# Research and Application of Data Mining Feature Selection Based on Relief Algorithm

Lun Gao[a], Taifu Li[a], Lizhong Yao[b], Feng Wen[b]
[a] Scientific Research Office, Chongqing University of Science and Technology, Chongqing, 40133,China
Email: cqgaolun@163.com; litaifuemail@qq.com
[b] School of Electric and Information Engineering, Chongqing University of Science and Technology, Chongqing, 40133,China
Email: {yaolizhong225, 2006631428}@163.com

***Abstract*—To choose the best features in data mining issues, the Relief Feature Selection Algorithm is proposed to implement the feature selection in this paper. Firstly, the data of Ionosphere from the UCI (University of California - Irvine) database is used to do a simulation experiment; secondly, the proposed method is employed to do feature selection for voice signal. In this case study, the study starts from the 24-dimensional parameters of MFCC (Mel Frequency Cepstrum Coefficient), the most important parameters of MFCC can be found in the voice signal; then, the 24-dimensional parameters of MFCC can be combined and optimized in the case of recognition rate not much changed. The experimental results show that the method extracts out the best features. Therefore, the research provides a new direction to feature extraction for speech recognition process.**

***Index Terms*—relief algorithm, feature selection, data mining, speech recognition**

## I. Introduction

Feature selection is not only the classic problems in the field of statistics, but also the important problems in the field of data mining [1], for instance, text categorization [2], bioinformatics [3], computer vision [4], information retrieval [5] and time series prediction [6], etc. Before predicting the unknown samples using training samples based on a data mining algorithm, some information must be obtained that which features should be adopted and which features should be ignored.

In the 1960s, feature selection is mainly studied from the viewpoint of statistics and information processing, and usually involved problems don't have much feature numbers. Since the 90s, data experts pay an unprecedented attention to the study on feature selection[7] [8] [9]. The main reasons are the following two aspects: 1) the performance from many of the data mining algorithm is adversely affected by irrelevant and redundant features [10]. Research results show the number of training samples based on data mining algorithms will increases exponentially with the growth of unrelated characteristics. 2) Appearing large-scale data processing problems [11]. Large-scale data includes two aspects: a) Large number of samples; b) Feature dimension of samples is high. However, it is an empirical "axioms" of data mining field that the dimension of the feature space should not be too high. Therefore, it puts forward severe challenges to the existing feature selection algorithm how to reduce the dimensions for high dimensional data.

Usually the feature selection method is divided into two categories from the perspective of search strategy and evaluation guidelines [12]. 1) According to search strategy to divide the feature selection algorithm. Feature selection search strategies can be divided into global optimal strategy [13], heuristic search strategy [14] and random search strategy [15]. a) Global optimal strategy. Global optimal strategies have exhaustive method [16], branch and bound method [17]. Banch and bound method saves computing time compared with the exhaustion method. However, time order of magnitude is not enough advantage than the exhaustion method, and it is often difficult to design an evaluation function which can meet the monotony in practical application. b) Heuristic search strategy. It includes optimal strategy alone, sequence forward selection method(SFS)[18], generalized sequence forward selection method(GSFS)[19], sequence backward selection method (SBS) [20], etc. These methods are difficult to select optimal feature for nonlinear data, dut to the nonlinear data shows distorted image in the low dimensional space. c) Random search strategy. Ggenetic algorithm is typical methods of andom search strategy [21]. 2) According to the evaluation standard to divide the feature selection algorithm. It can be divided into Embedded [22], Filter [8] and Wrapper [23].

Among them, the evaluation standard of Filter feature selection is divided into four kinds, as distance measurement, information measurement and correlation measurement, as well as consistency measure. Scholars believe that evaluation standard of Filter type feature selection can be obtained from data set itself inherent nature, relevant major features or sub assemblies can get high accuracy on the classifier, and Filter feature selection has better versatility. Therefore, the best Relief

Manuscript received June 20, 2013; revised July 1, 2013; accepted July 10, 2013.

