JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 34, 535-550 (2018)
DOI: 10.6688/JISE.201803_34(2).0014

A Game Theory Based Feature Word Selection Model for Chinese Texts

JING-TAO SUN[1] AND QIU-YU ZHANG[2]
[1]*School of Computer Science and Technology*
Xi'an University of Posts and Telecommunications
Xi'an, 710121 P.R. China
[2]*School of Computer and Communication*
Lanzhou University of Technology
Lanzhou, 730050 P.R. China
E-mail: sunjingtao@xupt.edu.cn; zhangqylz@163.com

Feature word selection plays an important role in the classification of Chinese texts, and its result has a direct influence on the precision of text classification. Existing methods are generally deficient in processing the fuzzy and uncertain information contained in natural language. To overcome such a deficiency, a novel game theory based feature word selection model for Chinese texts was proposed in this paper. This model applies game theory method to the selection of feature words, by using the combined contribution functions of the feature subsets to text classification and the fuzzy membership functions of samples defined by compatibility measurement, compatibility feature payoff functions are constructed in order to select the optimal feature subset with Nash equilibrium. Through comparative experiments on datasets from the CDSCE corpus, it is validated that the proposed model is able to perform effective spam email feature word selection, and its generalization performance is better than those of other commonly used feature word selection methods.

Keywords: game theory, Chinese feature word selection, compatibility measurement, fuzzy membership function, combined contribution function

1. INTRODUCTION

As we step into the era of big data, text content analysis has been becoming an effective means to interpret big data and discover underlying valuable information. Text classification, as a key technology to the content mining of big data, is applied to various areas, such as internet public opinion monitoring and early warning, hazardous information filtering, and sentiment analysis [1]. A crucial step in text classification is feature selection, which directly affects the model construction and the efficiency and accuracy of classification.

Currently, vector space model [2] (VSM)-based text annotation and machine learning methods [3] are commonly used for text classification. There are two types of problems in such practice. First, since the target is text content information, as many feature words as possible contained in each sample are selected to be the sampling feature points, in order to represent the content of each sample completely and avoid omitting important information. However, feature points selected in this way are often strongly correlated, increasing the redundancy of feature points and the dimensionality of the sample vector

Received September 18, 2016; revised May 16, 2017; accepted June 8, 2017.
Communicated by Shyi-Ming Chen.

space. Therefore, feature selection is needed in processing text dataset. At present, document frequency (DF) method [4, 5] is one of the relatively more influential feature selection methods. DF method counts the occurrences of feature words in the texts so as to determine the importance of features. As a feature's frequency of occurrences is not necessarily positively related to the class information it contains, the performance of DF method is not satisfying. Information gain (IG) method [6, 7] is another popular feature selection method; it ranks text features based on IG. This method, however, has significantly reduced performance on the corpora with uneven distributions of text classes. The mutual information (MI) method [8, 9] reflects the degree of correlation between each feature word and the classes in its results. However, the method does not consider the frequency of each feature word's occurrences in the text dataset and this may cause that the words with different frequencies contain similar mutual information. The χ^2 statistic (CHI) feature selection method [10] is able to measure the correlations between features and classes, and the importance of each feature is decided by its χ^2 value. This method is based on a distribution assumption between feature words and text classes; if the assumption does not stand, the low-frequency feature words are more likely to be chosen. So the existing feature selection methods are either based on term frequency or DF. This suggests that these methods are using limited information for calculating utility of a feature. In [11], this paper introduced a Game-theoretic rough set based method for combining Term frequency and DF in a novel way. The method incorporates the measures as players in a game where each player employs a three-way decision in selecting features. In [12], this paper improved global feature selection scheme (IGFSS) where the last step in a common feature selection scheme is modified in order to obtain a more representative feature set is proposed. Thus, a local feature selection method is used in IGFSS to label features according to their discriminative power on classes and these labels are used while producing the feature sets. Besides, better achievements have been gained by Granular Computing techniques to solve the feature word selection problem [13-23]. The other problem of VSM-based classification methods is that the inherent fuzziness and uncertainty of samples are often conventionally ignored, so are the resultant impacts on the selection of feature words. As a result, some noise and redundant feature points may be introduced for the second time, so as to reduce the classification performance in solving practical problems. Some studies have proposed algorithms for the feature selection of such datasets with uncertainty. In [24], fuzzy mutual information is introduced in conventional feature selection to solve the new problems caused by data uncertainty in using existing methods. However, the application of MI method often has such a problem that the feature terms of different frequencies contain the same quantity of MI, and the problem persists. In [25], on the basis of improving MI method, the concept of information entropy is introduced to propose a feature evaluation function TFMIIE. The improvement of MI helps avoid bias to low-frequency rare words, and the introduction of entropy helps eliminate the feature words which belong to uncertain classes. In [26], based on the Hilbert-Schmidt independence criterion, a feature selection algorithm FSUNT is proposed. The algorithm investigates the Hilbert-Schmidt correlation between features and uncertain class labeling to rank features, in order to select the final feature subset. In [27], this paper depicts the uncertainty of feature information by using mutual information entropy based on the game theory. The paper has provided a balance solution for this strategy by the payoff matrix so that the optimum feature subset can be obtained, the above-mentioned methods have been used to conduct feature selection as applied in vehicle selection, it may state that vehicle feature dimensionality can be compressed. In [28], a novel feature selection method based on coalition game theory has been proposed that enable the prediction of the severity of hemorrhage. The proposed feature selection method enhances the accuracy of prediction by optimally selecting the features compared to the state of the art.

