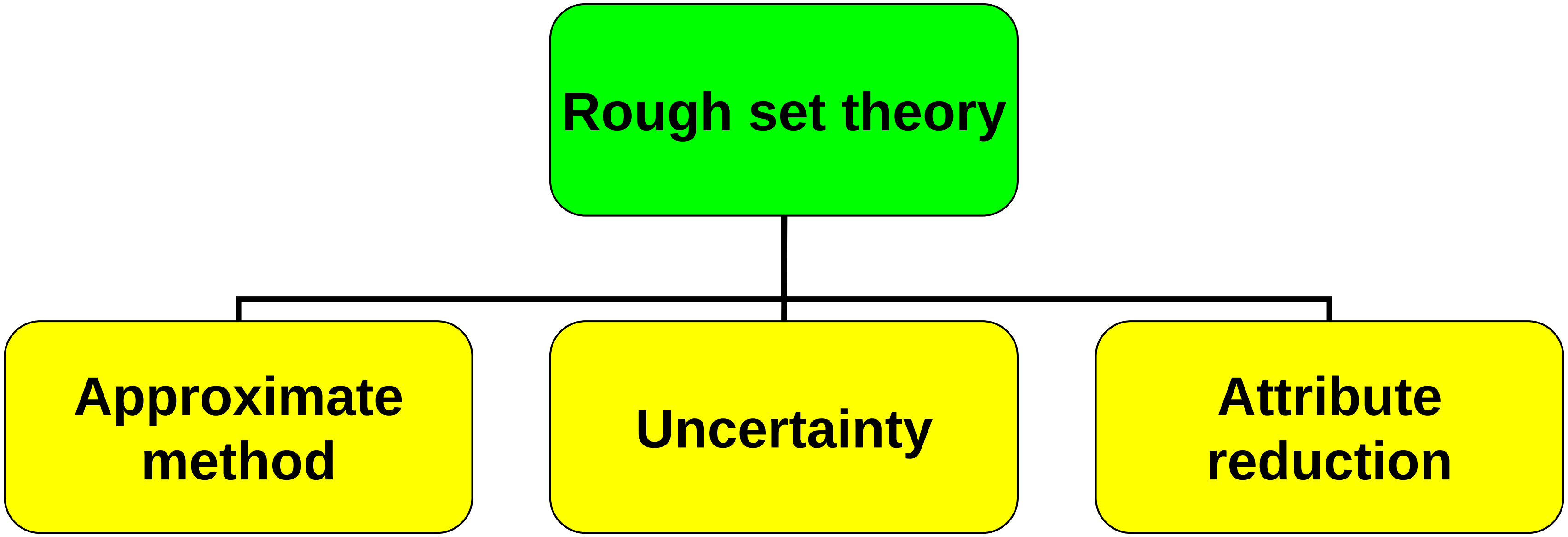
2. Accelerator of Feature Selection

3. Conclusion and Further Work

1. Uncertainty in Rough Set Theory

Key Issues in Rough Set Theory

Rough set theory is a usual soft computing tool for dealing with imprecise, uncertain, and vague information.



Uncertainty

Concepts

Certainty is regular, certain, crisp and exact properties in the development process of objective things.

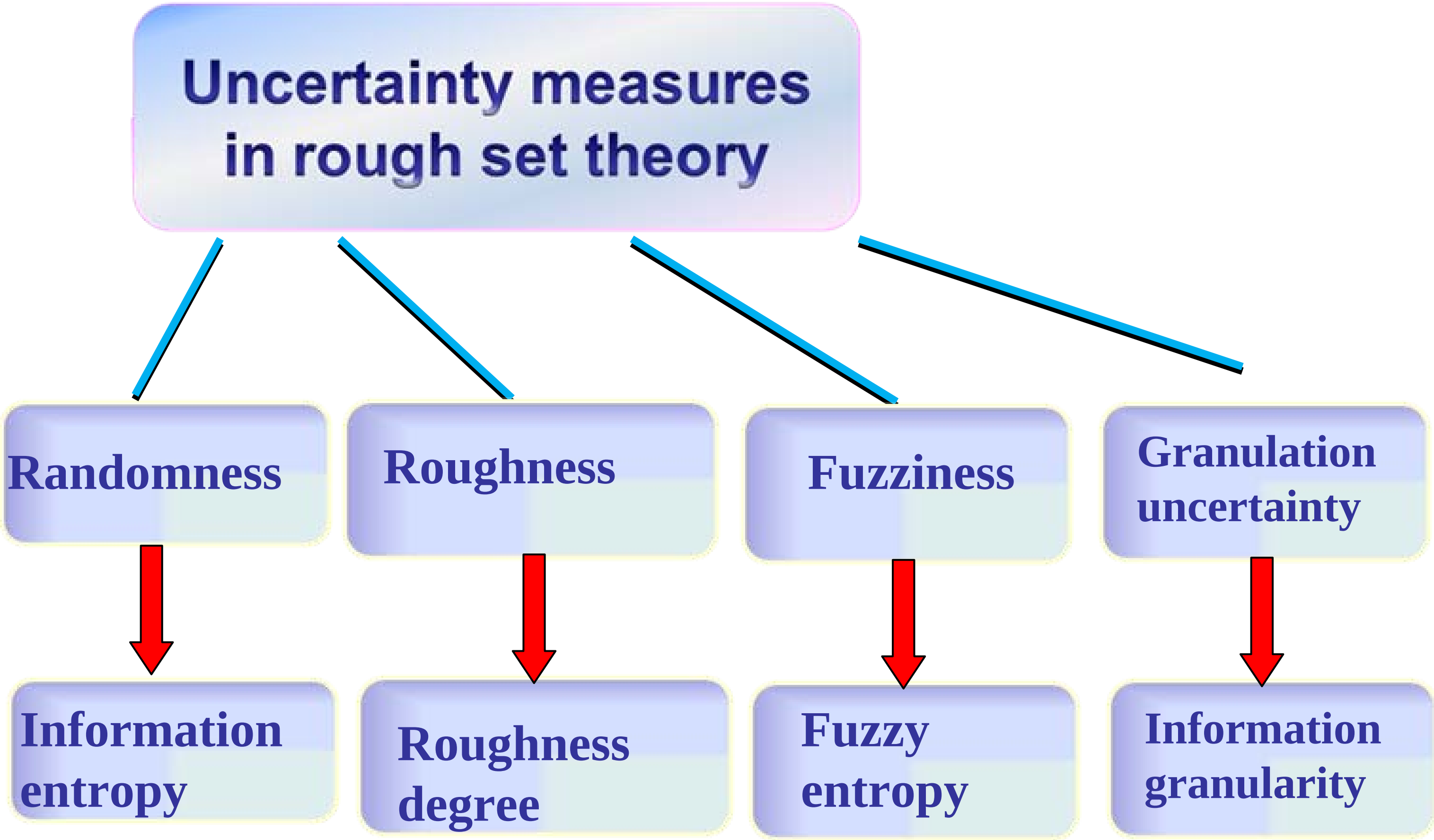
Uncertainty is unordered, casual, fuzzy and approximate properties in the development process of objective things.

Uncertainty from Four Views

- **Randomness** is an uncertainty caused by that the condition can not determine the result.
- **Fuzziness** is an uncertainty caused by the unclearness of object's classification.
- **Roughness** refers to the uncertainty of concept approximation in rough set theory, which is caused by the inequality between the upper approximation and the lower approximation.
- **Granulation uncertainty** argues that a cognitive subject is uncertain on the current information granulation.

Randomness and **fuzziness** are two basic natures of uncertainty.

Three Types of Measures



Information Entropy

- **Shannon's entropy**

Let $S = (U, A)$ be an information system, $U / A = \{X_1, X_2, \ldots, X_n\}$ is a partition on U and $p_i = p(X_i) = \frac{|X_i|}{|U|}$.

$$H(A) = -\sum_{i=1}^{n} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|}$$

Complementary entropy

$$E(A)=\sum_{i=1}^{n}\frac{|X_i|}{|U|}\frac{|X_i^c|}{|U|}=\sum_{i=1}^{n}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)$$

where X_i^c *is the completment set of* X_i .

Combination entropy

$$CE(A)=\sum_{i=1}^{n}\frac{|X_i|}{|U|}\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}$$

Where $\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}$ denotes the probability of pairs of the elements which are distinguishable each other within the whole number of pairs of the elements on the universe.

Information Entropy's Applications

The above three measures of randomness can be used to measure the significance of attributes in an information system.

