International Journal of Database Theory and Application
Vol.9, No.5 (2016), pp.77-90
http://dx.doi.org/10.14257/ijdta.2016.9.5.08

A Novel Soft Set Approach for Feature Selection

Daoli Yang[a], Zhi Xiao[a,b,1,2], Wei Xu[a], Xianning Wang[a] and Yuchen Pan[a]

[a] *School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China*

[b] *Chongqing Key Laboratory of Logistics, Chongqing University, Chongqing 400044, PR China*

xiaozhicqu@163.com

Abstract

Feature selection is an important preprocess for data mining. The soft set theory is a new mathematical tool to deal with uncertainties. Merging the soft set theory into the feature selection process based on rough sets facilitates the computation with equivalence classes and improves the efficiency. We propose a paired relation soft set model based on equivalence classes of the information system. Then we use it to present the lower approximate set in the form of soft sets and calculate the degree of dependency between relations. Furthermore, we give a new mapping to obtain equivalence classes of indiscernibility relations and propose a feature selection algorithm based on the paired relation soft set model. Compared to the algorithm based NSS, this algorithm shows 18.17% improvement on an average. Meantime, both of the algorithms show a good scalability.

Keywords*: soft set, feature selection, rough set, information system*

1. Introduction

Feature selection is an important issue in data mining. In the big data era, the scale of databases grows rapidly with an increasing number of features due to availability of more detailed information. The existence of irrelative features is inevitable. Feature selection can eliminate those features so that it reduces the dimensionality, improves the predictive accuracy and enhances comprehensibility of the induced concepts [1]. Besides, features may have approximate relationships with each other. Using a subset of features to represent other features and deduce the classification results is an imprecise process. The rough set theory is a powerful tool to process such kind of imprecision [2-4]. Thus, it is incorporated into feature selection algorithms. However, the computation with equivalence classes in these algorithms is inconvenient and hard to understand. The soft set theory can facilitate the computation and improve the efficiency.

Recent literatures of rough set based feature selection algorithms mainly focus on three aspects: feature evaluation, search strategies and practical application. Feature evaluation provides the significance of the candidate features which is represented by dependency degree in the rough set theory. For different nature of data, calculation methods of the dependency degree are different. Several types of data have been studied, such as fuzzy data [5], data with noise [6], dynamic incomplete data [7], hybrid data [8], *etc.* As for search strategies, there are commonly used methods like sequential forward selection and sequential backward elimination [6], as well as many heuristic methods which are hybrid genetic algorithm [9], granular neural network [5], water drops algorithm [10], *etc.* Some literatures explore feature selection based on rough set for practical application in

[1] Corresponding Author, email: xiaozhicqu@163.com
[2] Supported by the National Science Foundation of China (Grant No. 71171209)

domains like computer security [11] and assessment of power systems [12]. But researches focusing on the improvement in efficiency when calculating the dependency degree of features are few. This kind of improvement is also crucial in improving performance of rough set based feature selection.

The soft set theory was proposed by D. MOLODTSOV in 1999 [13]. It offers a rigid mathematical theory to deal with uncertainties. Unlike the probability theory and the fuzzy set theory suffering from the inadequacy of the parameterization tools, the soft set theory is free of those problems. This makes the theory easily applicable in practice [13] and very convenient to combine with other theories. In recent years, studies about the soft set theory are growing rapidly in many fields such as modern algebra [14], forecasting [15], decision making [16], data analysis [17], *etc*. It has been proved that every rough set is a soft set, which shows the necessity to combine two theories together [18]. As for the dependency degree in the rough set theory, Qin *et al*. [19] had found an efficient way to calculate it by constructing a soft set model (NSS) for an information system. Later, Mamat *et al*. [20] gave a new clustering method of attributes based on the same soft set model. Both of them suggest that the method based on the soft set theory is efficient and easy to apply. However, there are two remaining issues they don't tackle. One is how to get equivalence classes of the indiscernibility relation of features based on the soft set theory and the other is how to apply the soft set model to feature selection which is different from feature cluster.

Our main contributions are summarized as follows:

(a) We propose a paired relation soft set model (PRSS) based on the information system. We use it to construct the lower approximate set, present the approximate set in the form of soft set and furthermore calculate the dependency degree between features.

(b) We propose a method to acquire equivalence classes of indiscernibility relations based on the PRSS.

(c) We present a feature selection algorithm based on the PRSS and apply it to seventeen UCI benchmark data sets. The result shows improvement in efficiency compared to the method based on NSS.

The rest of our paper contains four sections. Section 2 presents the fundamental theory. Section 3 proposes the paired relation soft set model, a method to acquire equivalence classes of indiscernibility relations and the relative feature selection algorithm. Section 4 compares experimental results with the NSS method. The final section concludes our work.

2. Essential Rudiments

2.1. Rough Set Theory

Definition 1 (Information system [4]): Suppose there is a pair $\mathcal{A}=(U,\mathbb{A})$ of non-empty, finite sets U and $\mathbb{A}$, where U is the universe set of objects, and $\mathbb{A}$ is a set consisting of attributes, i.e. functions $a: U \to V_a$, where V_a is the set of values of attribute a, called the domain of a. The pair $\mathcal{A}=(U,\mathbb{A})$ is called an *information system*.

Sometimes, we distinguish in an information system $\mathcal{A}=(U,\mathbb{A})$ a partition of A into two classes C, $D \subseteq A$ of attributes, called condition and decision (action) attributes, respectively. The tuple $\mathcal{A} = (U, C, D)$ called a decision system [4].

Definition 2 (Indiscernibility relation [3]): In an information system $\mathcal{A}$, if $\mathbb{P} \subseteq \mathbb{A}$ and $\mathbb{P} \neq \emptyset$, then $\bigcap \mathbb{P}$ (intersection of all equivalence relations belonging to $\mathbb{P}$), denoted by $IND(\mathbb{P})$, will be called an *indiscernibility relation* over $\mathbb{P}$.

The equivalence class of $IND(\mathbb{P})$ involving element x is denoted by $[x]_{\mathrm{IND}(\mathbb{P})} = \bigcap_{r\in\mathbb{P}}[x]_r$. The $U/IND\,(\mathbb{P})$ denotes the family of all equivalence classes of

$IND(\mathbb{P})$. Meantime we denote the family of all equivalence relations defined in $\mathcal{A}$ as $IND(\mathcal{A}) = \{IND(\mathbb{P}): \emptyset \neq \mathbb{P} \subseteq \mathbb{A}\}$.