Based on the above discussion and the current research results, it is found that, as the mathematical theory and method to determine whether there is the most reasonable action plan between competitors in the research on game behaviors, the game theory is highly consistent with the idea of picking out the optimal subset in feature selection. Therefore, we introduce a new idea in this paper to overcome the deficiencies of the conventional feature word selection methods – a game theory based feature word selection model for Chinese texts. This model introduces the samples' membership functions, which are defined by means of compatibility measurement, to the weight definition of feature points. A discrimination function of features is constructed to eliminate noisy feature points. On this basis, game theory is introduced to establish a text feature word selection model. By using the combined contribution function of features contained in each feature subset to text classification as well as the discrimination function of feature points for text classification, the payoff functions of feature points are constructed. The optimal feature subset with Nash equilibrium is chosen to reduce the amount of feature points. In addition, the selected feature points can comprehensively reflect the content of the texts and improving the efficiency of classification. The experiments with the china education and research computer network emergency response team (CCERT) Data Sets of Chinese Emails (CDSCE) demonstrate that the proposed game theory based feature word selection model significantly improves the performance of the spam email classifiers constructed using Naïve Bayes algorithm [29].

The rest of this paper is organized as follows. Section 2 introduces the key technologies and methods involved in this study. Section 3 proposes the game theory based text feature word selection model in terms of model definition, player design, and payoff function. Section 4 analyzes the model's handling process through a simple numerical example. Section 5 details a case study of the algorithm and analyzes the result. Section 6 concludes the paper and proposes future works.

2. PROBLEM STATEMENT AND PRELIMINARIES

2.1 Game Theory

Studies on game theory can be traced back to the early 19th century or even earlier. Game theory investigates the actions of decision makers, their decision-making upon direct interaction, and how equilibrium is reached during a decision-making process. A decision maker is influenced by other factors (natural or human factors), and can influence decisions on other factors (especially human factors). The logic of game theory originates from Bayesian decision theory. Generally, a game consists of four basic elements, game players, players' actions, players' strategies, and payoffs [30, 31].

The mathematical model of game theory can be expressed by a triad $G = \langle \Gamma; (S_i); (u_i) \rangle$, where $\Gamma = \{1, 2, \ldots, K\}$ is the set of all players, $i = 1, 2, \ldots, K$ denotes the K players in the game, $S_i = \{s_i\}$ is the set of all the strategies of player i ($i = 1, 2, \ldots, K$), s_i is one strategy of player i, and $u_i = u_i(s_i, s_{-i})$ is the payoff of player i.

Solving a game theory problem is the process of searching for a strategy set s^* [32]. In 1950s, John Nash proved the existence of equilibrium solutions, which transforms the solving of a game theory problem into the search for its Nash equilibrium. Therefore, if a strategy set is not a Nash equilibrium, it can not be the solution to the problem.

The definition of Nash equilibrium is as follows [33]. For a Given K-player strategic game $G = \langle \Gamma; (S_i); (u_i) \rangle$, strategy set $s^* = (s_1^*, ..., s_i^*, \ldots, s_K^*)$ is a Nash equilibrium if and only if when $\forall i \in \Gamma$ and $\forall s_i \in S_i$, or $\forall i \in \Gamma$ and $s_i^* \in \arg \underset{s_i \in S_i}{Max}\, u_i(s_i, s_{-i}^*)$.

2.2 Dynamic Clustering Algorithm based on Fuzzy Equation Matrix Algorithm

The dynamic clustering algorithm based on fuzzy equation matrix (DCAFEM) is a practical fuzzy clustering analysis algorithm proposed for the case where the number of classes is unknown and the clustering needs to be dynamic upon different requirements [34-36].

Definition 1: Feature index matrix. Assuming a finite sample set $A = \{a_1, a_2, ..., a_n\}$ to be classified, and in the sample set, each sample a_i has m feature indices, which means a_i can be represented by an m-dimensional feature index vector $a_i = \{a_{i1}, a_{i2}, ..., a_{im}\}$ ($i = 1, 2, \ldots, n$), where a_{ij} denotes the jth feature index of the ith sample, then all the feature indices of the n samples constitute a matrix $A^* = (a_{ij})_{n\times m}$, which is called the feature index matrix of A.

Definition 2: Fuzzy similarity matrix. Assuming an n-order fuzzy matrix $R = (r_{ij})_{n\times n}$ on a finite domain of discourse $A = \{a_1, a_2, \ldots, a_n\}$, it is a fuzzy similarity matrix if and only if it is 1) reflexive, *i.e.* $r_{ii} = 1$, and 2) symmetric, *i.e.* $r_{ij} = r_{ji}$.

Definition 3: Fuzzy equivalence matrix. Assuming an n-order fuzzy matrix $R' = (r'_{ij})_{n\times n}$ on a finite domain of discourse $A = \{a_1, a_2, \ldots, a_n\}$, it is a fuzzy equivalence matrix if and only if it is 1) reflexive, *i.e.* $r_{ii} = 1$, and 2) symmetric, *i.e.* $r_{ij} = r_{ji}$, and 3) transitive, *i.e.* $R' \circ R' \subseteq R'$.

The DCAFEM algorithm mainly consists of the following steps [37].

Step 1: Construct the feature index matrix for fuzzy clustering analysis;

Step 2: Data normalization. Since the m feature indices may not have the same dimensionality and order of magnitude, their values must be normalized in order to eliminate potential impacts of such differences;

Step 3: Construct the fuzzy similarity matrix. After all a_{ij} ($i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$) have been normalized, multivariate analysis is applied to determine the similarity $r_{ij} = R(a_i, a_j) \in [0,1]$ ($i, j = 1, 2, \ldots, n$) between $a_i = \{a_{i1}, a_{i2}, ..., a_{im}\}$ and $a_j = \{a_{j1}, a_{j2}, ..., a_{jm}\}$, and thereby construct the fuzzy similarity matrix $R = (r_{ij})_{n\times n}$ between samples;

Step 4: Construct the fuzzy equivalence matrix. The fuzzy similarity matrix constructed through the above steps is not necessarily transitive and needs to be transformed into a fuzzy equivalence matrix. Specifically, a transitive closure with similarity is constructed by squaring R, *i.e.* $R \circ R = R^2$, and squaring it again, *i.e.* $R^2 \circ R^2 =$

R^4, and then repeating the process to obtain R^8, R^{16}, and so on until $R^{2K} = R^K$. Then, R^K is a fuzzy equivalence matrix that can be used for fuzzy clustering analysis.