- $Sig_H(a, A) = H(A) - H(A \cup \{a\})$
- $Sig_E(a, A) = E(A) - E(A \cup \{a\})$
- $Sig_{CE}(a, A) = CE(A) - CE(A \cup \{a\})$

Conditional Entropy

◆ Conditional entropy in decision tables

➢ **Shannon's conditional entropy**

$$H(D|C) = -\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{i=1}^{n}\frac{|X_i \cap Y_j|}{|X_i|}\log\frac{|X_i \cap Y_j|}{|X_i|}$$

➢ **Complementary conditional entropy**

$$E(D|C) = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|Y_j \cap X_i|}{|U|}\frac{|Y_j^c - X_i^c|}{|U|}$$

➢ **Combination conditional entropy**

$$CE(D|C) = \sum_{i=1}^{n}\left(\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2} - \sum_{i=1}^{n}\frac{|X_i \cap Y_j|}{|U|}\frac{C_{|X_i \cap Y_j|}^2}{C_{|U|}^2}\right)$$

Conditional Entropy's Applications

The above three measures of randomness can be used to define the significance of attributes in a decision table, which are as follows.

- $Sig_H(a,C,D) = H(D|C) - H(D|C \cup \{a\})$
- $Sig_E(a,C,D) = E(D|C) - E(D|C \cup \{a\})$
- $Sig_{CE}(a,C,D) = CE(D|C) - CE(D|C \cup \{a\})$

Roughness

The roughness of a target concept results from its boundary region induced by the lower approximation and the upper approximation.

- **Rough degree** $$\rho_A(X) = 1 - \frac{|\underline{R}X|}{|\overline{R}X|} = \frac{|BN(X)|}{|\overline{R}X|}$$

For the <u>different</u> approximation spaces, the rough degrees of a target concept <u>may be identical</u>.

Rough Entropy

- **Rough entropy of *A***

$$E_r(A) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|}$$

- **Rough entropy of *X***

$$\begin{aligned} E_A(X) &= \rho_A(X) E_r(A) \\ &= -\rho_A(X)\left(\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log \frac{1}{|X_i|}\right) \end{aligned}$$

The rough entropy possess the better depicting ability than the rough degree for measuring the roughness of a rough set.

◆ **Relationship between the rough entropy and Shannon's entropy in an information system**

$$E_r(A) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|} \qquad H(A) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|}$$

$$E_r(A) + H(A) = \log_2 |U|$$

The relationship between the rough entropy and Shannon's entropy is strict complement relationship, and they possess the same capability on depicting the uncertainty of an information system.

See: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 12 (1) : 37-46.

Fuzziness

Fuzzy entropy

A fuzzy entropy characterizes the fuzziness degree of a fuzzy set.

- $$e_1^p(A) = \frac{2}{n^{1/n}} d^p(A, A_{near})$$
- $$e_2^p(A) = \frac{d^p(A, A_{near})}{d^p(A, A_{far})}$$
- $$e_3(A) = -k\sum_{i=1}^{n}(\mu_A(u_i)\ln\mu_A(u_i) + (1-\mu_A(u_i))(1-\ln\mu_A(u_i))), k > 0$$
- $$e_4(A) = \frac{1}{n\sqrt{e}-1}\sum_{i=1}^{n}(\mu_A(u_i)e^{1-\mu_A(u_i)} + (1-\mu_A(u_i))e^{\mu_A(u_i)} - 1)$$
- $$e_5(A) = \frac{4}{n}\sum_{i=1}^{n}\mu_A(u_i)(1-\mu_A(u_i))$$

Fuzziness

◆ Fuzziness of a rough set

For any object $u \in U$, the membership function of $u \in X$ is denoted by

$$\delta_X(u) = \frac{|X \cap [u]_A|}{|X|}.$$

where $\delta_X(u)$ represents a fuzzy concept. The fuzziness of the rough set can be measured by the following fuzzy entropy

$$e_A(X) = \frac{4}{|U|}\sum_{i=1}^{|U|}\delta_X(u_i)(1-\delta_X(u_i)).$$

Fuzzy entropy can be employed to measure the fuzziness of a rough set or a rough decision in rough set theory.

Information Granularity

Information granularity denotes the average measure of a granular space induced by some information granules.

- **Knowledge granularity**

$$GK(A) = \frac{1}{|U|^2} \sum_{i=1}^{m} |X_i|^2$$

$$E(A) + GK(A) = 1$$

- **Combination granularity**

$$CG(A) = \sum_{i=1}^{m} \frac{|X_i|}{|U|} \frac{C_{|X_i|}^2}{C_{|U|}^2}$$

$$CE(A) + CG(A) = 1$$

See: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 2008.

Partial Order Relation

◆ Rough partial-order relation

Let $K = (U, \boldsymbol{R})$ be a group of granular spaces and $P, Q \in \boldsymbol{R}$. $K(P) = \{ N_P(x), x \in U \}$ and $K(Q) = \{ N_Q(x), x \in U \}$ the granular structure induced by P and Q, where $N_P(x)$ and $N_Q(x)$ are the neighborhood induced by object x with respect to P and Q.

➢ Rough partial order relation $\preceq$ is defined as:

$$K(P) \preceq K(Q)(P, Q \in R) \Leftrightarrow N_P(x) \subseteq N_Q(x), x \in U.$$

If $K(P) \preceq K(Q)$ and $K(P) \neq K(Q)$, we say $K(P)$ is strictly finer than $K(Q)$, denoted by $K(P) \prec K(Q)$.

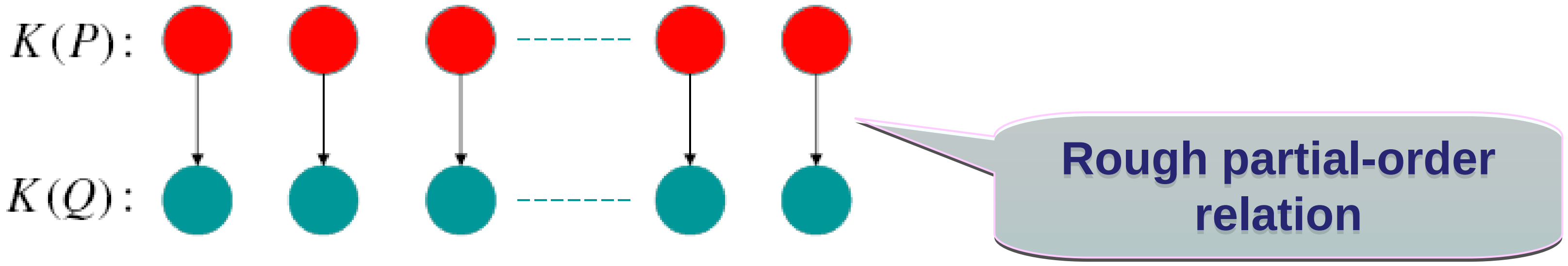
◆ Granulation partial-order relation

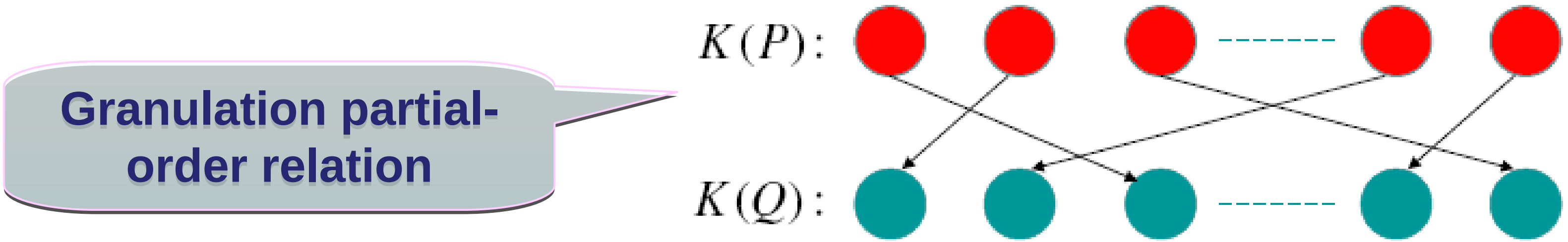
- Granulation partial order relation $\leqslant$ is defined as:

$K(P) \leqslant K(Q) \Leftrightarrow$ There exists a bijective mapping function $f: K(P) \rightarrow K(Q)$ such that $|N_P(x)| \leq |f(N_P(x))|, x \in U$.

If there is a bijective mapping function $f: K(P) \rightarrow K(Q)$ such that $|N_P(x)| = |f(N_P(x))|, x \in U$, denoted by $K(P) \approx K(Q)$.

If $K(P) \leqslant K(Q)$ and $K(P) \not\approx K(Q)$, we say $K(P)$ is strictly granulation than $K(Q)$, denoted by $K(P) \ll K(Q)$.





See: IEEE Transactions on Fuzzy Systems, 2011, 19(2): 253-264.