Definition 3 (The lower and upper approximations [3]): With each subset $X \subseteq U$ and an equivalence relation $\mathbb{r} \in IND(\mathcal{A})$, we associate two subsets:

$$\begin{cases} \underline{\mathbb{r}}X = \bigcup\{x \in U: [x]_{\mathbb{r}} \subseteq X\}, \\ \overline{\mathbb{r}}X = \bigcup\{x \in U: [x]_{\mathbb{r}} \cap X \neq \emptyset\}. \end{cases} \tag{1}$$

Called the $\mathbb{r}$-lower and $\mathbb{r}$-upper approximations of X respectively.

Definition 4 (Degree of partial dependency [3]): Let $\mathcal{A}$=$(U, \mathbb{A})$ be an information system and $\mathbb{P}, \mathbb{Q} \subset \mathbb{A}$. We say that $\mathbb{Q}$ depends in a degree $k(0 \leq k \leq 1)$ from $\mathbb{P}$, if and only if

$$k_{\mathbb{P}}(\mathbb{Q}) = \frac{|\mathrm{POS}_{IND(\mathbb{P})}IND(\mathbb{Q})|}{|U|} \tag{2}$$

Where $|U|$ denotes the cardinality of the set U and the $IND(\mathbb{P})$-positive region of $IND(\mathbb{Q})$ is

$$\mathrm{POS}_{IND(\mathbb{P})}(IND(\mathbb{Q})) = \bigcup_{X \in U/IND(\mathbb{Q})} \underline{IND(\mathbb{P})}\, X. \tag{3}$$

Dependency degree shows the importance of $IND(\mathbb{P})$ for $IND(\mathbb{Q})$.

2.2. Soft Set Theory

Let U be the initial universe set and E be a set of parameters.

Definition 5 (Soft set [13]): A pair of (F, E) is called a soft set (over U) if and only if F is a mapping of E into the set of all subset of the set U.

In other words, the soft set is a parameterized family of subsets of the set U. Every set $F(\varepsilon)$, $\varepsilon \in E$, from this family may be considered as the set of ε-elements of the soft set (F, E), or as the set of ε- approximate elements of the soft set [13].

Definition 6 (Soft subset [21]): For two soft sets (F, A) and (G, B) over a common universe set U, the (F, A) is a soft subset of (G, B) if $A \subset B$ and $\forall \varepsilon \in A$, $F(\varepsilon)$ and $G(\varepsilon)$ are identical approximations. It is denoted as $(F, A) \tilde{\subset} (G, B)$.

Apparently, for a set of ε- approximate elements of (G, B), we can define a corresponding soft set (G, ε) which has only one parameter ε. $(G, \varepsilon) \tilde{\subset} (G, B)$.

Let (F, A) and (G, B) be soft sets over U, then operation * for soft sets [13] is

$$(F, A) * (G, B) = (H, A \times B), \tag{4}$$

Where $H(\alpha, \beta) = F(\alpha) * G(\beta)$, $\alpha \in A$, $\beta \in B$, and $A \times B$ is the Cartesian product of the sets A and B.

Example 1: There are five houses $U = (h_1, h_2, h_3, h_4, h_5)$ which are ready to be evaluated by three criterions $E = \{\text{expensive}(\varepsilon_1), \text{beautiful}(\varepsilon_2), \text{wooden}(\varepsilon_3)\}$. Suppose $F(\varepsilon_1) = \{1,2,3\}$,, $F(\varepsilon_2) = \{3,4,5\}$ and $F(\varepsilon_3) = \{5\}$. Then we get the soft set

$$(F, E) = \{(\{1,2,3\}, \varepsilon_1), (\{3,4,5\}, \varepsilon_2), (\{5\}, \varepsilon_3)\}. \tag{5}$$

If we want to know which houses are both expensive and beautiful, we get (F, ε_1), $(F, \varepsilon_2) \tilde{\subset} (F, E)$, and execute the intersection operation as follows

$$(F, \varepsilon_1) \cap (F, \varepsilon_2) = (F(\varepsilon_1) \cap F(\varepsilon_2), \varepsilon_1 \times \varepsilon_2) \tag{6}$$

That is $(\{3\}, \varepsilon_1 \times \varepsilon_2)$.

2.3. A Soft Set Model on Equivalence Class

Computation in the rough set theory always involves intersection, union of the sets of equivalence classes and the dependency degree between attributes. It is inconvenient to execute these operations directly on the information system. Thus, Qin *et al*. [19] constructed a soft set model over equivalence classes to facilitate the computation.

Given an information system $\mathcal{A}$=$(U, \mathbb{A})$, $U/\mathbb{A}$ denotes the set of all equivalence classes in the partitions U/r_i, where $r_i \in \mathbb{A}$. Let U'=$U/\mathbb{A}$ be the initial universe set of objects and E= $U/\mathbb{A}$ be the set of parameters. $P(U')$ denotes the power set of U'. Meantime, define mapping $F: E \to P(U')$. The pair(F, E) is a soft set model over equivalence classes (NSS) [19]. Table 1 shows the tabular representation of (F, E), where $x_i = e_i$ and $F(e_i)(x_j)$ is either 1 or 0, $i = 1,2, \cdots, m$, $j = 1,2, \cdots, m$.

Table 1. The Tabular Representation of the Soft Set over Equivalence Classes

$U' \backslash E$	e_1	e_2	…	e_m
x_1	$F(e_1)(x_1)$	$F(e_2)(x_1)$	…	$F(e_m)(x_1)$
x_2	$F(e_1)(x_2)$	$F(e_2)(x_2)$	…	$F(e_m)(x_2)$
…	…	…	…	…
x_m	$F(e_1)(x_m)$	$F(e_2)(x_m)$	…	$F(e_m)(x_m)$

Then they defined two soft set (F_1, E) and (F_2, E) where $F_1, F_2: E \to P(U')$ and

$$\begin{cases} F_1(e) = \{A | A \in U' \text{ and } A \subset e\}, \\ F_2(e) = \{A | A \in U' \text{ and } A \cap e \neq \emptyset\}. \end{cases} \tag{7}$$

The lower and upper approximations of parameter e_j with respect to attribute r_i are defined as

$$\begin{cases} \underline{r_i}(e_j) = \{x | x \in x_k \text{ and } F_1(e_j)(x_k) = 1, k = D+1, \cdots, D + |U/r_i|\} \\ \bar{r_i}(e_j) = \{x | x \in x_k \text{ and } F_2(e_j)(x_k) = 1, k = D+1, \cdots, D + |U/r_i|\} \end{cases}, \tag{8}$$

where $D = \sum_{l=1}^{i-1} |U/r_l|$.