3. GAME THEORY BASED TEXT FEATURE WORD SELECTION MODEL

3.1 Model Definition

In game theory, each player's strategy is dependent on the other players' strategies. We employ this idea to select feature points from text sample sets. By doing this, we hope to obtain the optimal feature subset and improve the text classification performance. DCAFEM, expected cross entropy (ECE) [38], and characteristic feature subset (CFS) evaluation criterion [39] are used in the model, so as to make it more reliable for feature word selection and adapt to practical problems more easily. Its processing flow is shown in Fig. 1.

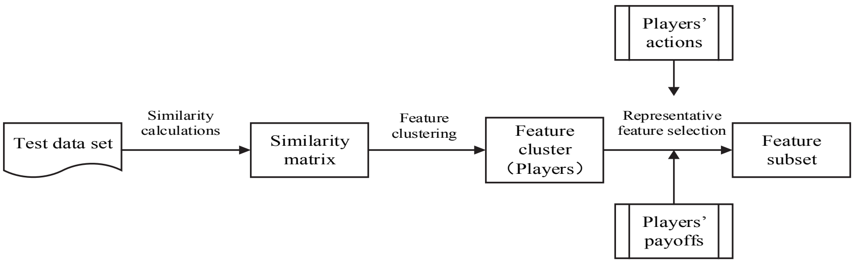

Fig. 1. Processing flow.

As shown in Fig. 1, based on the definition of game theory and related theorems, we define the model as follows.

(1) Players

It is pointed out in [40] that the feature selection in a feature space with redundancy and complementary relation can be treated as a game between different features. Based on this idea, we first cluster the feature points in the sample set by using the DCAFEM algorithm. The feature points in the same class, as in the clustering result, are treated as a whole and defined as one player of the game. Then, the K classes resulted from the clustering are defined as K players of the game.

(2) Players' actions

Based on the definition of players, each player i ($i = 1, 2, \ldots, K$) is composed of N_i feature points of the same class in the ith clustering result. Therefore, we define the feature points of the same cluster of player i as the action set $D_i = \{d_1, d_2, \ldots, d_{N_i}\}$ of the player.

(3) Players' strategies

With feature word selection being a process of an information game, the game players all know other players' information gain functions. Meanwhile, the information game is a process where decisions are made at the same time. Therefore, the feature word selection process can be considered as a complete static information game. In such a game, since there are no timing differences in decision making, all players make decisions at the same decision point, *i.e.* the time point where the game begins. Therefore, all game players face the same and only decision situation, and the strategy set and action set of each player are the same.

(4) Players' payoffs

In game theory, payoffs represent the players' gains. Generally, payoff functions are used to represent the payoffs in specific game situations. Based on the definitions of players and players' actions/strategies, for the $Q = \prod_{i=K} D_i$ action sets generated by the K players, the corresponding payoffs ($u_1, u_2, \ldots, u_Q$) are associated with the feature set selected by each player. For example, in the action set $T_j^* = (t_1^*, t_2^*, \ldots, t_K^*)$ ($j = Q$), player i takes action $t_i^*(t_i^* \in D_i)$, and the other players take actions $t_{-i}^* = (t_1^*, t_{i-1}^*, t_{i+1}^*, \ldots, t_K^*)$, then the payoff of player i is denoted by $u(t_i^*, t_{-i}^*)$. For the selection of feature words, as a player needs to consider the degree that his/her action discriminates the class when choosing an action/strategy, the correlation $M_{t_i^*}$ between feature points and classes becomes a key factor in the definition of $u(t_i^*, t_{-i}^*)$. For the definition of payoff $u(t_i^*, t_{-i}^*)$, in addition to each player's own choice, the choices of other players should be considered as well. Therefore, the conflict relation $R_{T_j^*}$ of players with the action set T_j^* should be included as well. In summary, with action set T_j^*, the formal definition of payoff function is $u(t_i^*, t_{-i}^*) = M_{t_i^*} + R_{T_j^*}$, where $M_{t_i^*}$ is the correlation, and $R_{T_j^*}$ is the conflict relation.

3.2 Design of Player

The players of the feature word selection game we study are defined in the previous section. In this section, the DCAFEM clustering result is used to design the game players further by the following process.

We assume a sample set A consisting of G pieces of texts, denoted as $A = \{a_1, a_2, \ldots, a_G\}$ where sample a_i contains n_i feature points. Sample set A contains in total $N = \sum_{i=1}^{G} n_i$ feature points. Thus, feature point matrix $C = (c_{ij})_{N\times G} = (C_1, C_2, \ldots, C_N)^T$ can be constructed for the sample set. In order to eliminate the impacts resulted from the differences in feature points' units and orders of magnitude, the values in the matrix need to be normalized. By using equations $\overline{C}_i = \frac{1}{G}\sum_{j=1}^{G} c_{ij}$ and $\sigma_i^2 = \frac{1}{G}\sum_{j=1}^{G}(c_{ij} - \overline{C}_i)^2$ ($i = 1, 2, \ldots, N, j = 1, 2, \ldots, G$), each row of data is normalized to $c'_{ij} = \frac{c_{ij} - \overline{C}_i}{\sigma_i}$ ($i = 1, 2, \ldots, N, j = 1, 2, \ldots, G$), where $\overline{C}_i$ and σ_i are the mean and standard deviation of C_i, respectively. Then, clustering is performed to the N feature points. Through multivariate analysis, the degree of similarity $r_{ij} = R(C_i, C_j) \in [0, 1]$ ($i, j = 1, 2, \ldots, N$) between feature points C_i ($i \in N$) and C_j ($j \in N$) is determined, thereby constructing a fuzzy similarity matrix $R = \begin{pmatrix} r_{11} & \cdots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \cdots & r_{NN} \end{pmatrix}$. The degree of similarity is calculated by the following equation:

$$r_{ij} = \frac{1}{G}\sum_{k=1}^{G}\exp\{-\frac{3}{4}(\frac{c_{ik}-c_{jk}}{\delta_k})^2\}, \tag{1}$$

where $\delta_k^2 = \frac{1}{N}\sum_{i=1}^{N}(c_{ik}-\bar{a}_k)^2$ and $\bar{a}_k = \frac{1}{N}\sum_{i=1}^{N}c_{ik}$.

The fuzzy similarity matrix $R = (r_{ij})_{N\times N}$ constructed through the above process is not necessarily transitive. Therefore, before clustering the feature points, the matrix needs to be transformed into a fuzzy equivalence matrix R^*. Then, dynamic clustering is performed based on the fuzzy equivalence matrix. In order to improve the algorithm's universality, the fuzzy transitive closure method is adopted for clustering. The process is detailed as follows.

(a) Calculate the transitive closure $t(R)$ of fuzzy similarity matrix R by using square method, *i.e.* $R^2 \Rightarrow R^4 \Rightarrow R^{2k} = t(R)$, where $k \leq [\log_2 N] + 1$;
(b) Choose a proper confidence level $\lambda \in [0, 1]$ and calculate the λ-cutest matrix $t(R)_\lambda$ of $t(R)$. With the value of λ varying in the range of [0, 1], the classification result varies as well.

3.3 Design of Players' Payoff Functions

When defining the model, we gave formal definition to the payoff function $u(t_j^*, t_{-i}^*)$ of player i with the action set T_j^*. In order to make the model more practical, mathematical abstraction of the formal definition is needed. For the calculation of $u(t_j^*, t_{-i}^*)$, to obtain the correlation $M_{t_i^*}$ between feature points and classes, methods such as MI, IG, and expected cross entropy are commonly used. However, the contribution of each sample's own degree of class membership to the calculation of the correlation is ignored in such methods, thus the description of the correlation is unreasonable. In this study, in addition to calculating the correlation between feature words and classes, we introduce the inherent degree of class membership. Meanwhile, for the measurement of conflict relation $R_{T_j^*}$ in the calculation of $u(t_j^*, t_{-i}^*)$, we define the correlation between action set T_j^* and the class as the conflict based on CFS evaluation criterion.

$$M_{t_i^*} = \overline{f(a_l)}p(t_i^*)\sum_{h=1}^{|c|}p(c_h \mid t_i^*)\log(\frac{p(c_h \mid t_i^*)}{p(c_h)}) \tag{2}$$

In Eq. (2), $M_{t_i^*}$ measures the correlation between feature word t_i^* and its class, $\overline{f(a_l)}$ is the average degree of class membership of samples that contain feature word t_i^*, $p(t_i^*)$ is the frequency of texts where feature word t_i^* occurs, $p(c_h)$ is the frequency of texts belonging to class c_h, $p(c_h|t_i^*)$ is the frequency of texts belonging to class c_h and containing word t_i^*, and $|c|(|c| \geq 2)$ is the total amount of classes.

$$R_{T_j^*} = \frac{K\overline{M_q}}{\sqrt{K + K(K-1)\overline{r_{qq}}}} \tag{3}$$

In Eq. (3), $R_{T_j^*}$ measures the correlation between a feature subset T_j^* containing K feature words and its class, $\overline{M_q}$ is the average correlation between feature word $q \in T_j^*$ and class c_h, and $\overline{r_{qq}}$ is the average correlation coefficient between features. In the previous section, the calculation of $\overline{r_{qq}}$ has been described.

After mathematical abstraction, payoff function $u(t_j^*, t_{-i}^*)$ is expressed as Eq. (4) below.

$$u(t_i^*, t_{-i}^*) = \overline{f(a_l)} p(t_i^*) \sum_{h=1}^{|c|} p(c_h \mid t_i^*) \log(\frac{p(c_h \mid t_i^*)}{p(c_h)}) + \frac{K\overline{M_q}}{\sqrt{K + K(K-1)\overline{r_{qq}}}} \tag{4}$$

In Eq. (4), the calculation of $\overline{f(a_l)}$ is a key to this study. This variable is proposed to reflect a potential problem in practice that sample a_l does not explicitly belong to a specific class but shows certain fuzziness. We define the membership function through compatibility measurement. For convenience of description and simplicity of symbols, text sample a_l is denoted by x when describing the calculation process. The calculation process is detailed as below.

Assuming $k \in N+$ and $k > 1$; sample x has k neighboring samples, p of which belong to the lth class, denoted as $x_{l1}^*, x_{l2}^*, \ldots, x_{lp}^*$, then the local dispersion $V_l(x)$ of sample x about class l is defined as

$$V_l(x) = \frac{1}{p} \sum_{i=1}^{p} (x_{li}^* - x)^T (x_{li}^* - x), \tag{5}$$

the compatibility $Co_l(x)$ between sample x and class l is defined as

$$Co_l(x) = \frac{\underset{1 \le i \le p}{Max}\{V_l(x_{li}^*)\}}{V_l(x)}, \tag{6}$$

thus the membership function $f(x)$ of sample x is defined as

$$f(x) = Co_l(x). \tag{7}$$

3.4 Algorithm Description

Input: $A = \{a_1, a_2, \ldots, a_G\}$//Test data set
Output: T_{best}^*//The set of selected features
Begin

a) Initialization: $T_{\text{best}}^* = \varnothing$
b) Constructing a feature point matrix $C = (c_{ij})_{N \times G}$;
c) Calculating the degree of similarity $r_{ij} = R(C_i, C_j) \in [0,1]$ $(i, j = 1, 2, \ldots, N)$ between feature points $C_i (i \in N)$ and $C_j (j \in N)$;
d) Constructing a fuzzy similarity matrix R;
e) By the DCAFEM algorithm get K clusters;
f) The feature set T_j^* selected by the K clusters;
g) By Using Eq. (4) to calculate the payoff function $u(T_j^*)$;
h) If a certain condition is satisfied, $T_{\text{best}}^* = T_j^*$, otherwise, Proceed to Step f).