Axiom Approach of Information Granularity

Axiom 1: Let $K = (U, \boldsymbol{R})$ be a group of granular spaces, if for $\forall P \in \boldsymbol{R}$, there is a real number $G(P)$ with the following properties:

(1) $G(P) \geq 0$; (Non-negative)

(2) $\forall P, Q \in \boldsymbol{R}, if\ K(P) = K(Q), then\, G(P) = G(Q)$; (Invariability)

(3) $\forall P, Q \in \boldsymbol{R}, if\ K(P) \prec K(Q), then\, G(P) < G(Q)$. (Rough monotonicity)

Then G is called a **rough granularity** on K.

Axiom 2: Let $K = (U, \boldsymbol{R})$ be a group of granular spaces, if for $\forall P \in \boldsymbol{R}$, there is a real number $G(P)$ with the following properties:

(1) $G(P) \geq 0$; (Non-negative)

(2) $\forall P, Q \in \boldsymbol{R}, if\ K(P) \approx K(Q), then\, G(P) = G(Q)$; (Invariability)

(3) $\forall P, Q \in \boldsymbol{R}, if\ K(P) \ll K(Q), then\, G(P) < G(Q)$. (Granulation monotonicity)

Then G is called an **information granularity** on K.

See: Information granules and entropy theory in information systems. Sci. China., Ser. F. 51, 1427-1444 (2008)

Related Properties

It has been proved that some existing definitions are various special forms of information granularity.

(1) $GK(A)$ is an information granularity, $\frac{1}{|U|} \leq GK(A) \leq 1$.

(2) $CG(A)$ is an information granularity, $0 \leq CG(A) \leq 1$.

(3) $E_r(A)$ is an information granularity, $0 \leq E_r(A) \leq \log_2 |U|$.

See: Information granules and entropy theory in information systems. Sci. China., Ser. F. 51, 1427-1444 (2008)

Knowledge Distance

- In rough set theory, information entropy and knowledge granulation are two main approaches to measuring the uncertainty of a knowledge structure in knowledge bases.
- If the knowledge granulation (or information entropy) of one knowledge structure is equal to that of the other knowledge structure, we say that these two knowledge structures have the same uncertainty.
- However, it does not mean that these two knowledge structures are equivalent each other.
- Information entropy and knowledge granulation cannot characterize the difference between any two knowledge structures in a knowledge base.

For the information system $S = (U, A), P, Q \subseteq A$. The knowledge distance between $K(P)$ and $K(Q)$ is defined as

$$D(K(P), K(Q)) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|}$$

The knowledge distance aims to reveal the geometrical structure in granular spaces.

See: International Journal of Approximate Reasoning, 2009, 50 : 174 - 188.

Basic Properties

- Several properties of the knowledge distance:

(1) $0 \le D(K(P), K(Q)) \le 1 - \frac{1}{|U|}$; (Extremum)

(2) $D(\omega, \delta) = 1 - \frac{1}{|U|}$, and $D(K(P), \omega) + D(K(P), \delta) = 1 - \frac{1}{|U|}$, (Complement)

where $\omega = \{N_P(x_i) \mid N_P(x_i) = \{x_i\}, x_i \in U\}$, and $\delta = \{N_P(x_i) \mid N_P(x_i) = U, x_i \in U\}$;

(3) $D(K(P), K(Q)) = D(\neg K(P), \neg K(Q))$; (Symmetry)

(4) $if\ K(P) \preceq K(Q) \preceq K(R),\ then\ D(K(P), K(R)) = D(K(P), K(Q)) +$ $D(K(Q), K(R))$; (Linearity)

(5) $if\ K(P) \preceq K(Q),\ then\ D(K(P), \omega) \le D(K(Q), \omega)\ and\ D(K(P), \delta)$ $\ge D(K(Q), \delta)$. (Monotonicity)

Let $\mathbf{K}(U)$ be the set of all granular spaces induced by U, then $(\mathbf{K}(U), D)$ is a distance space.

- **Non-negtive**
- **Symmetry**
- **Triangle inequality**

- **The knowledge distance can be used to distinguish the divergence between two granular structures with the same information granularity (or information entropy).**
- **The knowledge distance characterizes the essence of uncertainty of granular structures.**

Axiom approach of generalized information granularity

Axiom 3: Let $K = (U, \boldsymbol{R})$ be a group of granular spaces, if for $\forall P \in \boldsymbol{R}$, there is a real number $G(P)$ with the following properties:

(1) $G(P) \geq 0$; (Non-negative)

(2) $\forall P, Q \in \boldsymbol{R}, if\ D(K(P), \omega) = D(K(Q), \omega), then\ G(P) = G(Q)$; (Invariability)

(3) $\forall P, Q \in \boldsymbol{R}, if\ D(K(P), \omega) < D(K(Q), \omega), then\ G(P) < G(Q)$. (Monotonicity)

Then G is called a **generalized information granularity** on K.

Rough granularity → **Information granularity** → **Generalized information granularity**

2. Accelerator of Feature Selection

Accelerator of Feature Selection

Feature selection

Feature selection is a challenging problem in areas such as pattern recognition, machine learning and data mining.

To select feature subset efficiently, many heuristic feature selection algorithms have been developed. The common approach is the forward greedy search strategy to select a subset of features , which has a wide variety of applications.

In feature selection, there are two general strategies, namely wrappers and filters.

In rough set theory, feature selection (also called attribute reduction) aims to retain the discriminatory power of original features.

Reduction for two types of data

- **Symbolic values**——Discernibility matrix approach and heuristic approach.
- **Numerical values**——Relying on fuzzy rough set theory or doing discretization of the numerical attributes.

Two Reduction Tasks

Attribute reduction

- **Complete reduction**——Employing discernibility matrix approach to obtain all reducts of an information system (or a decision table).
- **A single reduct**——Finding a single reduct from a given data set by using heuristic search strategy.
- Some other reduction approaches, such as optimal reduction, approximation reduction, and so on.

Discernibility Matrix Approach

- Skowron proposed a discernibility matrix approach to obtain all attribute reducts of an information system.

- Many other scholars studied the discernibility matrix approach in extended rough set models.

Heuristic Attribute Reduction Approach

- Grzymala Busse proposed the idea of attribute reduction using positive region.
- Hu and Cercone proposed the positive-region reduction algorithm for a decision table.
- Ziarko developed the β-reduct based on the variable precision rough set model.
- Yao et al. gave the attribute reduction approach in decision theoretic rough set.
- Many other techniques of heuristic attribute reduction approaches are provided.

Limitation of Existing Algorithms

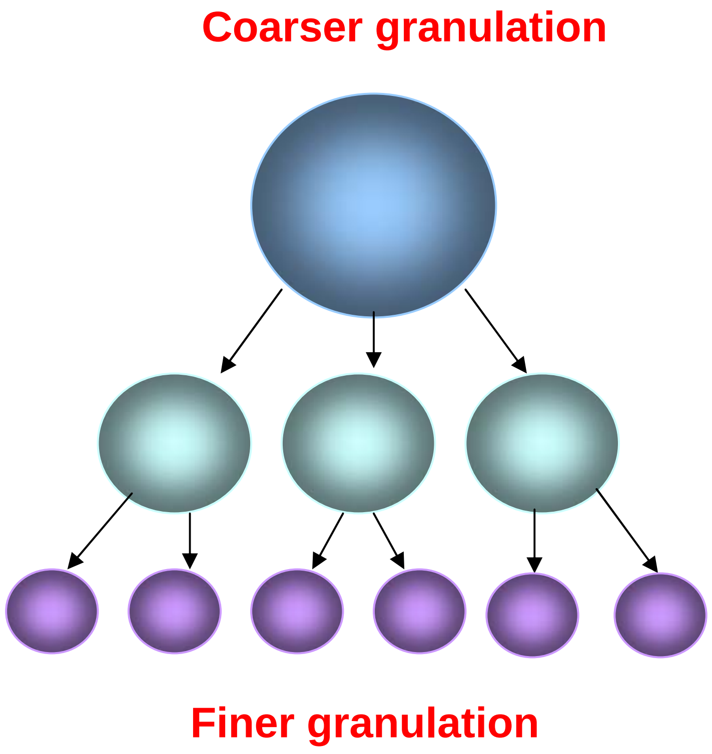
- **Each of the existing methods preserves a particular property of a given information system or a given decision table.**

- **The existing algorithms are still computationally very expensive, which are intolerable for dealing with large-scale data sets with high dimensions.**

Granulation Order

The partition induced by equivalence relation provides a granulation world of describing the target concept. Hence, one can structure a granulation world ordered from coarser to finer.