The work is supported by National Natural Science Foundation of China (No.51075418), Application Development Major Projects of Chongqing (No.cstc2013yykfC0034), Natural Science Foundation Key Project (No.CQ CSTC2013jjB40007, Chongqing Educational Committee Science and Technology Research Project (No.KJ111417).


doi:10.4304/jsw.9.2.515-522

algorithm [24] [25] in Filter method is presented in this paper. Series of simulations are made using the Relief algorithm; then, the Relief algorithm is used to research speech recognition and evaluation index. Starting from 24-dimensional MFCC parameters, the study finds out the most important MFCC parameters in speech signal. Meanwhile, the 24-dimensional parameters of MFCC can be combined and optimized in the case of recognition rate not much changed.s.

## II. Study on Relief Feature Selection Algorithm

### A. Preliminaries

Relief algorithm was first proposed by Kira [24], which was applied to classification of two kinds of data. Relief is a kind of feature weighting algorithm, which gives different weights according to the relevance of features and categories. Also, the relevance of features and categories in relief algorithms is based on the ability for features to distinguish between close-sample.

Algorithm selects a sample $R$ randomly from the training sets $D$, and then search for the nearest neighbor $H$ from the similar samples with $R$, called $\text{Near Hit}$; search for the nearest neighbor M from different types of samples with $R$, called $\text{Near Miss}$. Afterwards, update the weight of each feature following the rules: the feature is beneficial to distinguish nearest neighbor as similar or not, when the distance between $R$ and $\text{Near Hit}$ in a feature is less than it between $R$ and $\text{Near Miss}$, and then increase the weights of features. Otherwise, decrease the weight of the feature. The above process is repeated m times to obtain the average weights of each feature.The greater weight of a feature, the stronger classification ability of it. The running time of the relief algorithm will linearly slow down, with the increase of sampling frequency N of the sample and the number N of original features, Therefore, the efficiency is very high.

The concept of hypothesis interva is used to the process of feature weight calculation, a measure of trust in the classifier decision, which refers to maximum distance of decision surface's moving, in the case of the same sample classification.

It is expressed as

$$\theta=\frac{1}{2}\left(\|x-M(x)\|-\|x-H(x)\|\right) \qquad (1)$$

Where $H(x)$ and $M(x)$ are similar and non-similar to nearest neighbor of $x$, respectively. Supposed that interval can evaluate the features classification ability of each dimension. We can approximately evaluate the features about classification value, by calculating the interval size of the training sample.

### B. Algorithm:Relief algorithm

Specific steps of relief algorithm are as follows:

Input: feature vector set of training examples and class label; /* $N$ -feature dimensions, $m$ - iterations */

Output: weight vector $W$ corresponding to the various components of eigenvector

1) $W[A]:=0.0$; (2)

2) $for\ i:=1\ to\ m$ (3)

3) select practical examples $R$ randomly;

4) Find out a similar nearest neighbor example $H$ and a non-similar nearest neighbor example $M$;

5) $for\ A:=1\ to\ all\ attributes$ (4)

$$W[A]:=W[A]-diff(A,R,H)/m+diff(A,R,M)/m \qquad (5)$$

6) end; (6)

7) end; (7)

Where: function $diff(Feature, Ins\tan ce_1, Ins\tan ce_2)$ is used to calculate the feature difference of the two different samples, which is defined as

For discrete features:

$$diff(F,I_1,I_2)=\begin{cases}0; & value(F,I_1)=value(F,I_2)\\ 1; & others\end{cases} \qquad (8)$$

For continuous features:

$$diff(F,I_1,I_2)=\frac{|value(F,I_1)-value(F,I_2)|}{\max(F)-\min(F)} \qquad (9)$$

Where $I_1$ 、 $I_2$ are two samples, $value(F,I_1)$ refers to the $F$ eigenvalue of samples $I_1$.

After solving the relevance weights $W$ of various features and categories, we can sort to them. Further, characters whose relevance is greater than a threshold value constitute the final feature subset, eliminating invalid feature.