The lower and upper approximations of attribute r_j with respect to attribute r_i are defined as

$$\begin{cases} \underline{r_i}(r_j) = \{x | x \in \underline{r_i}(e_k), k = D+1, \cdots, D + |U/r_j|\} \\ \bar{r_i}(r_j) = \{x | x \in \bar{r_i}(e_k), k = D+1, \cdots, D + |U/r_j|\} \end{cases} \tag{9}$$

where $D = \sum_{l=1}^{i-1} |U/r_l|$.

In the following, we just present an example for soft set (F_1, E). (F_2, E) is similar.

Example 2: we consider the information system of house evaluation as shown in Table 2.

Then, Then they

The lower and upper approximations of parameter e_j with respect to attribute r_i are defined

h

The lower and upper approximations of attribute r_j with respect to attribute r_i are

he

In the following, we just present an example for soft set (F_1, E). (F_2, E) is similar.

Example 2: we consider the information system of house evaluation as shown in Table 2.

partitions of each attribution are $U/\text{price} = \{\{h_1\}, \{h_2, h_3\}\}$, $U/\text{appearance} = \{\{h_1, h_2\}, \{h_3\}\}$ and $U/\text{material} = \{\{h_1\}, \{h_2\}, \{h_3\}\}$. The soft set (F_1, E) is shown in Table 3.

For e_3, $e_4 \in U/\text{appearance}$, and x_1, $x_2 \in U/\text{price}$, $\underline{r_{price}}(e_3)$= $\{h_1\}$ and $\underline{r_{price}}(e_4)$= ϕ. Thus, $\underline{r_{price}}(r_{appearance})$= $\{h_1\}$.

Table 2. An Information System of House Evaluation

	price	appearance	material
h_1	expensive	beautiful	wood
h_2	cheap	beautiful	metal
h_3	cheap	ugly	brick

Table 3. The Tabular Representation of Soft Set (F_1, E) Based on NSS

	$e_1\{h_1\}$	$e_2\{h_2,h_3\}$	$e_3\{h_1,h_2\}$	$e_4\{h_3\}$	$e_5\{h_1\}$	$e_6\{h_2\}$	$e_7\{h_3\}$
$x_1\{h_1\}$	1	0	1	0	1	0	0
$x_2\{h_2,h_3\}$	0	1	0	0	0	0	0
$x_3\{h_1,h_2\}$	0	0	1	0	0	0	0
$x_4\{h_3\}$	0	1	0	1	0	0	1
$x_5\{h_1\}$	1	0	1	0	1	0	0
$x_6\{h_2\}$	0	1	1	0	0	1	0
$x_7\{h_3\}$	0	1	0	1	0	0	1

3. Proposed Paired Relation Soft Set Technique

3.1. Paired Relation Soft Set

Given an information system $\mathcal{A}=(U, \mathbb{A})$ and two relations $r_i, r_j \in IND(\mathcal{A})$, $i \neq j$. When examining the upper and lower approximation sets of r_i with respect to r_j, it is unnecessary to include all the equivalence classes of them into U'. As shown in Table 3, computation in cells with color is redundant. Thus, we change the initial universe set U' and parameter E of NSS.

Definition 7 (Paired relation soft set): Let $U' = U/r_i$ and $E = U/r_j$. Define the mapping $F: E \to P(U')$. A pair of (F, E) is called paired relation soft set for its universe set and parameters stemming from two relations r_i, and r_j, respectively.

Table 4. The Tabular Representation of PRSS

	appearance		
price	$u'\backslash E$	$e_1\{h_1,h_2\}$	$e_2\{h_3\}$
	$x_1\{h_1\}$	$F(e_1)(x_1)$	$F(e_2)(x_1)$
	$x_2\{h_2,h_3\}$	$F(e_1)(x_2)$	$F(e_2)(x_2)$

Example 3: For example 2, let $r_1 = \text{appearance}$ and $r_2 = \text{price}$. $U/r_1 = \{\{h_1,h_2\},\{h_3\}\}$, $e_1 = \{h_1,h_2\}$ and $e_2 = \{h_3\}$. $U/r_2 = \{\{h_1\},\{h_2,h_3\}\}$, $x_1 = \{h_1\}$ and $x_2 = \{h_2,h_3\}$.

The tabular representation of PRSS for the appearance and price is shown in Table 4. Then we can construct the soft sets (F_1, E) and (F_2, E), where mappings F_1 and F_2 are in the formula 7. The tabular representation of soft set (F_1, E) and(F_2, E) for appearance and price in Example 2 is shown in Table 5.

Table 5. The Tabular Representation of Soft Sets (F_1, E) and (F_2, E) Based on PRSS

(F_1, E)	appearance			(F_2, E)	appearance		
price	$u'\backslash E$	$e_1\{h_1, h_2\}$	$e_2\{h_3\}$	price	$u'\backslash E$	$e_1\{h_1, h_2\}$	$e_2\{h_3\}$
	$x_1\{h_1\}$	1	0		$x_1\{h_1\}$	1	0
	$x_2\{h_2, h_3\}$	0	0		$x_2\{h_2, h_3\}$	1	1

3.2. Computation of Dependency Degree Based on PRSS

According to formula 2, the computation of the dependency degree requires the lower approximate set and the positive region. We firstly give a soft set with two parameter sets. Then, based on that soft set and (F_1, E), we present the soft set representation of the lower approximate set and the computation of dependency degree.

A soft set with two parameter sets is shown as follows. Given an information system $\mathcal{A}=(U, \mathbb{A})$ and two relations $r_i, r_j \in IND(\mathcal{A})$, $i \neq j$. Let $E_1 = U/r_j$, $E_2 = U/r_i$. Define mapping $\Psi: E_1 \times E_2 \rightarrow P(U)$. We get soft set $(\Psi, E_1 \times E_2)$.

Meantime, let $R_1, R_2 = \mathrm{IND}(\mathcal{A})$, $r_j \in R_1$, $r_i \in R_2$, $e_t \in E_1$ and $x_k \in E_2$.