End

4. A SIMPLE NUMERICAL EXAMPLE

In order to further explain the proposed model's applications to practical problems, we choose 6 email texts to form a simple test set for numerical analysis. Since the number of samples is small and each sample belongs to an explicit class, the samples' own degree of membership is 1 or 0. Table 1 details the selected test set, consisting of 2 spam emails about making money on the Internet, 2 spam emails about pornography, and 2 regular emails about conference notices.

Table 1. Original texts.

Num	Text
Mail1	与其天天耗在网上闲聊玩游戏看美女写真,不如与我一道在网上边娱乐边赚钱,比闲聊泡美女强多了!
Mail2	让你在上网娱乐的同时也有一份不错的收入.二十一世纪,网上赚钱不在是神话,边游戏边赚钱成为可能.娱乐赚钱两不误
s-mail1	爱我美女为您提供日本美女,性感美女,美女写真，淫荡视频,绝对都是让你喷火的美女.点击进入观看。
s-mail2	太漂亮了,上网观看淫荡美女写真《漂亮日本美女写真集》《性感韩国美女写真》观看的美女视频和发表观看感言。
z-mail1	由我院承办的 2016 年全国通信软件学术会议于 6 月 25 日在逸夫教学楼报告厅举办，请全体老师按时参加
z-mail2	由科技厅主办、学会承办的"智能制造与智慧院所演进路线"专题培训于 2016 年 6 月 29 日下午 2:30 在省科技资源统筹中心举行，请各位老师踊跃参加。

Seventeen keywords are extracted from the 6 email samples to construct a word-text matrix X, as shown in Table 2.

Table 2. Original word-text matrix.

Word	Mail1	Mail2	s-mail1	s-mail2	z-mail1	z-mail2
网上	2	1	0	0	0	0
闲聊	2	0	0	0	0	0
游戏	1	1	0	0	0	0
娱乐	1	2	0	0	0	0
赚钱	1	3	0	0	0	0
上网	0	1	0	1	0	0
收入	0	1	0	0	0	0
美女	2	0	5	4	0	0
观看	0	0	1	3	0	0
写真	1	0	1	3	0	0
视频	0	0	1	1	0	0
淫荡	0	0	1	1	0	0
承办	0	0	0	0	1	1
会议	0	0	0	0	1	0
参加	0	0	0	0	1	1
培训	0	0	0	0	0	1
老师	0	0	0	0	1	1

The similarity coefficient matrix R of the feature words is computed through multivariate analysis.

$$
R=\begin{bmatrix}
1.0000 & 0.8944 & 0.9487 & 0.8000 & 0.7071 & 0.3162 & 0.4472 & 0.2667 & 0 & 0.2697 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.8944 & 1.0000 & 0.7071 & 0.4472 & 0.3162 & 0 & 0 & 0.2981 & 0 & 0.3015 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.9487 & 0.7071 & 1.0000 & 0.9487 & 0.8944 & 0.5000 & 0.7071 & 0.2108 & 0 & 0.2132 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.8000 & 0.4472 & 0.9487 & 1.0000 & 0.9899 & 0.6325 & 0.8944 & 0.1333 & 0 & 0.1348 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.7071 & 0.3162 & 0.8944 & 0.9899 & 1.0000 & 0.6708 & 0.9487 & 0.0943 & 0 & 0.0953 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.3162 & 0 & 0.5000 & 0.6325 & 0.6708 & 1.0000 & 0.7071 & 0.4216 & 0.6708 & 0.6396 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0 \\
0.4472 & 0 & 0.7071 & 0.8944 & 0.9487 & 0.7071 & 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.2667 & 0.2981 & 0.2108 & 0.1333 & 0.0943 & 0.4216 & 0 & 1.0000 & 0.8014 & 0.8540 & 0.9487 & 0.9487 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.6708 & 0 & 0.8014 & 1.0000 & 0.9535 & 0.8944 & 0.8944 & 0 & 0 & 0 & 0 & 0 \\
0.2697 & 0.3015 & 0.2132 & 0.1348 & 0.0953 & 0.6396 & 0 & 0.8540 & 0.9535 & 1.0000 & 0.8528 & 0.8528 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.5000 & 0 & 0.9487 & 0.8944 & 0.8528 & 1.0000 & 1.0000 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.5000 & 0 & 0.9487 & 0.8944 & 0.8528 & 1.0000 & 1.0000 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7071 & 1.0000 & 0.7071 & 0 & 0.7071 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7071 & 0 & 0.7071 & 1.0000 & 0.7071 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000
\end{bmatrix}
$$

The transitive closure $t(R)$ is computed by using the square self-forming method.