## III.Simulation Case Based on Relief Feature Selection Algorithm

### A. Experimental Data Sets

Ionosphere data of UCI database is used in the experiment, a total of 351 instances of samples, each sample with 32 eigenvalues and two categories including good (substitute 1) and bad (substitute 0). The data containing instances 1 to 30 is shown in the Table I. X1 ~ X32 respectively represent different eigenvalues and CLASS stands for their class.

### B. Experimental Method and Result Analysis

The experiment uses MATLAB to simulate and program Relief algorithm, and the data from the Ionosphere is imported to calculate feature weights. Sort feature weights, draw graph of weights and classification. The experimental results are shown in the Table II.

In the above data, weight value of characteristic X14, X18, X19, X20, X22, X27, X28, X32 are all above 0.2, bigger than weight value of X2, X3, X4, X5, X15, whose value are all below 0.1339. Therefore, X2, X3, X4, X5, X15 these five features are rejected from the feature subset, with the rest of the features to represent the classification feature subset.

In the paper, Relief algorithm is described by means of the classification accuracy and classification time, based on the SVM classifier. To make results more reliable,

TABLE I.
PART OF THE DATA IN IONOPSHERE

| X1 | X2 | X3 | X4 | X5 | ··· | X32 | CLASS |
|---|---|---|---|---|---|---|---|
| 1 | 0.9954 | -0.0589 | 0.8524 | 0.0231 | ··· | 0.1864 | 1 |
| 1 | 1 | -0.1883 | 0.9304 | -0.3616 | ··· | -0.1374 | 0 |
| 1 | 1 | -0.0336 | 1.0000 | 0.0049 | ··· | -0.2418 | 1 |
| 1 | 1 | -0.4516 | 1.0000 | 1.0000 | ··· | -0.3238 | 0 |
| 1 | 1 | -0.0240 | 0.9414 | 0.0653 | ··· | 0.0461 | 1 |
| 1 | 0.0234 | -0.0059 | -0.0992 | -0.1195 | ··· | -0.0004 | 0 |
| 1 | 0.9759 | -0.1060 | 0.9460 | -0.2080 | ··· | -0.1383 | 1 |
| 0 | 0 | 0 | 0 | 0 | ··· | 0 | 0 |
| 1 | 0.9636 | -0.0720 | 1.0000 | -0.1433 | ··· | 0.3890 | 1 |
| 1 | -0.0186 | -0.0846 | 0 | 0 | ··· | -0.0821 | 0 |
| 1 | 1 | 0.0665 | 1.0000 | -0.1839 | ··· | 0.2299 | 1 |
| 1 | 1 | -0.5421 | 1.0000 | -1.0000 | ··· | 1.0000 | 0 |
| 1 | 1 | -0.1632 | 1.0000 | -0.1017 | ··· | 0.6473 | 1 |
| 1 | 1 | -0.8670 | 1.0000 | 0.2228 | ··· | 1.0000 | 0 |
| 1 | 1 | 0.0738 | 1.0000 | 0.0342 | ··· | 1.0000 | 1 |
| 1 | 0.0334 | -0.0159 | -0.0899 | -0.219 | ··· | -0.0006 | 0 |
| 1 | 0.9759 | -0.1060 | 0.9460 | -0.2080 | ··· | -0.1383 | 1 |
| 0 | 0 | 0 | 0 | 0 | ··· | 0 | 0 |
| 1 | 0.9636 | -0.0720 | 1.0000 | -0.1433 | ··· | 0.3890 | 1 |
| 1 | -0.0186 | -0.0846 | 0 | 0 | ··· | -0.0821 | 0 |
| 1 | 1 | 0.0765 | 1.0000 | -0.1939 | ··· | 0.2389 | 1 |
| 1 | 1 | -0.6421 | 1.0000 | -1.0000 | ··· | 1.0000 | 0 |
| 1 | 1 | -0.1732 | 1.0000 | -0.1317 | ··· | 0.6573 | 1 |
| 1 | 1 | -0.8770 | 1.0000 | 0.2248 | ··· | 1.0000 | 0 |
| 1 | 1 | 0.0738 | 1.0000 | 0.0342 | ··· | 1.0000 | 1 |
| 1 | 0.9835 | -0.0443 | 0.8252 | 0.0245 | ··· | 0.1933 | 1 |
| 1 | 1 | -0.1883 | 0.9304 | -0.3616 | ··· | -0.1374 | 0 |
| 1 | 1 | -0.0536 | 1.0000 | 0.0049 | ··· | -0.2318 | 1 |
| 1 | 1 | -0.4716 | 1.0000 | 1.0000 | ··· | -0.3238 | 0 |
| 1 | 1 | -0.0234 | 0.9514 | 0.0753 | ··· | 0.0463 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