The lower approximation set of parameter e_t with respect to x_k is represented by soft set $(L_c, E_1 \times E_2)$ where

$$L_c(e_t, x_k) = \{x | x \in x_k \text{ and } F_1(e_t)(x_k) = 1\}, \tag{10}$$

$F_1(e_t)(x_k)$ is as shown in Table 5. Apparently, $(L_c, e_t \times x_k) \subset (L_c, E_1 \times E_2)$.

The lower approximation of parameter e_t with respect to r_i is represented by soft set $(L_r, E_1 \times R_2)$ where

$$L_r(e_t, r_i) = \{x | x \in x_k\ ,\ F_1(e_t)(x_k) = 1 \ and\ x_k \in U/r_i\}. \tag{11}$$

Then, $(L_r, e_t \times r_i) \subset (L_r, E_1 \times R_2)$.

Compare $(L_r, e_t \times r_i)$ with $(L_c, e_t \times x_k)$, we have

$$(L_r, e_t \times r_i) = \cup_{x_k \in U/r_i} (L_c, e_t \times x_k) = (\cup_{x_k \in U/r_i} L_c(e_t, x_k), e_t \times \underset{x_k \in U/r_i}{\times} x_k) \tag{12}$$

where $\underset{x_k \in U/r_i}{\times} x_k = x_1 \times x_2 \times \cdots \times x_{|U/r_i|}$.

The lower approximation of relation r_j with respect to r_i is represented by soft set $(L_R, R_1 \times R_2)$

$$L_R(r_j, r_i) = \{x | x \in x_k\ ,\ F_1(e_t)(x_k) = 1, x_k \in U/r_i \ and\ e_t \in U/r_j\}. \tag{13}$$

Then, $(L_R, r_j \times r_i) \subset (L_R, R_1 \times R_2)$.

Compare $(L_R, r_j \times r_i)$ and $(L_r, e_t \times r_i)$, we get,

$$(L_R, r_j \times r_i) = \cup_{e_t \in U/r_j} (L_r, e_t \times r_i) = \cup_{e_t \in U/r_j} \cup_{x_k \in U/r_i} (L_c, e_t \times x_k) =$$
$$(\cup_{e_t \in U/r_j} \cup_{x_k \in U/r_i} L_c(e_t, x_k), \underset{e_t \in U/r_j}{\times} e_t \times \underset{x_k \in U/r_i}{\times} x_k) \tag{14}$$

where $\underset{e_t \in U/r_j}{\times} e_t = e_1 \times e_2 \times \cdots \times e_{|U/r_j|}$.

Apparently, soft set $(L_R, r_j \times r_i)$ is the r_i-positive region of r_j (see formula 2 and 3).

To get the dependency degree, we also need the cardinality of $(L_R, r_j \times r_i)$.

$$|(L_R, r_j \times r_i)| = |\cup_{e_t \in U/r_j} \cup_{x_k \in U/r_i} L_c(e_t, x_k)|. \tag{15}$$

However, if we only want to get the dependency degree, it is unnecessary to calculate $(L_R, r_j \times r_i)$ first. In the following, we provide a direct way to get the dependency degree.

Based on soft set (F_1, E)(see table 5), the cardinality of $(L_c, e_t \times x_k)$ is

$$|(L_c, e_t \times x_k)| = F_1(e_t)(x_k) \times |e_t|. \tag{16}$$

The cardinality of $(L_r, e_t \times r_i)$ is

$$|(L_r, e_t \times r_i)| = \sum_{x_k \in U/r_i} |L_c, e_t \times x_k| = \sum_{x_k \in U/r_i} (F_1(e_t)(x_k) \times |e_t|), \tag{17}$$

The cardinality of $(L_R, r_j \times r_i)$ is

$$|(L_R, r_j \times r_i)| = \sum_{e_t \in U/r_j} |(L_r, e_t \times r_i)| = \sum_{e_t \in U/r_j} \sum_{x_k \in U/r_i} |(L_c, e_t \times x_k)| =$$
$$\sum_{e_t \in U/r_j} \sum_{x_k \in U/r_i} (F_1(e_t)(x_k) \times |e_t|). \tag{18}$$

The degree of dependency of relation r_j with respect to r_i is

$$k_{r_i}(r_j) = \frac{|(L_R, r_j \times r_i)|}{|U|}. \tag{19}$$

Given a decision system $\mathcal{A} = (U, C, D)$, $\forall B \subset C$, the positive region of D on B is represented by

$$(L_R, D \times IND(B)) \tag{20}$$

The dependency degree of decision D on B is represented by

$$k_{IND(B)}(D) = \frac{|(L_R, D \times IND(B))|}{|U|}. \tag{21}$$

Example 4: we present the low approximate set and the degree of dependency for r_1 (appearance) and r_2 (price) in Example 3 based on (F_1, E) in Table 5.

$L_c(e_1, x_1) = (\{h_1\}, e_1 \times x_1)$, $L_c(e_1, x_2) = (\phi, e_1 \times x_2)$, $L_c(e_2, x_1) = (\phi, e_2 \times x_1)$, $L_c(e_2, x_2) = (\phi, e_2 \times x_2)$.

$(L_r, e_1 \times r_2) = \bigcup_{x_k \in U/r_2} (L_c, e_1 \times x_k) = L_c(e_1, x_1) \cup L_c(e_1, x_2) = (\{h_1\}, e_1 \times x_1 \times x_2)$,

$(L_r, e_2 \times r_2) = \bigcup_{x_k \in U/r_2} (L_c, e_2 \times x_k) = L_c(e_2, x_1) \cup L_c(e_2, x_2) = (\phi, e_2 \times x_1 \times x_2)$,

$(L_R, r_1 \times r_2) = \bigcup_{e_t \in U/r_1} (L_r, e_t \times r_2) = (L_r, e_1 \times r_2) \cup (L_r, e_2 \times r_2) = (\{h_1\}, e_1 \times e_2 \times x_1 \times x_2)$

.

The lower approximation of relation r_1 with respect to r_2 is $\{h_1\}$.

For the calculation of cardinality,

$|L_c(e_1, x_1)| = 1 \times 1 = 1$, $|L_c(e_1, x_2)| = 0 \times 2 = 0$, $|L_c(e_2, x_1)| = 0$, $|L_c(e_2, x_2)| = 0$.