$$
t(R)=\begin{bmatrix}
0.8944 & 0.9487 & 0.9487 & 0.9487 & 0.7071 & 0.9487 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.0000 & 0.8944 & 0.8944 & 0.8944 & 0.7071 & 0.8944 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.8944 & 1.0000 & 0.9487 & 0.9487 & 0.7071 & 0.9487 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.8944 & 0.9487 & 1.0000 & 0.9899 & 0.7071 & 0.9487 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.8944 & 0.9487 & 0.9899 & 1.0000 & 0.7071 & 0.9487 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.7071 & 0.7071 & 0.7071 & 0.7071 & 1.0000 & 0.7071 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.8944 & 0.9487 & 0.9487 & 0.9487 & 0.7071 & 1.0000 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 1.0000 & 0.8944 & 0.8944 & 0.9487 & 0.9487 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.8944 & 1.0000 & 0.9535 & 0.8944 & 0.8944 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.8944 & 0.9535 & 1.0000 & 0.8944 & 0.8944 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.9487 & 0.8944 & 0.8944 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.6708 & 0.9487 & 0.8944 & 0.8944 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.7071 & 1.0000 & 0.7071 & 0.7071 & 0.7071 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.7071 & 0.7071 & 0.7071 & 1.0000 & 0.7071 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.7071 & 1.0000 & 0.7071 & 1.0000
\end{bmatrix}
$$

With $\lambda = 0.7$, the cutest set $t(R)_\lambda$ is obtained.

$$
t(R)_\lambda=\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

Therefore, we obtain three clusters, {网上, 闲聊, 游戏, 娱乐, 赚钱, 上网, 收入}, {美女, 观看, 写真, 视频, 淫荡}, and {承办, 会议, 参加, 培训, 老师}. By using the payoff functions, the payoff matrix of each player is computed. With the strategy of repeatedly eliminating weak strategies, the payoff matrices are screened. Finally, the action set {赚钱, 美女, 参加} is obtained as the optimal feature selection strategy of the players, *i.e.* the optimal feature subset of the action/strategy set.

5. CASE STUDY

Evaluation on the application effect of the model is analyzed and researched by choosing the experimental data sets which can satisfy different situations such as different parameters, different training data scale and different training text length. Therefore, Chinese spam content classification is taken as an example in this paper. The model is implemented on Matlab R2015a software platform. The PC used to compile and run the code is an HP Pavilion 15 computer with Intel i7-6500U CPU, 8 GB memory, and Win-10 64-bit operating system. The CDSCE email corpus is used as the test dataset. Some supplementation and optimization are performed to the corpus, such as supplementing new types of emails, excluding outdated types of emails, and balancing the numbers of small-sample types of emails. The updated corpus contains 18593 spam emails and 9956 non-spam emails. For the preprocessing of texts, the Chinese lexical analysis system ICTCLAS of the Institute of Computing Technology is used. The Naïve Bayes classification algorithm, which is simple and efficient, is employed for classification.

The classification is evaluated based on the recall, precision, and F_1 value, as defined below.

1. Recall is the rate of spam emails that are detected. This indicator reflects the model's ability of detecting spam emails. A higher recall means fewer spam emails being missed.
2. Precision is the rate of correctly detected spam emails. This indicator reflects the system's ability to correctly detect spam emails. A higher precision means fewer non-spam emails being classified as spam emails.
3. F_1 value is a harmonic mean of recall and precision. It is in fact an integration of the two indicators above.

The following formulae calculate the three indicators.

Recall: $R = (N_A/N_S)*100\%$
Precision: $P = (N_A/(N_A+ N_B))*100\%$
F_1 value: $F_1 = (2RP/(R+P))*100\%$

In the above formulae, N_A is the number of spam emails being correctly detected, N_S is the actual number of spam emails, and N_B is the number of legit emails that are detected as spam emails.

Table 3. Training sets.

	Number of spam emails	Number of non-spam emails	Total
TR-CDSCE1	849	672	1521
TR-CDSCE2	1371	1365	2736
TR-CDSCE3	2080	1894	3974

Table 4. Test sets.

	Number of spam emails	Number of non-spam emails	Total
TE-CDSCE1	346	167	513
TE-CDSCE2	672	452	1124
TE-CDSCE3	944	739	1683

To make sure that the corpus objectively reflects the model's email filtering performance, a certain amount of emails are randomly picked from the CDSCE corpus to form three training sets and three test sets. Tables 3 and 4 detail the distributions of spam and non-spam emails in the training and test sets. Through 5-fold cross validations on the three training sets, the proposed feature selection model and conventional feature selection method are evaluated with Naïve Bayes classifiers.

Table 5. Experimental results of DF method, IG method, MI method, and proposed game theory based feature selection model with Naïve Bayes classifiers.

Results / Methods	Classification result on TE-CDSCE1			Classification result on TE-CDSCE2			Classification result on TE-CDSCE3		
	Recall (%)	Precision (%)	F_1 value (%)	Recall (%)	Precision (%)	F_1 value (%)	Recall (%)	Precision (%)	F_1 value (%)
Game theory based feature selection model	**<u>83.36</u>**	**<u>92.87</u>**	**<u>87.86</u>**	81.71	**<u>89.93</u>**	**<u>85.62</u>**	82.17	**<u>91.95</u>**	**<u>86.79</u>**
DF	79.24	82.43	80.80	78.82	83.41	81.05	76.97	81.74	79.28
IG	80.49	84.91	82.64	**<u>81.92</u>**	85.26	83.55	**<u>82.63</u>**	84.95	83.77
MI	70.76	72.25	71.5	68.63	71.97	70.26	68.11	72.4	70.19

DF method, IG method, MI method, and the proposed game theory based feature selection model are evaluated on the test sets with Naïve Bayes classifiers. The results are presented in Table 5, in which the bold and underlined numbers are the largest under each indicator. As can be seen, the method used for feature selection has significant influence on the classification performance. MI method performs worse than the other methods in terms of all three indicators, IG method is slightly better than DF method, and the proposed model performs only slightly worse than other methods in terms of recall. Overall, the proposed game theory based feature selection model has better average performance than those of the other three methods.