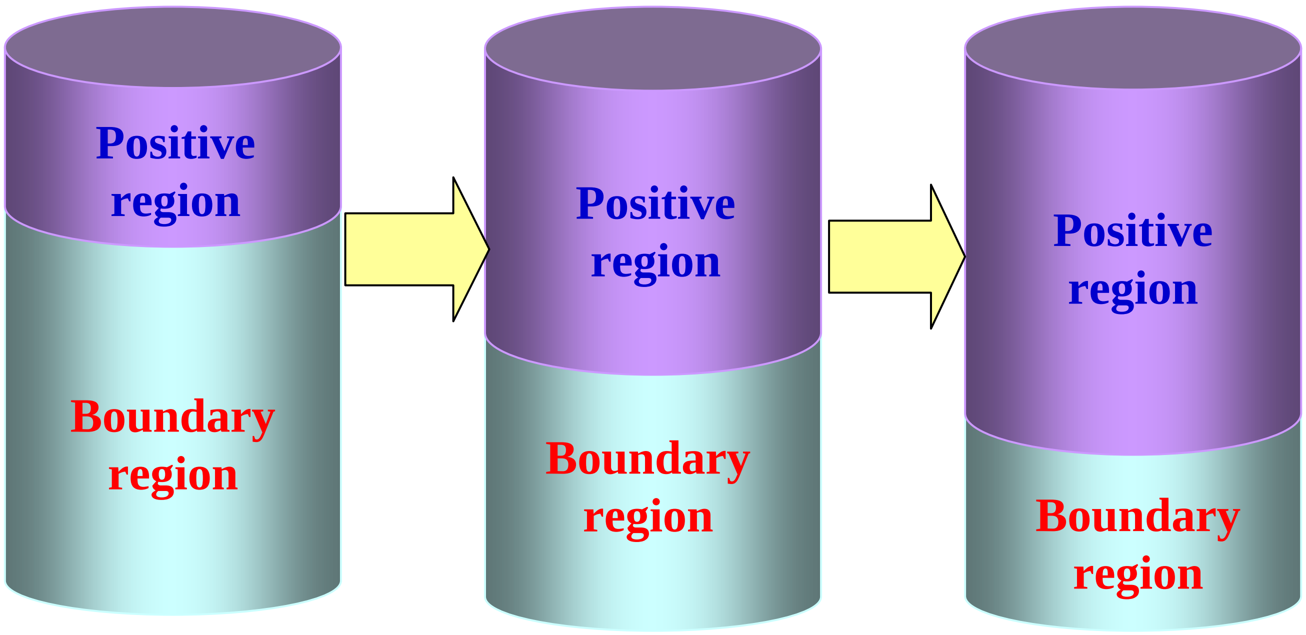
In an information system, by adding attributes or attribute values, one can get a ordered granulation world (from coarser to finer).



Positive Region Varying with Granulation

$$R_1 \succeq R_2 \succeq \cdots \succeq R_n$$

Coarser granulation ⟶ Finer granulation



Recursive Expression Principle

Let $S=(U,C\cup D)$ be a decision table, $X\subseteq U$ and $P=\{R_1,R_2,\cdots,R_n\}$ with $R_1\preceq R_2\preceq\cdots\preceq R_n\ (R_i\in 2^C)$. Given $P_i=\{R_1,R_2,\cdots,R_i\}$, then

$$POS^U_{P_{i+1}}(D)=POS^U_{P_i}(D)\cup POS^{U_{i+1}}_{R_{i+1}}(D),$$

where $U_1=U$ and $U_{i+1}=U-POS^U_{P_i}(D)$.

The principle shows that a target decision can be positively approximated by using a granulation order from coarser to finer. This mechanism implies the idea of the accelerator for improving the computing performance of a heuristic attribute algorithm.

Representative Significance Measures

Significance measures of attributes

- ✓ Attribute dependent degree $\gamma_C(D)$
- ✓ Shannon's conditional entropy $H(D|C)$
- ✓ Complementary conditional entropy $E(D|C)$
- ✓ Combination conditional entropy $CE(D|C)$

Inner Importance

For the decision table $S=(U,C\cup D)$ and $B\subseteq C$, the significance measure of $a\in B$ is defined as

$$Sig_1^{inner}(a,B,D)=\gamma_B(D)-\gamma_{B-\{a\}}(D),$$

$$Sig_2^{inner}(a,B,D)=H(D\,|\,B-\{a\})-H(D\,|\,B),$$

$$Sig_3^{inner}(a,B,D)=E(D\,|\,B-\{a\})-E(D\,|\,B),$$

$$Sig_4^{inner}(a,B,D)=CE(D\,|\,B-\{a\})-CE(D\,|\,B).$$

Outer Importance

For the decision table $S=(U,C\cup D)$ and $B\subseteq C$, the significance measure of $a\in C-B$ is defined as

$$Sig_1^{outer}(a,B,D)=\gamma_{B\cup\{a\}}(D)-\gamma_B(D),$$

$$Sig_2^{outer}(a,B,D)=H(D\,|\,B)-H(D\,|\,B\cup\{a\}),$$

$$Sig_3^{outer}(a,B,D)=E(D\,|\,B)-E(D\,|\,B\cup\{a\}),$$

$$Sig_4^{outer}(a,B,D)=CE(D\,|\,B)-CE(D\,|\,B\cup\{a\}).$$

Rank Preservation

- **Rank preservation of the significance of attributes**

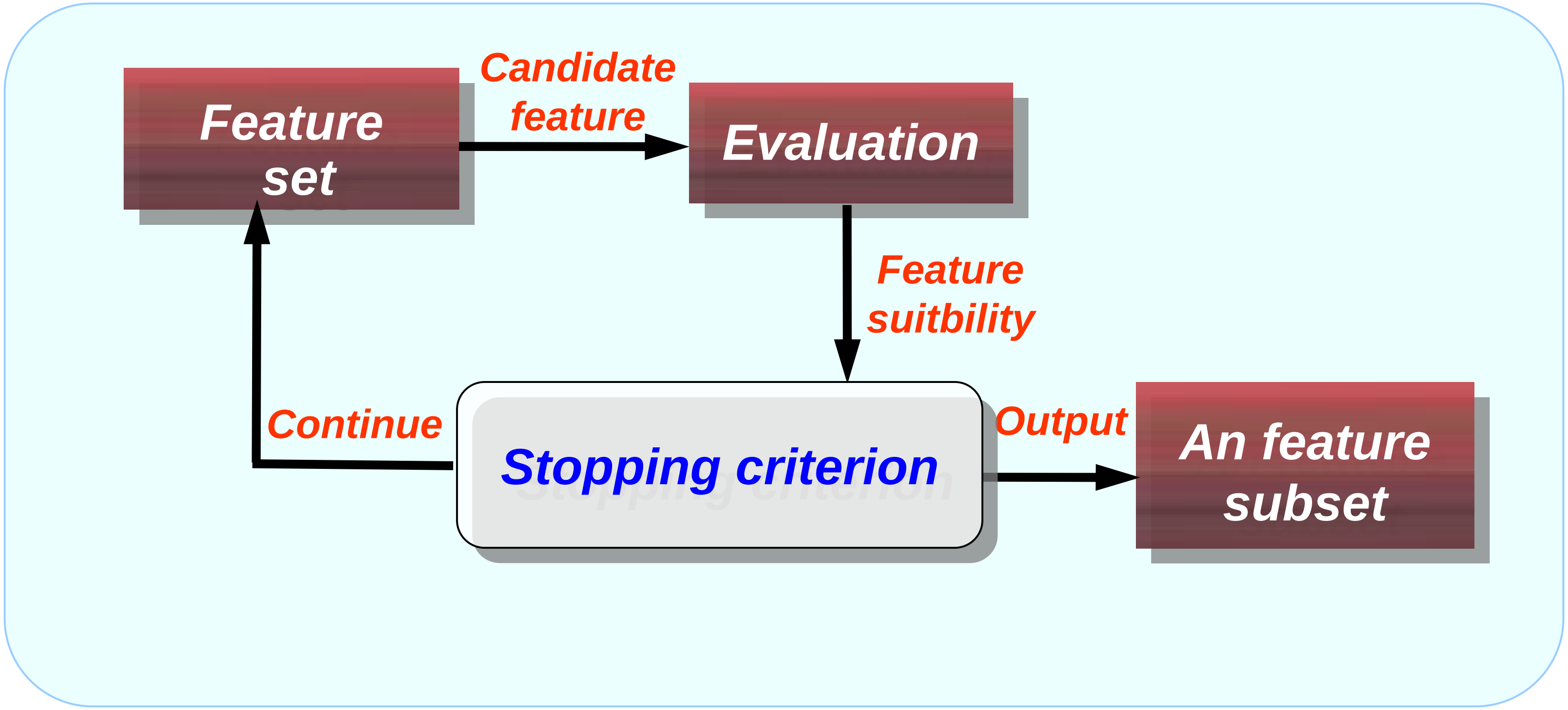
$$sig^{outer}(a,B,D,U) \geq sig^{outer}(b,B,D,U)$$