$|(L_r, e_1 \times r_2)| = \sum_{x_k \in U/r_2} |(L_c, e_1 \times x_k)| = |L_c(e_1, x_1)| + |L_c(e_1, x_2)| = 1$,

$|(L_r, e_2 \times r_2)| = \sum_{x_k \in U/r_2} |(L_c, e_2 \times x_k)| = |L_c(e_2, x_1)| + |L_c(e_2, x_2)| = 0$,

$|(L_R, r_1 \times r_2)| = \sum_{e_t \in U/r_1} |L_r, e_t \times r_2| = |(L_r, e_1 \times r_2)| + |(L_r, e_2 \times r_2)| = 1$.

Then the dependency degree of relation r_1 with respect to r_2 is $k_{r_2}(r_1) = \frac{1}{3}$.

3.3. Computation of Equivalence Classes of Indiscernibility Relations

We first introduce a new mapping F_3 as a supplement to mappings in formula 7. Then, represent equivalence classes of the indiscernibility relation formed by two relations using a soft set based on F_3.

Given a PRSS (F, E') where the universal set $U' = U/r_i$, parameter set $E' = U/r_j$ and $r_j, r_i \in \mathrm{IND}(\mathcal{A})$, $i \neq j$, we propose Definition 8.

Definition 8: Let $P(U)$ be the power set of U, $G_1 = E'$ and $G_2 = U'$.We define mapping $F_3: G_1 \times G_2 \to P(U)$, for

$$F_3(A, e) = \{x | x \in A \cap e \text{ and } A \in G_2, e \in G_1\}. \tag{22}$$

The soft set based on mapping F_3 is denoted by $(F_3, G_1 \times G_2)$. Let $e_t \in G_1$ and $x_k \in G_2$, $(F_3, e_t \times x_k) \tilde{\subset} (F_3, G_1 \times G_2)$.

Suppose $U^{''} = P(U)$, $R_1, R_2 = \mathrm{IND}(\mathcal{A})$, $r_j \in R_1$ and $r_i \in R_2$. Define mapping $I: R_1 \times R_2 \to P(U^{''})$. The equivalence classes of the indiscernibility relation formed by r_j and r_i are represented by $(I, R_1 \times R_2)$ where

$$I(r_j, r_i) = \{F_3(A, e) | A \in U/r_i, e \in U/r_j \text{ and } F_3(A, e) \neq \phi\}. \quad (23)$$

Then $(I, r_j \times r_i) \tilde{\subset} (I, R_1 \times R_2)$.

Example 5: For r_1 (appearance) and r_2 (price) in example 3, the tabular representation of $(F_3, G_1 \times G_2)$ is shown in Table 6. Then, $(I, r_1 \times r_2) = (\{\{h_1\}, \{h_2\}, \{h_3\}\}, r_1 \times r_2)$.

Table 6. The Tabular Representation of $(F_3, G_1 \times G_2)$

$(F_3, G_1 \times G_2)$	appearance		
price	$G_2 \backslash G_1$	$e_1\{h_1, h_2\}$	$e_2\{h_3\}$
	$x_1\{h_1\}$	$\{h_1\}$	ϕ
	$x_2\{h_2, h_3\}$	$\{h_2\}$	$\{h_3\}$

3.4. Feature Selection Algorithm Based on PRSS

Given a decision system $\mathcal{A} = (U, C, D)$, the feature selection algorithm based on PRSS is shown in Table 7.

The algorithm selects features according to the dependency degree of decision D provided by Algorithm 2 and uses the sequential forward selection [6] as the search strategy provided by Algorithm 1. The output is a feature ranking $F' = \{f_1', f_2', \cdots, f_{|F'|}'\}$.

Table 7. The Feature Selection Algorithm Based on PRSS

	Algorithm 1		Algorithm 2	
Input	X, F	X is a sample set and $F = C$	F', f, D	F', f and D are in step 2 of algorithm 1
Output	F	F is a feature ranking	$k_{IND(F' \cup \{f\})}(D)$	k is the dependency degree

	Begin
	Step 1: Initialize $F' = \emptyset$
	While $F \neq \emptyset$
	Step 2: Find
	$f = \arg_f \max_{f \in F}\{k_{IND(F' \cup \{f\})}(D)\}$
	Step 3: $F' = F' \cup \{f\}$
	Step 4: $F = F - \{f\}$
	End
	Return F'
	End
			Begin
			If $F' = \emptyset$, then
			Construct (F_1, E) of D and f based on PRSS
			Calculate $k = \|(L_R, D \times f)\|/\|U\|$
			Else
			Construct soft set $(F_3, F' \times f)$ based on PRSS
			Calculate $IND(F' \cup \{f\}) = (I, F' \times f)$
			Construct (F_1, E) of D and $IND(F' \cup \{f\})$ based on PRSS
			Calculate $k = \|(L_R, D \times IND(F' \cup \{f\})\|/\|U\|$
			End
			End	

Given the set F_{k-1}' with $k-1$ features selected, the kth feature is determined by

$$\max_{f \in F - F_{k-1}'} \{k_{IND(F' \cup \{f\})}(D)\}. \quad (24)$$

Based on the ranking, we can get a feature ranking soft set (f_c, C), where $C = \{1,2,\cdots,|F'|\}$ and the f_c is defined as

$$f_c(c) = \{\{f_1', f_2', \cdots, f_c'\} | f_i' \in F', c \in C\}. \tag{25}$$

Next, we use KNN classifier to cross-validate the classification accuracy of the data with these feature subsets. Those with the highest classification accuracy are the final feature subsets.

Table 8. The Algorithm for the Degree of Dependency Based on NSS

	Algorithm 3	
Input	F', f, D	F', f and D are in step 2 in table
Output	$k_{IND(F' \cup \{f\})}(D)$	k is the dependency degree

	Begin
	If $F' = \emptyset$, then
	Construct $(F_1, E)_{\text{NSS}}$ of D and f based on NSS.
	Calculate $k_f(D) = \lvert\underline{f}(D)\rvert / \lvert U\rvert$ according to formula 9.
	Else
	Construct soft set $(F_3, N_1 \times N_2)_{\text{NSS}}$ based on NSS.
	Calculate $IND(F' \cup \{f\}) = (I, F' \times f)$.
	Construct $(F_1, E)_{\text{NSS}}$ of D and $IND(F' \cup \{f\})$ based on NSS.
	Calculate $k_{IND(F' \cup \{f\})}(D) = \lvert\underline{IND(F' \cup \{f\})}(D)\rvert / \lvert U\rvert$.
	End
	End	

3.5. Feature Selection Algorithm Based on NSS

Feature selection algorithm based on NSS still uses Algorithm 1 as the search strategy. But the Algorithm 3 in Table 8 which provides the dependency degree of decision D is different from Algorithm 2.