TABLE II.
DATA CHARACTERISTICS OF THE WEIGHT

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
|---|---|---|---|---|---|---|---|
| 0.1709 | 0.1307 | 0.1171 | 0.1312 | 0.1141 | 0.1446 | 0.1743 | 0.1512 |
| X9 | X10 | X11 | X12 | X13 | X14 | X15 | X16 |
| 0.1669 | 0.1720 | 0.1712 | 0.1972 | 0.1465 | 0.2241 | 0.1338 | 0.1867 |
| X17 | X18 | X19 | X20 | X21 | X22 | X23 | X24 |
| 0.1917 | 0.2379 | 0.2077 | 0.2088 | 0.1797 | 0.2062 | 0.1859 | 0.1990 |
| X25 | X26 | X27 | X28 | X29 | X30 | X31 | X32 |
| 0.1839 | 0.1819 | 0.2169 | 0.2044 | 0.1821 | 0.1846 | 0.1789 | 0.2072 |

TABLE III.
RELIEF IN THE CALSSIFICATION PERFORMANCE OF SVM TRAINING

| Data set | The accuracy of classification (%) | The time of classification(s) | The data dimension |
|---|---|---|---|
| Ionosphere | 91.3907 | 25.393502 | 32 |
| Ionosphere | 89.1412 | 20.27722 | 27 |

each experiment training set and test set are randomly generated, randomly selecting 50% of data set as the training set, the remaining 50% as test set. The simulation results show classification accuracy before feature selection is 91.3907%, and classification time 25.393502s, classification accuracy after feature selection is 89.1412%, and classification time is 20.27722s.

The results from the Table III show that, in the case of without feature selection the accuracy of classification is relatively higher but the classification time is longer .However, in the case of using Relief algorithm to calculate the feature weights and reduce feature subset, the accuracy of classification does not significantly reduce but the classification time is decreased by 5s. The classification results prove the accuracy of Relief algorithm, and the classification time proves the efficient performance of algorithm Relief.

## IV. Relief Method and Its Application in Feature Extraction of Speech Recognition

In the voice recognition, the original waveform of voice signal can't be directly used for recognition; certain change must be undergone to extract the characteristic parameters of reflecting the voice essential to perform the recognition. Currently, most common voice recognition mode covers Linear Predictive Cepstrum Coefficients (LPCC) and Mel Frequency Cepstrum Coefficient (MFCC) . Within recent years, the scholars at home and abroad put forward the PLP (Perceptual Linear Predictive PLP) and (zero-crossings with peak amplitudes, ZCPA) methods to extract the characteristic parameters of voice and ensure that it is provided with better recognition performance, these improve the voice recognition ratio on certain degree. However, the parameters of voice signal are numerous, it is an issue required to be solved urgently how to extract and remain the most important characteristic parameters in the process of characteristic extraction. In order to solve this problem, the Relief characteristic selection algorithm is employed for attempting. It is started from 24-dimensional MFCC parameters to perform the characteristic selection study for voice signal via Relief Algorithm and find out the most important MFCC parameters in the voice signal, under the condition that the recognition ratio doesn't change greatly, the optimized combination is performed for 24-dimensional MFCC parameters to provide a new orientation for characteristic of voice recognition.