$$\Rightarrow sig^{outer}(a,B,D,U') \geq sig^{outer}(b,B,D,U')$$

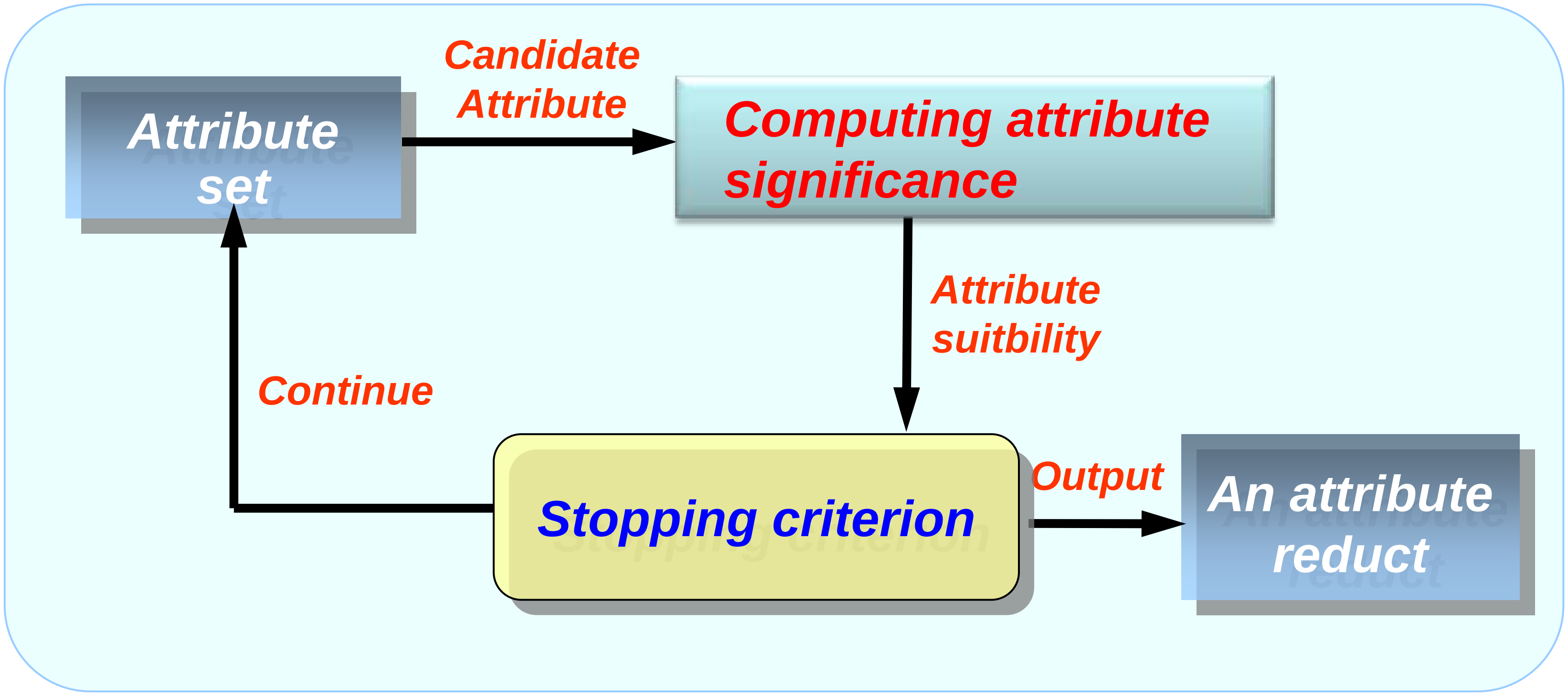
$where,\ U' = U - POS_B^U(D).$

From above equation, one can see that the rank of attributes in the process of attribute reduction will retain unchanged after reducing the lower approximation of positive approximation.

This mechanism can be used to improve the computational performance of a heuristic attribute reduction algorithm, while retaining the same selected feature subset.



The process of forward greedy algorithm



The process of forward greedy algorithm based on rough set theory

Feature selection algorithm based on rough set theory

Algorithm 1. A general forward greedy attribute reduction algorithm.

Input: Decision table $S = (U, C \cup D)$;
Output: One reduct *red*.
Step 1: $red \leftarrow \emptyset$; //*red* is the pool to conserve the selected attributes;
Step 2: Compute $Sig^{inner}(a_k, C, D)$, $k \leqslant |C|$; //$Sig^{inner}(a_k, C, D)$ is the inner importance measure of the attribute a_k;
Step 3: Put a_k into *red*, where $Sig^{inner}(a_k, C, D, U) > 0$;
Step 4: While $EF(red, D) \neq EF(C, D)$ Do //This provides a stopping criterion.
$\{red \leftarrow red \cup \{a_0\}$, where $Sig^{outer}(a_0, red, D) = \max\{Sig^{outer}(a_k, red, D), a_k \in C - red\}\}$; //$Sig^{outer}(a_k, C, D)$ is the outer importance measure of the attribute a_k;
Step 5: Return *red* and end.

Accelerated Feature Selection Algorithm

Algorithm Q1. A general improved feature selectio

Input: Decision table $S = (U, C \cup D)$;
Output: One reduct *red*.
Step 1: $red \leftarrow \emptyset$; //*red* is the pool to conserve the s
Step 2: Compute $Sig^{inner}(a_k, C, D, U)$, $k \leqslant |C|$;
Step 3: Put a_k into *red*, where $Sig^{inner}(a_k, C, D, U)$ >
Step 4: $i \leftarrow 1$, $R_1 = red$, $P_1 = \{R_1\}$ and $U_1 \leftarrow U$;
Step 5: While $EF^{U_i}(red, D) \neq EF^{U_i}(C, D)$ Do

{Compute the positive region of positive ap
$U_i = U - POS^U_{P_i}(D)$,
$i \leftarrow$
$red \leftarrow red \cup \{a_0\}$, where $Sig^{outer}(a_0, red, D, U$
$R_i \leftarrow R_i \cup \{a_0\}$,
$P_i \leftarrow \{R_1, R_2, \ldots, R_i\}\}$;

Accelerator

Step 6: Return *red* and end.

{Compute $POS^U_{P_i}(D)$,

$U_{i+1} = U - POS^U_{P_i}(D)$,

$i \leftarrow i+1$,

$red \leftarrow red \cup \{a_0\}$, where $Sig^{outer}(a_0, red, D, U_i)$

$= \max\{Sig^{outer}(a_k, red, D, U_i), a_k \in C - red\}$,

$R_i \leftarrow R_{i-1} \cup \{a_0\}$,

$P_i \leftarrow \{R_1, R_2, \cdots, R_i\}$.

}

See: Positive approximation: an accelerator for attribute reduction in rough set theory, ***Artificial Intelligence***, 2010, 174(9-10): 597-618.

- The complexities description

Algorithms	Step 2	Step 3	Step 5	Other steps
Each of original algorithms	$O(\|C\|\|U\|)$	$O(\|C\|)$	$O(\sum_{i=1}^{\|C\|}\|U\|(\|C\|-i+1)))$	Constant
FSPA	$O(\|C\|\|U\|)$	$O(\|C\|)$	$O(\sum_{i=1}^{\|C\|}\|U_i\|(\|C\|-i+1))$	Constant

Experimental Analysis

Data sets description

	Data sets	Cases	Features	Classes
1	Mushroom	5644	22	2
2	Tic-tac-toe	958	9	2
3	Dermatology	358	34	6
4	Kr-vs-kp	3196	36	2
5	Breast-cancer-wisconsin	683	9	2
6	Backup-large.test	376	35	19
7	Shuttle	58000	9	7
8	Letter-recognition	20000	16	26
9	Ticdata2000	5822	85	2