Let N_1 and N_2 be the set of all equivalence classes in the partitions U/F' and U/f.

Example 6: Suppose F' = appearance and f = price. $N_1 = \{e_1\{h_1\}, e_2\{h_2, h_3\}, e_3\{h_1, h_2\}, e_4\{h_3\}\}$, $N_2 = x_1\{h_1\}, x_2\{h_2, h_3\}, x_3\{h_1, h_2\}, x_4\{h_3\}$. Then $(F_3, N_1 \times N_2)_{\text{NSS}}$ is shown in Table 9. The construct of $(F_1, E)_{\text{NSS}}$ follows the same logic of Table 3.

Table 9. The Tabular Representation of $(F_3, N_1 \times N_2)_{\text{NSS}}$

$(F_3, N_1 \times N_2)_{\text{NSS}}$	price			appearance	
price	$N_2 \backslash N_1$	$e_1\{h_1\}$	$e_2\{h_2, h_3\}$	$e_3\{h_1, h_2\}$	$e_4\{h_3\}$
	$x_1\{h_1\}$	$\{h_1\}$	Φ	$\{h_1\}$	Φ
	$x_2\{h_2, h_3\}$	Φ	$\{\{h_2, h_3\}$	$\{h_2\}$	$\{h_3\}$
appearance	$x_3\{h_1, h_2\}$	$\{h_1\}$	$\{h_2\}$	$\{h_1, h_2\}$	Φ
	$x_4\{h_3\}$	Φ	$\{h_3\}$	Φ	$\{h_3\}$

4. Experimental Results

We compare the execute time of Algorithm 1 and 2 based on PRSS with that of Algorithm 1 and 3 basd on NSS using seventeen datasets abtained from the benchmark UCI Machine Learning Repository [22]. Those datasets are discribed in Table 10. The field named "Number of Classes" in this table shows the aggregation of the number of equvalence classes formed by each feature.

We run the algorithms in MATLAB R2015a on a pc with a processor Intel(R) Core(TM) i5-2410M, a 8.0GB memory and Widows 7 operating system. Figure 1 shows datasets on which ranges of the executive time are (0 , 0.1) and (0.1, 1) in seconds. Figure 2 shows datasets on which ranges of the executive time are (1, 20) and (20, 120) in seconds.

Table 11 summrizes the comparison of feature selection algorithm based on PRSS and NSS. It shows that PRSS improves efficiency to a certain degree on each dataset and achieves 18.17% improvement on an average.

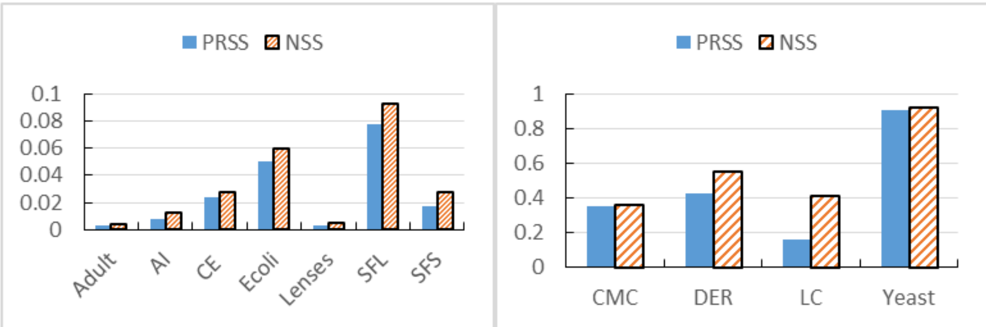


Figure 1. Excutive Time (Second) of Feature Selection Based on PRSS and NSS

Table 10. Description of Datasets

No	Data sets	Number of instances	Number of features	Number of Classes
1	Abalone	4177	9	6077
2	Adult	20	4	10
3	AI(Acute Inflammations)	120	8	58
4	CE(Car Evaluation)	1728	7	25
5	Chess(King-Rook vs. King-Pawn)	3196	37	75
6	CMC(Contraceptive Method Choice)	1473	10	74
7	DER(Dermatology)	366	34	135
8	Ecoli	336	9	707
9	LC(Lung Cancer)	32	55	154
10	Lenses	24	6	36
11	Mushroom (MR)	8124	23	119
12	Musk	476	169	25760
13	Nursery	12960	9	32
14	SFL(Solar Flare large)	1066	13	48
15	SFS(Solar Flare small)	323	13	40
16	Statlog (statlog_satimage)	4435	37	2752