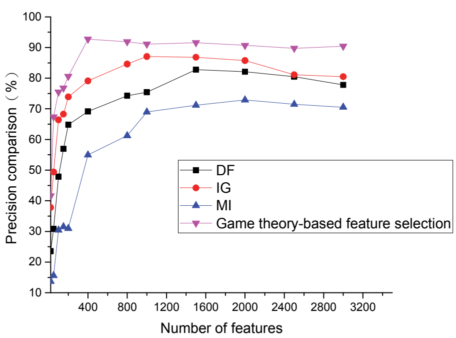

Fig. 2. Precision comparison.

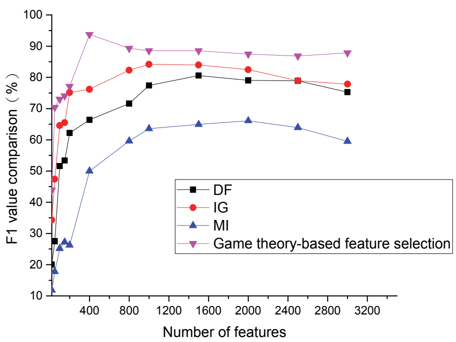

Fig. 3. F_1 value comparison.

As seen in Figs. 2 and 3, it is apparent that the evaluation indicator values increase with the number of features. However, every method shows an inflection point when the number of features increased to a certain value, indicating that the values of evaluation indicators would eventually stabilize or decrease. Therefore, an overly large amount of features is not desirable. When we use the feature selection model based on game theory with 400 features, the accuracy of Naïve Bayes classifier reaches a maximum of 93.72, and then the value tends to be stable. The accuracy and F_1 values of DF and IG method are relatively close, while MI method is the worst-performing among these 4 methods. As can be seen from Fig. 2, when the number of features is small (less than 200), each method has low classification accuracy, and there is little difference in the classification accuracy between the feature selection model based on game theory and IG method then. But with the addition of more characteristic words, the classification effect of each method is improved a lot, and the feature selection model based on game theory gets a faster promotion compared to other methods. As Fig. 3, when the feature number of DG and MI method is greater than 2400, the value of F_1 drops significantly. According to the experimental results of Figs. 2 and 3 and Table 5, we can see that the feature subset selected by the feature selection model based on game theory proposed in this paper can obtain better classification results faster, owning better generalization performance than that of DF, IG and MI method.

6. CONCLUSIONS

In this paper, we propose a game-theory based feature word selection model for Chinese texts. This model considers not only the influence of the correlation among feature words on classification, but also the influence of samples' own fuzzy membership on the selection of feature subsets. By calculating the feature payoff, the optimal feature subset with Nash equilibrium is chosen, so as to overcome the defect that the text discrimination power of only single features is analyzed in feature selection, as well as the deficiencies of current feature word selection methods in processing the fuzzy and uncertain information contained in natural language texts. Experiments were performed to validate the proposed feature selection model by applying it to the classification of spam emails. With the maximum amount of information of the training samples preserved, the optimal feature subset is selected to reduce the amount of features and improve the email filtering efficiency. As big data technology and application become more and more common, feature selection has become a crucial step for processing the big data in Internet public opinion analysis and social network data analysis. The proposed feature word selection model will play an important role in those rising application fields of big data. However, further studies are needed to improve the model, such as its convergence during feature selection and its time complexity.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61363078), the Science and Technology Project in Shaanxi Province of China (No. 2016KTZDGY04-01, No. 2016GY-092). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

1. A. K. Uysal, "An improved global feature selection scheme for text classification," *Expert Systems with Applications*, Vol. 43, 2016, pp. 82-92.
2. L. Ma and H. F. Liu, "Study on the extraction of characteristics of Chinese text based on the LDA model," *Journal of Xi'an University of Posts and Telecommunications*, Vol. 20, 2015, pp. 79-81.
3. X. Lu, Y. Li, Q. Wang, *et al.*, "An approach to sentiment analysis of short Chinese texts based on SVMs," *IEEE Computer Society*, 2015, pp. 9115-9120.
4. Y. Chen, Q. Zhou, W. Luo, *et al.*, "Classification of Chinese texts based on recognition of semantic topics," *Cognitive Computation*, Vol. 8, 2016, pp. 114-124.
5. B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," *Procedia Engineering*, Vol. 69, 2014, pp. 1356-1364.
6. B. Azhagusundari and A. Thanamani, "Feature selection based on information gain," *International Journal of Innovative Technology and Exploring Engineering*, Vol. 2, 2013, pp. 18-21.
7. H. Dan, G. U. Daibeng, and P. Miao, "Improved information gain text feature selection algorithm based on word frequency information," *Journal of Computer Applications*. Vol. 34, 2014, pp. 3279-3282.
8. M. A. Sulaiman and J. Labadin, "Feature selection based on mutual information," in *Proceedings of IEEE International Conference on IT in Asia*, 2015, pp. 1-6.
9. H. Liu, Z. Yao, and S. U. Zhan, "Optimization mutual information text feature selection method based on word frequency," *Computer Engineering*, Vol. 40, 2014, pp. 179-182.
10. Y. F. Qiu, W. Wang, and D. Y. Liu, "Research on an improved CHI feature selection method," *Applied Mechanics and Materials*, Vol. 241-244, 2012, pp. 2841-2844.
11. N. Azam and J. T. Yao, "Game-theoretic rough sets for feature selection," *Intelligent Systems Reference Library*, Vol. 43, 2013, pp. 61-78.
12. A. K. Uysal, "An improved global feature selection scheme for text classification," *Expert Systems with Applications*, Vol. 43, 2016, pp. 82-92.
13. Y. Yao, "A triarchic theory of granular computing," *Granular Computing*, Vol. 1, 2016, pp. 145-157.
14. D. Dubois and H. Prade, "Bridging gaps between several forms of granular computing," *Granular Computing*, Vol. 1, 2016, pp. 115-126.
15. G. Peters and R. Weber, "DCC: a framework for dynamic granular clustering," *Granular Computing*, Vol. 1, 2016, pp. 1-11.
16. L. Livi and A. Sadeghian, "Granular computing, computational intelligence, and the analysis of non-geometric input spaces," *Granular Computing*, Vol. 1, 2016, pp. 13-20.
17. P. Lingras, F. Haider, and M. Triff, "Granular meta-clustering based on hierarchical, network, and temporal connections," *Granular Computing*, Vol. 1, 2016, pp. 1-22.
18. A. Skowron, A. Jankowski, and S. Dutta, "Interactive granular computing," *Granular Computing*, Vol. 1, 2016, pp. 95-113.
19. M. Antonelli, P. Ducange, B. Lazzerini, *et al.*, "Multi-objective evolutionary design of granular rule-based classifiers," *Granular Computing*, Vol. 1, 2016, pp. 37-58.