Experimental Design

Each of these nine data sets is divided into twenty parts equally, denoted by

$$x_i \ (i = 1, 2, \cdots, 20).$$

The twenty data sets using in the experiment is the combination of x_i , denoted by

$$X_i \ (i = 1, 2, \cdots, 20),$$

where,

$$X_1 = x_1,$$

$$X_2 = x_1 + x_2,$$

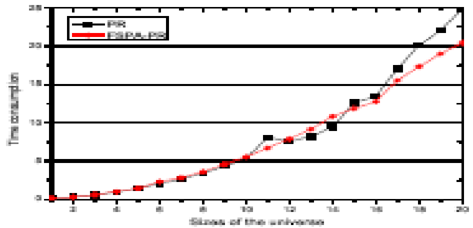
$$\vdots$$

$$X_{20} = x_1 + x_2 + \cdots + x_{20}.$$

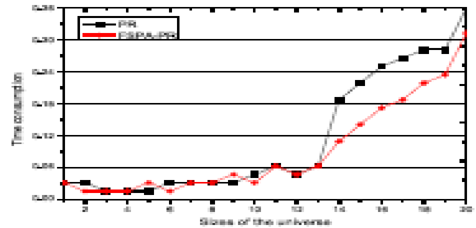
Comparison of Algorithms

The time and reduct of the classic algorithm and accelerated algorithm based on positive region

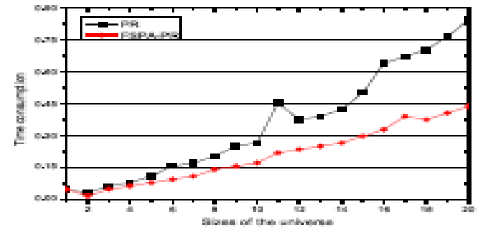
Data sets	Original features	PR algorithm		FSPA-PR algorithm	
		Selected features	Time (s)	Selected features	Time (s)
Mushroom	22	3	24.8750	3	20.4531
Tic-tac-toe	9	8	0.3594	8	0.3125
Dermatology	34	10	0.8438	10	0.4375
Kr-vs-kp	36	29	28.0313	29	21.5781
Breast-cancer-wisconsin	9	4	0.1250	4	0.0938
Backup-large.test	35	10	0.6563	10	0.4219
Shuttle	9	4	906.0625	4	712.2500
Letter-recognition	16	11	282.6406	11	112.6250
Ticdata2000	85	24	886.4531	24	296.3750



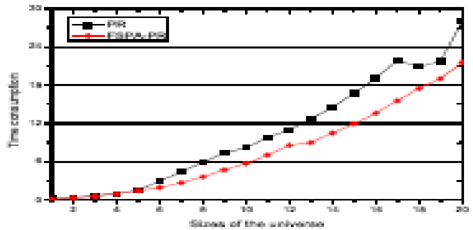
(a) Mushroom



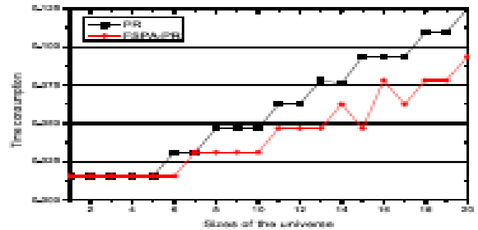
(b) Tic-tac-toe



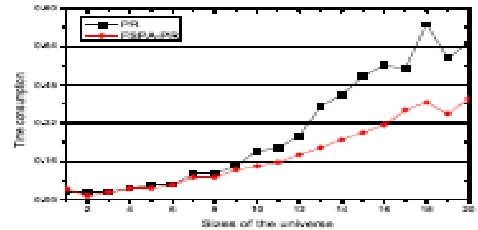
(c) Dermatology



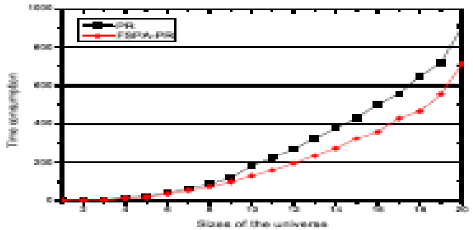
(d) Kr-vs-kp



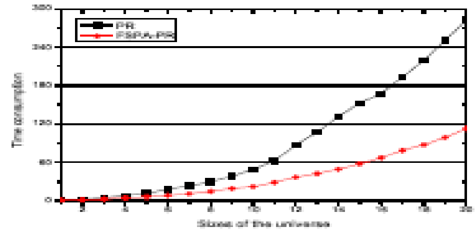
(e) Breast-cancer-wisconsin



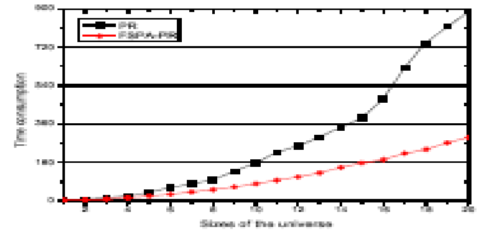
(f) Backup-large.test



(g) Shuttle



(h) Letter-recognition

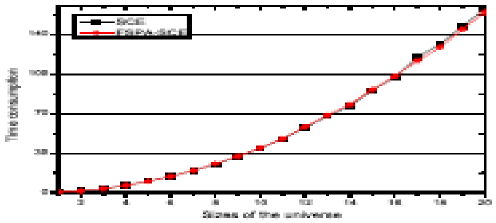


(i) Ticdata2000

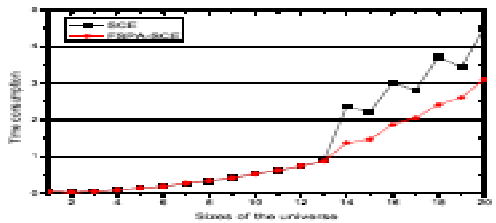
The time of the classic algorithm and accelerated algorithm based on positive region

The time and reduct of the algorithms based on Shonnon’s entropy

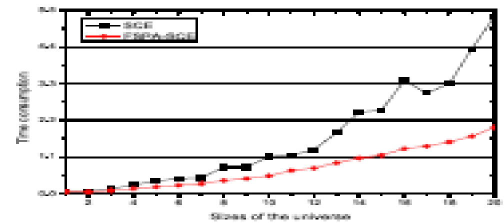
Data sets	Original features	SCE algorithm		FSPA-SCE algorithm	
		Selected features	Time (s)	Selected features	Time (s)
Mushroom	22	4	162.6406	4	159.5938
Tic-tac-toe	9	8	4.5000	8	3.1094
Dermatology	34	11	5.3125	11	1.9844
Kr-vs-kp	36	29	149.6250	29	105.9844
Breast-cancer-wisconsin	9	4	1.3438	4	0.8438
Backup-large.test	35	10	4.3594	10	1.7656
Shuttle	9	4	12665.3906	4	10153.1719
Letter-recognition	16	11	7015.7031	11	2740.2500
Ticdata2000	85	24	8153.6563	24	1043.8906