17	Yeast	1484	10	1884

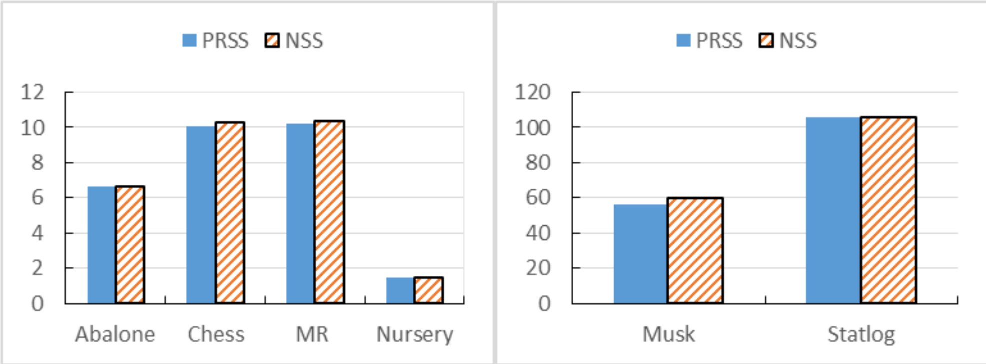


Figure 2. Excutive Time (Second) of Feature Selection Based on PRSS and NSS

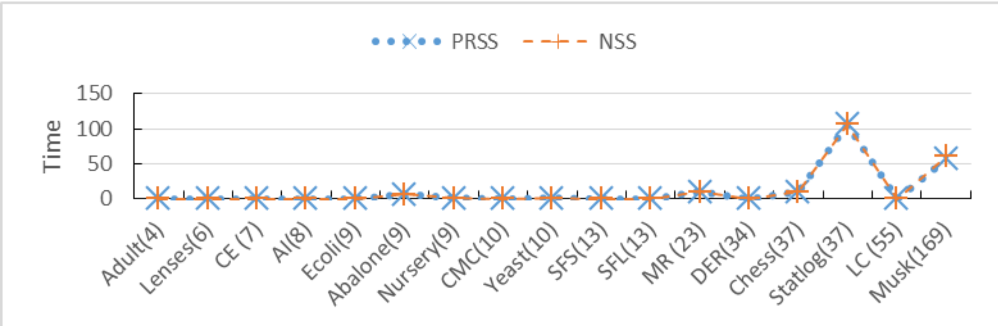


Figure 3. The Scalability of PRSS and NSS to the Number of Features

Table 11. Comparison of Feature Selection Algorithm of PRSS and NSS

No	Data sets	Executing time (s)		Improvement
		NSS	PRSS	
1	Abalone	6.645	6.620825	0.37%
2	Adult	0.0047	0.0032	33.50%
3	AI(Acute Inflammations)	0.0128	0.008	37.80%
4	CE(Car Evaluation)	0.0282	0.0239	15.40%
5	Chess(King-Rook vs. King-Pawn)	10.3424	10.0919	2.42%
6	CMC(Contraceptive Method Choice)	0.3662	0.3511	4.10%
7	DER(Dermatology)	0.5524	0.4227	23.49%
8	Ecoli	0.06	0.0506	15.68%
9	LC(Lung Cancer)	0.4183	0.1632	60.99%
10	Lenses	0.0052	0.0027	47.93%
11	Mushroom	10.42	10.2518	1.61%
12	Musk	60.4035	56.4015	6.63%
13	Nursery	1.5058	1.4757	2.00%
14	SFL(Solar Flare large)	0.093	0.078	16.09%
15	SFS(Solar Flare small)	0.0283	0.0173	38.95%
16	Statlog(statlog_satimage)	106.3707	105.8892	0.45%
17	Yeast	0.9246	0.9113	1.44%
Average improvement				18.17%

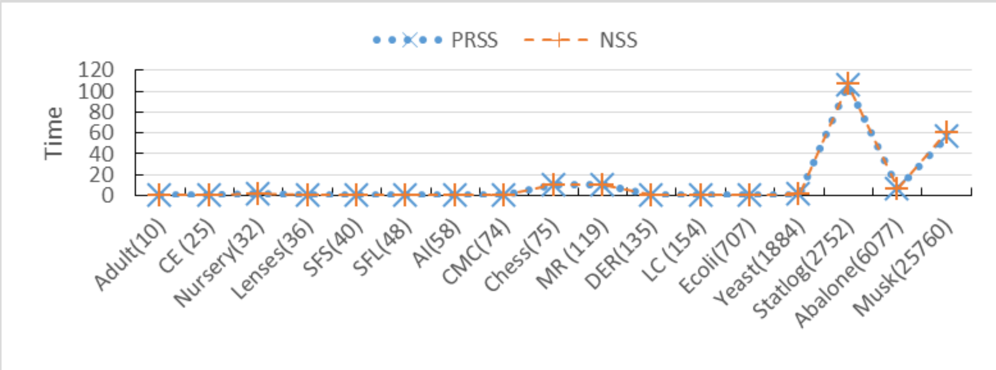


Figure 4. The Scalability of PRSS and NSS to the Number of Instances

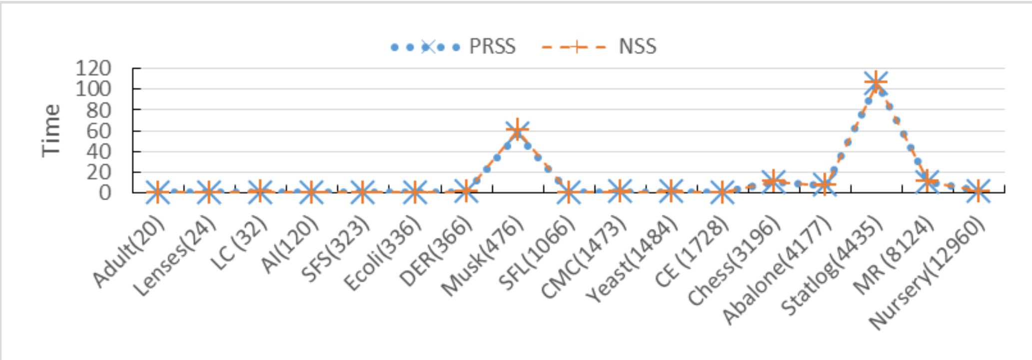


Figure 5. The Scalability of PRSS and NSS to the Number of Equvalence Classes

We also analyze the scalability of both algorithms on three aspects, the number of features, the number of instances and the number of equvalence classes. Figure 3 shows the execute time of both algorithms with regard to the number of features in each dataset. Figure 4 shows the execute time of both algorithms with regard to the number of instances in each dataset. We also provide the change of the execute time with regard to the aggregation of the number of equvalence classes formed by each feature in Figure 5. In general, three figures show the execute time increase linearly as the number of features, instances and equvalence classes increase, respectively. Though the execute time varies at several points, it is at an acceptabe level. Thus, those three figures demonstrate a good scalibility of both feature selection algorithms.

5. Conclusion

Merging the soft set theory into some core conceptions of the rough set theory facilitates the computation based on equvalence classes. We proposed a paired relation soft set model to construct the lower approximate set, presented the lower approximate set in the form of soft sets, and furthermore calculated the dependency degree of features. We also provided a new mapping to calculate equivalence classes of indiscernibility relations. Based on PRSS, the feature selection algorithm was given. Compared to the feature selection algorithm based on NSS, the PRSS algorithm shows 18.17% improvement on an average. Experimental results also show a good scalibility of both feature selection algorithms. Theoretically, the paper shows a way to combine the soft set theory and the rough set theory together to improve the feature selection algorithm. Practically, the novel method based on PRSS can accelerate data processing procedure and is helpful in dealing with big data.

In this paper, we only combined the soft set theory with the classic rough set model which is based on equivalence classes. To deal with data with fuzzy attributes, a fuzzy rough set model based on fuzzy similarity relations is more suitable. How to apply PRSS into a fuzzy rough set model and facilitate the computation will be a topic of furture reasearches.