20. D. Ciucci, "Orthopairs and granular computing," *Granular Computing*, Vol. 1, 2016, pp. 159-170.
21. H. Liu, A. Gegov, and M. Cocea, "Rule-based systems: a granular computing perspective," *Granular Computing*, Vol. 1, 2016, pp. 259-274.
22. F. Min and J. Xu, "Semi-greedy heuristics for feature selection with test cost constraints," *Granular Computing*, Vol. 1, 2016, pp. 199-211.
23. S. S. S. Ahmad and W. Pedrycz, "The development of granular rule-based systems: a study in structural model compression", *Granular Computing*, Vol. 2, 2017, pp. 1-12.
24. J. Grande, M. D. R. Suárez, and J. R. Villar, "A feature selection method using a fuzzy mutual information measure," *Innovations in Hybrid Intelligent Systems*, Vol. 2, 2007, pp. 56-63.
25. W. Q. Chen and X. Tang, "A text feature selection method using the improved mutual information and information entropy," *Journal of Nanjing University of Posts and Telecommunications* (*Natural Science*), Vol. 33, 2013, pp. 63-68.
26. B. Wang, Y. Jia, S. Q. Yang, *et al.*, "Feature selection algorithm for uncertain text classification," *Journal on Communications*, Vol. 30, 2009, pp. 32-38.
27. M. Fu, Y. Zhang, and F. Lian, "The research of feature selection model based on the game theory," in *Proceedings of IEEE International Conference on Wireless Communications*, *Networking and Mobile Computing*, 2009, pp. 3624-3627.
28. A. Razi, F. Afghah, A. Belle, *et al.*, "Blood loss severity prediction using game theoretic based feature selection," in *Proceedings of IEEE International Conference on Biomedical and Health Informatics*, 2014, pp. 776-780.
29. S. D. Sarkar, S. Goswami, A. Agarwal, *et al.*, "A novel feature selection technique for text classification using naïve bayes," *International Scholarly Research Notices*, Vol. 2014, 2014, pp. 717092.
30. L. Zhou, K. Lü, P. Yang, *et al.*, "An approach for overlapping and hierarchical community detection in social networks based on coalition formation game theory," *Expert Systems with Applications*, Vol. 42, 2015, pp. 9634-9646.
31. S. Mohebbi and X. Li, "Coalitional game theory approach to modeling suppliers' collaboration in supply networks," *International Journal of Production Economics*, Vol. 169, 2015, pp. 333-342.
32. Z. Bu, H. Li, J. Cao, *et al.*, "Game theory based emotional evolution analysis for Chinese online reviews," *Knowledge-Based Systems*, Vol. 103, 2016, pp. 60-72.
33. D. Fudenberg and J. Tirole, *Game Theory*, China Renmin University Press, Beijing, 2010.
34. W. J. Zhao and J. L. Liu, "Rough set methods based on fuzzy equivalence relationship," *Journal of Tonghua Teachers College*, Vol. 30, 2009, pp. 35-38.
35. Y. Zhan, Q. W. Shen, and L. Zhang, "The application of fuzzy equivalence relation based on the quotient space in cluster analysis," *Journal of Anhui Institute of Architecture and Industry*, Vol. 21, 2013, pp. 83-86.
36. X. Q. Geng, H. Y. Yang, and F. M. Tao, "DGFCM: A new dynamic clustering algorithm," *Applied Mechanics & Materials*, Vol. 556-562, 2014, pp. 3945-3948.
37. J. T. Sun, Q. Y. Zhang, Z. T. Yuan, *et al.*, "Application of game theory for email feature selection," *Journal of University of Electronic Science and Technology of China*, Vol. 40, 2011, pp. 95-99.

38. L. L. Shan, B. Q. Liu, and C. J. Sun, "Comparison and improvement of feature selection method for text categorization," *Journal of Harbin Institute of Technology*, Vol. 43, 2011, pp. 319-324.
39. M. A. Hall, "Correlation-based feature selection for machine learning," Department of Computing and Mathematical Sciences, The University of Waikato, New Zealand, 1999.
40. G. Wang, "Problems on the game fusion under the conflict environment of multiple source information," Department of Computer Science and Engineering, Northeastern University, 2004.

Jing-Tao Sun (孙晶涛) received her Ph.D. degree in Control Theory and Control Engineering from Lanzhou University of Technology in 2010. He is currently a Senior Engineer at Xi'an University of Posts and Telecommunications. His research interests include algorithm design and analysis, data mining, and Chinese text classification.

Qiu-Yu Zhang (张秋余) Researcher/Ph.D. supervisor, graduated from Gansu University of Technology in 1986, and then worked at School of Computer and Communication in Lanzhou University of Technology. He is Vice Dean of Gansu Manufacturing Information Engineering Research Center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, image understanding and recognition, multimedia communication technology.