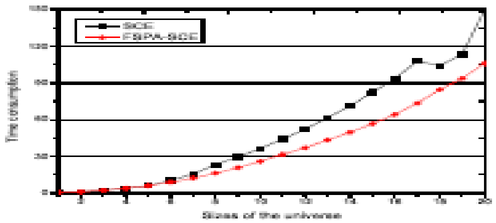
(a) Mushroom



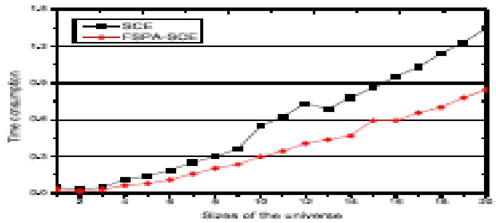
(b) Tic-tac-toe



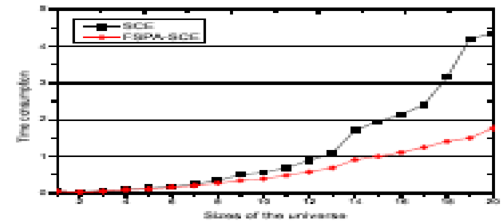
(c) Dermatology



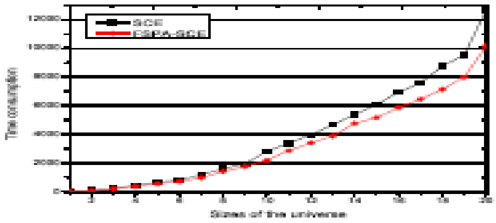
(d) Kr-vs-kp



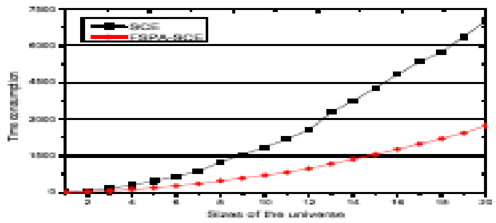
(e) Breast-cancer-wisconsin



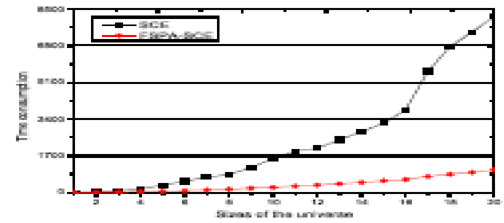
(f) Backup-large.test



(g) Shuttle



(h) Letter-recognition

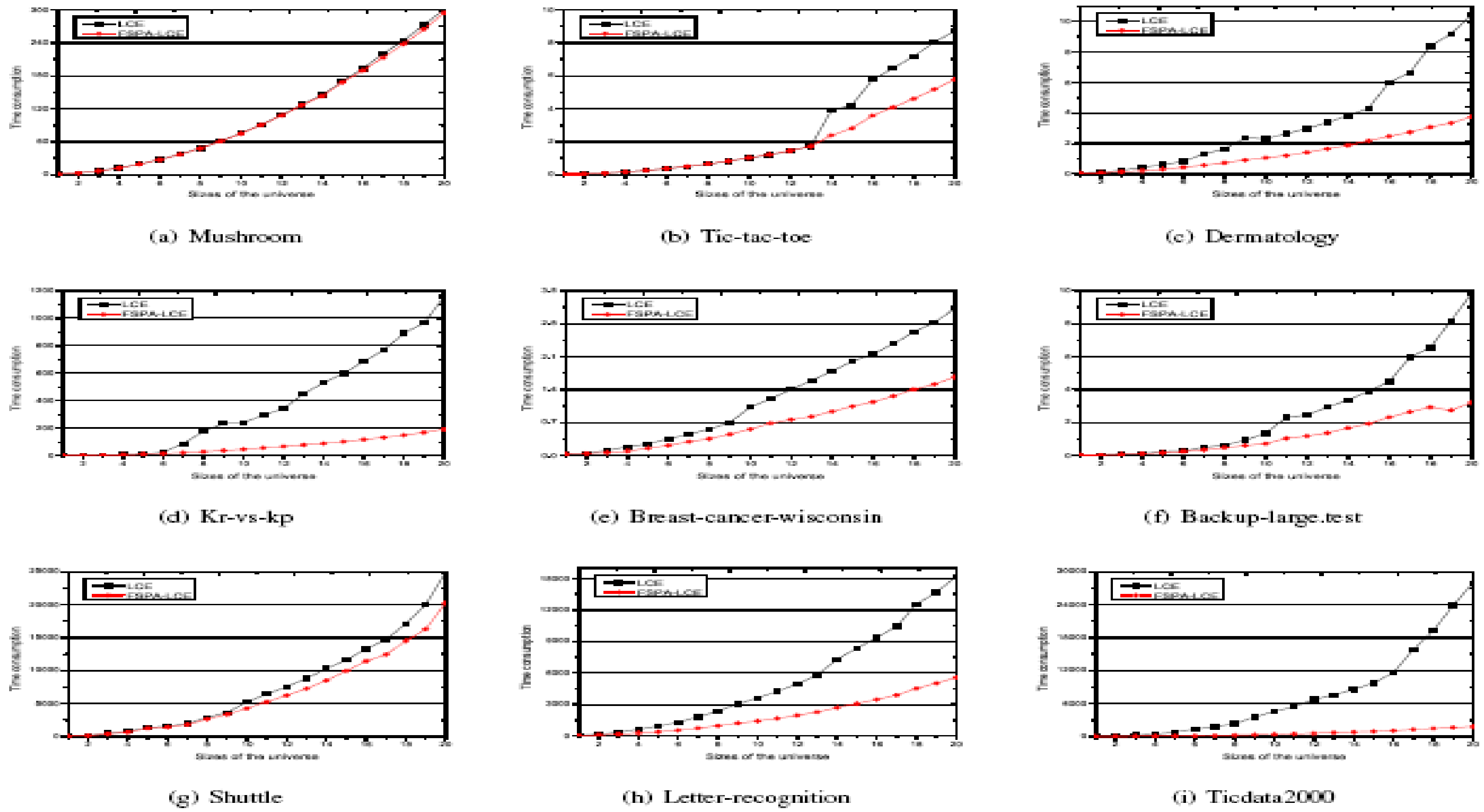


(i) Ticdata2000

The time of the algorithms based on Shannon’s entropy

The time and reduct of the algorithms based on complementary entropy

Data sets	Original features	LCE algorithm		FSPA-LCE algorithm	
		Selected features	Time (s)	Selected features	Time (s)
Mushroom	22	4	300.2188	4	294.0000
Tic-tac-toe	9	8	8.7344	8	5.7813
Dermatology	34	10	10.4531	10	3.7500
Kr-vs-kp	36	29	1156.1250	29	191.1250
Breast-cancer-wisconsin	9	5	3.1250	5	1.6719
Backup-large.test	35	10	9.8438	10	3.2188
Shuttle	9	4	24883.6250	4	20228.3906
Letter-recognition	16	12	15176.7656	12	5558.7813
Ticdata2000	85	24	27962.6250	24	1805.5625

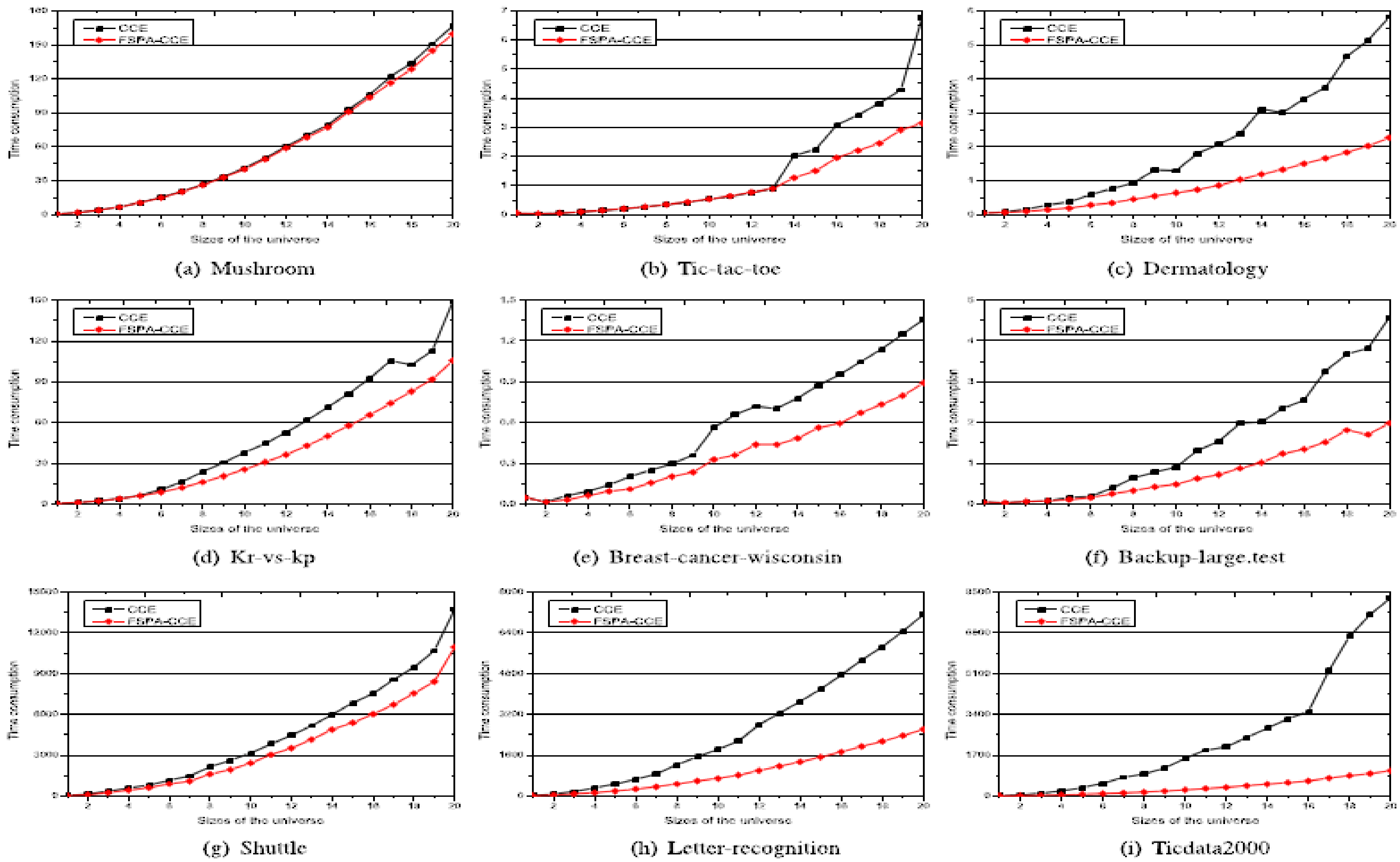


The time of the algorithms based on complementary entropy

The time and reduct of the algorithms based on combination entropy

Table 6. The time and attribute reduction of the algorithms CCE and FSPA-CCE

Data sets	Original features	CCE algorithm		FSPA-CCE algorithm	
		Selected features	Time (s)	Selected features	Time (s)
Mushroom	22	4	166.9219	4	159.6406
Tic-tac-toe	9	8	6.7656	8	3.1406
Dermatology	34	10	5.8281	10	2.2656
Kr-vs-kp	36	29	149.7500	29	105.7500
Breast-cancer-wisconsin	9	4	1.3594	4	0.8906
Backup-large.test	35	9	4.5781	9	1.9844
Shuttle	9	4	13718.8750	4	10948.9219
Letter-recognition	16	11	7118.2656	11	2610.3594
Ticdata2000	85	24	8262.0469	24	1048.5781



The time of the algorithms based on combination entropy

Stability Analysis of Accelerator Algorithm

- **Experiment design**

The stability of a heuristic attribute reduction algorithm determines the stability of its classification accuracy.