Acknowledgment

We are very grateful to referees and editors. Their comments are very helpful in improving the paper. We also appreciate the research team members.

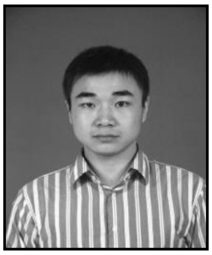
References

[1] L. H. and M. H, "Feature Selection for Knowledge Discovery and Data Mining", Kluwer, Boston, **(1998)**.
[2] Z. Pawlak, "Rough Sets", International Journal of Computer & Information Sciences, vol. 11, no. 5, **(1982)**, pp. 341-356.
[3] Z. Pawlak, "Rough Sets: A Theoretical Aspect of Reasoning About Data", Kluwer Academic Publisher, **(1991)**.
[4] Z. Pawlak and A. Skowron, "Rudiments of Rough Sets", Information Sciences, vol. 177, no. 1, **(2007)**, pp. 3-27.
[5] A. Ganivada, S. S. Ray and S. K. Pal, "Fuzzy Rough Sets, and a Granular Neural Network for Unsupervised Feature Selection", Neural Networks, vol. 48, **(2013)**, pp. 91-108.
[6] Q. H. Hu, S. A. An and D. R. Yu, "Soft Fuzzy Rough Sets for Robust Feature Evaluation and Selection", Information Sciences, vol. 180, no. 22, **(2010)**, pp. 4384-4400.
[7] W. H. Shu and H. Shen, "Incremental Feature Selection Based on Rough Set in Dynamic Incomplete Data", Pattern Recognition, vol. 47, no. 12, **(2014)**, pp. 3890-3906.
[8] A. P. Zeng, T. R. Li, D. Liu, J. B. Zhang and H. M. Chen, "A Fuzzy Rough Set Approach for Incremental Feature Selection on Hybrid Information Systems", Fuzzy Sets and Systems, vol. 258, **(2015)**, pp. 39-60.
[9] S. Y. Jing, "A Hybrid Genetic Algorithm for Feature Subset Selection in Rough Set Theory", Soft Computing, vol. 18, no. 7, **(2014)**, pp. 1373-1382.
[10] B. O. Alijla, L. C. Peng, A. T. Khader and M. A. Al-Betar, "Intelligent Water Drops Algorithm for Rough Set Feature Selection", Intelligent Information and Database Systems (Aciids 2013), Pt Ii, vol. 7803, **(2013)**, pp. 356-365.
[11] V. Rampure and A. Tiwari, "A Rough Set Based Feature Selection on Kdd Cup 99 Data Set", International Journal of Database Theory and Application, vol. 8, no. 1, **(2015)**, pp. 149-156.
[12] X. P. Gu, Y. Li and J. H. Jia, "Feature Selection for Transient Stability Assessment Based on Kernelized Fuzzy Rough Sets and Memetic Algorithm", International Journal of Electrical Power & Energy Systems, vol. 64, **(2015)**, pp. 664-670.
[13] D. Molodtsov, "Soft Set Theory - First Results", Computers & Mathematics with Applications, vol. 37, no. 4-5, **(1999)**, pp. 19-31.
[14] G. Muhiuddin, F. Feng and Y. B. Jun, "Subalgebras of Bck/Bci-Algebras Based on Cubic Soft Sets", The Scientific World Journal, vol. 2014, **(2014)**, pp. 9.
[15] W. Xu, Z. Xiao, X. Dang, D. L. Yang and X. L. Yang, "Financial Ratio Selection for Business Failure Prediction Using Soft Set Theory", Knowledge-Based Systems, vol. 63, **(2014)**, pp. 59-67.
[16] Z. Xiao, W. J. Chen, and L. L. Li, "A Method Based on Interval-Valued Fuzzy Soft Set for Multi-Attribute Group Decision-Making Problems under Uncertain Environment", Knowledge and Information Systems, vol. 34, no. 3, **(2013)**, pp. 653-669.
[17] Y. Zou and Z. Xiao, "Data Analysis Approaches of Soft Sets under Incomplete Information", Knowledge-Based Systems, vol. 21, no. 8, **(2008)**, pp. 941-945.
[18] T. Herawan and M. M. Deris, "A Direct Proof of Every Rough Set Is a Soft Set", in 2009 3rd Asia International Conference on Modelling and Simulation, Bandung, Bali, Indonesia, **(2009)**, pp. 119-124.
[19] H. W. Qin, X. Q. Ma, J. M. Zain, and T. Herawan, "A Novel Soft Set Approach in Selecting Clustering Attribute", Knowledge-Based Systems, vol. 36, **(2012)**, pp. 139-145.
[20] R. Mamat, T. Herawan and M. M. Denis, "Mar: Maximum Attribute Relative of Soft Set for Clustering Attribute Selection", Knowledge-Based Systems, vol. 52, **(2013)**, pp. 11-20.
[21] P. K. Maji, R. Biswas, and A. R. Roy, "Soft Set Theory", Computers & Mathematics with Applications, vol. 45, no. 4-5, **(2003)**, pp. 555-562.
[22] Uci Machine Learning Repository. Available: http://archive.ics.uc*i.e*du/ml/datasets.html

Authors

Daoli Yang, He received the B.A. degree in Electronic Business from Shandong University in 2006. Currently, he is a Ph.D. candidate in the School of Economics and Business Administration at Chongqing University. His research interests include supply chain management, information intelligent analysis and soft set theory.

Zhi Xiao, He received the B.S. degree in Department of Mathematics from Southwest Normal University in 1982. And he received his Ph.D. degree in Technology Economy and Management from Chongqing University in 2003. He is currently a Professor in the School of Economics and Business Administration at Chongqing University. His research interests include information intelligent analysis, data mining, and soft set theory.



Wei Xu, He received the B.A. degree in Logistic Management from Jiangxi University of Finance and Economic in 2008. He received his Ph.D. degree in Quantitative Economics from Chongqing University in 2014. His research interests include intelligent decision-making, information intelligent analysis and business Intelligence.

Xianning Wang, He is a Ph.D. candidate in the School of Economics and Business Administration at Chongqing University from 2012. His research interests mainly include dynamic economics modeling and regional economy development.

Yuchen Pan, She is a Ph.D. candidate in the School of Economics and Business Administration at Chongqing University. Her research interests include information intelligent analysis and prediction, and financial information application.