The objective of this suite of experiments is to compare the stability of the computing time and attribute reduction of each of the modified algorithms with those obtained when running the original methods.

In the experiments, in order to evaluate the stability of feature subset selected with 10-fold cross validation, we partition a given data set to 10 subsets with the same size. The standard deviation is used to the stability of each algorithm. The lower the value of the standard deviation, the higher the stability of the algorithm.

The stabilities of the time and attribute reduction of algorithms PR and FSPA-PR.

Data sets	PR's time	FSPA-PR's time	PR's stability	FSPA-PR's stability
Mushroom	16.8359 ± 0.2246	14.8438 ± 0.2130	0.0000 ± 0.0000	0.0000 ± 0.0000
Tic-tac-toe	0.3234 ± 0.0222	0.2391 ± 0.0262	0.0000 ± 0.0000	0.0000 ± 0.0000
Dermatology	0.8234 ± 0.0494	0.3922 ± 0.0109	0.2142 ± 0.1692	0.2142 ± 0.1692
Kr-vs-kp	25.0781 ± 4.3400	16.2438 ± 0.2232	0.0675 ± 0.0652	0.0675 ± 0.0652
Breast-cancer-wisconsin	0.1156 ± 0.0104	0.0813 ± 0.0094	0.1733 ± 0.2736	0.1733 ± 0.2736
Backup-large.test	0.6344 ± 0.0788	0.3891 ± 0.0331	0.4187 ± 0.1830	0.4187 ± 0.1830
Shuttle	778.6959 ± 29.4587	551.6750 ± 10.6770	0.0250 ± 0.0750	0.0250 ± 0.0750
Letter-recognition	224.1219 ± 7.3887	90.5797 ± 1.5252	0.2222 ± 0.2020	0.2222 ± 0.2020
Ticdata2000	698.1016 ± 54.8386	248.8391 ± 6.5261	0.2058 ± 0.0862	0.2058 ± 0.0862

The stabilities of the time and attribute reduction of algorithms SCE and FSPA-SCE.

Data sets	SCE's time	FSPA-SCE's time	SCE's stability	FSPA-SCE's stability
Mushroom	130.6234 ± 0.9870	126.1625 ± 0.8873	0.0000 ± 0.0000	0.0000 ± 0.0000
Tic-tac-toe	3.8359 ± 0.0614	2.5045 ± 0.0617	0.1111 ± 0.1111	0.1111 ± 0.1111
Dermatology	4.0500 ± 0.3197	1.6266 ± 0.0422	0.5312 ± 0.1000	0.5312 ± 0.1000
Kr-vs-kp	126.7734 ± 15.7752	83.2891 ± 0.9501	0.0675 ± 0.0652	0.0675 ± 0.0652
Breast-cancer-wisconsin	1.2156 ± 0.0894	0.7500 ± 0.0677	0.3562 ± 0.3099	0.3562 ± 0.3099
Backup-large.test	3.7234 ± 0.3919	1.4188 ± 0.0655	0.3599 ± 0.2521	0.3599 ± 0.2521
Shuttle	9749.1705 ± 308.8128	8158.8490 ± 209.5685	0.0250 ± 0.0750	0.0250 ± 0.0750
Letter-recognition	5891.5906 ± 181.0442	2282.8141 ± 73.0362	0.1689 ± 0.1823	0.1689 ± 0.1823
Ticdata2000	7107.3904 ± 105.7970	861.2000 ± 9.7081	0.2485 ± 0.0830	0.2485 ± 0.0830

The stabilities of the time and attribute reduction of algorithms LCE and FSPA-LCE.

Data sets	LCE's time	FSPA-LCE's time	LCE's stability	FSPA-LCE's stability
Mushroom	241.9891 ± 1.3425	236.0313 ± 1.6868	0.0000 ± 0.0000	0.0000 ± 0.0000
Tic-tac-toe	7.3328 ± 0.0601	4.7531 ± 0.1007	0.1778 ± 0.0889	0.1778 ± 0.0889
Dermatology	8.2875 ± 0.6289	3.0938 ± 0.0617	0.1852 ± 0.1783	0.1852 ± 0.1783
Kr-vs-kp	228.9547 ± 27.4934	154.4984 ± 2.0417	0.0675 ± 0.0652	0.0675 ± 0.0652
Breast-cancer-wisconsin	2.5969 ± 0.0493	1.4031 ± 0.0554	0.2333 ± 0.1528	0.2333 ± 0.1528
Backup-large.test	7.9094 ± 0.4949	2.7109 ± 0.1746	0.1617 ± 0.1630	0.1617 ± 0.1630
Shuttle	17717.9594 ± 391.4628	14392.4496 ± 99.2163	0.0250 ± 0.0750	0.0250 ± 0.0750
Letter-recognition	12334.2729 ± 80.6504	4252.5578 ± 71.4054	0.1914 ± 0.1436	0.1914 ± 0.1436
Ticdata2000	19582.6515 ± 385.2873	1463.7391 ± 14.5646	0.1744 ± 0.1192	0.1744 ± 0.1192

The stabilities of the time and attribute reduction of algorithms CCE and FSPA-CCE.

Data sets	CCE's time	FSPA-CCE's time	CCE's stability	FSPA-CCE's stability
Mushroom	133.9672 ± 0.9331	129.3531 ± 1.2343	0.0000 ± 0.0000	0.0000 ± 0.0000
Tic-tac-toe	3.8391 ± 0.0297	2.5172 ± 0.0439	0.1778 ± 0.0889	0.1778 ± 0.0889
Dermatology	4.6469 ± 0.3029	1.8016 ± 0.0335	0.2735 ± 0.1698	0.2735 ± 0.1698
Kr-vs-kp	130.0047 ± 17.5668	86.3641 ± 0.9297	0.0733 ± 0.0780	0.0733 ± 0.0780
Breast-cancer-wisconsin	1.1969 ± 0.0865	0.7406 ± 0.0298	0.1200 ± 0.1600	0.1200 ± 0.1600
Backup-large.test	3.8016 ± 0.3155	1.5875 ± 0.1018	0.3426 ± 0.1780	0.3426 ± 0.1780
Shuttle	9564.8752 ± 68.5368	7440.3281 ± 25.0001	0.0250 ± 0.0750	0.0250 ± 0.0750
Letter-recognition	5956.0833 ± 43.7866	2171.0000 ± 36.5273	0.1370 ± 0.1450	0.1370 ± 0.1450
Ticdata2000	6726.4778 ± 42.1287	859.4672 ± 10.7790	0.1742 ± 0.0894	0.1742 ± 0.0894

Advantage of Accelerator Algorithm

- Each of the accelerated algorithms preserves the attribute reduct induced by the corresponding original one.
- Each of the accelerated algorithms usually comes with a substantially reduced computing time when compared with amount of time used by the corresponding original algorithm.
- The performance of these modified algorithms is getting better in presence of larger data sets; the larger the data set, the more profound computing savings.

3. Conclusion and Further Work

Conclusion

- The uncertainty measures can be used to measure the significance of attributes, design heuristic feature selection algorithms, etc.

- The granular space distance can be used to distinguish the divergence between two granular structures with the same information granulation (or information entropy).

- The accelerated algorithm can choose the same attribute reduct as its original version, which possesses the same classification accuracy.

- The accelerated algorithm is high-efficiency, especially for large-scale data sets.

Further Work

- ✓ **Uncertainty measures for generalized rough set models.**
- ✓ **Feature selection for the large-scale data sets by separating and fusing data sets.**
- ✓ **Efficient accelerated feature selection mechanism for hybrid data sets.**
- ✓ **Incremental feature selection algorithms for dynamic data sets.**

It is our wish that this study provides new views on dealing with large-scale and complicated data sets in applications.


Thank you!