### *A. Voice Data Collection*

Such two kinds of different music as folk song and Koto are respectively served as Category 1 and Category 2, LPCC is employed to extract 500 groups of 24-dimensional voice characteristic signal for each section of music, the signal characteristic is replaced by employing X1~X24, and the category is expressed by CLASS, the specific data collection is shown as in Table IV.

### *B. Weight Calculation on Voice Data*

In MATLAB, the data is introduced to analyze the weight calculation; the results are such as in Table V. Fig.1 shows vividly feature weighting curve.

It can be known from above-said weight that weight of

TABLE IV.
PART OF VOICE DATA

| X1 | X2 | X3 | X4 | X5 | ··· | X24 | CLASS |
|---|---|---|---|---|---|---|---|
| -14.8271 | -3.00109 | 1.520908 | 3.955348 | -1.09918 | ··· | 0.213977 | 1 |
| -16.2289 | -2.80187 | -0.41082 | 1.475467 | -1.89839 | ··· | 0.320708 | 1 |
| -15.1243 | -2.59871 | -0.35997 | 1.345838 | -0.34883 | ··· | 0.596002 | 1 |
| -7.00807 | -3.04678 | -4.24805 | -2.73724 | -3.68622 | ··· | 0.193837 | 2 |
| -6.47052 | -2.93598 | -3.13517 | -3.86577 | -5.31144 | ··· | 0.114851 | 2 |
| -8.13019 | -3.33951 | -6.67022 | -4.29618 | -2.09678 | ··· | -0.12204 | 2 |
| -7.00807 | -3.04678 | -4.24805 | -2.73724 | -3.68622 | ··· | 0.193837 | 2 |
| -7.12468 | -3.29122 | -1.98949 | -2.29103 | -1.61213 | ··· | 0.596002 | 1 |
| -7.49117 | -3.87656 | -6.28446 | -4.83163 | 0.013242 | ··· | 0.491487 | 1 |
| -7.00815 | -3.24678 | 4.32483 | -2.72324 | -3.6845 | ··· | 0.284901 | 1 |
| -6.42254 | -2.93448 | -3.15617 | -3.82377 | -5.33244 | ··· | -0.19032 | 2 |
| 8.20231 | -3.43395 | -6.57022 | -5.29618 | -3.22678 | ··· | -0.23404 | 2 |
| -7.12468 | -3.29122 | -1.98949 | -2.29103 | -1.61213 | ··· | 0.193837 | 2 |
| -7.49117 | -3.87656 | -6.28446 | -4.83163 | 0.013242 | ··· | -0.19032 | 2 |
| -7.00567 | -3.12678 | -4334805 | -2.89724 | -3.66722 | ··· | -0.31404 | 2 |
| -6.47052 | -2.93598 | -3.13517 | -3.86577 | -5.31144 | ··· | 0.193837 | 2 |
| -8.15419 | -3.36751 | -6.67552 | -4.23218 | -2.67678 | ··· | 0.234851 | 2 |
| -7.15607 | 4.3458 | -4.655805 | -2.76724 | -3.23622 | ··· | 0.156837 | 2 |
| -7.12458 | -3.29342 | -1.98129 | -2.45103 | -1.61563 | ··· | 0.597702 | 1 |
| -7.49145 | -3.876656 | -6.22846 | -4.82163 | 0.019242 | ··· | 0.495687 | 1 |
| -7.032080 | -3.345678 | -4.66 805 | -2.43 724 | -3.76 622 | ··· | 0.283701 | 1 |
| -6.47052 | -2.93598 | -3.13517 | -3.86577 | -5.31144 | ··· | -0.19032 | 2 |
| -7.49117 | -3.87656 | -6.28446 | -4.83163 | 0.013242 | ··· | -0.16404 | 2 |
| -7.210347 | -3.3348 | -4.66505 | 2.87724 | -3.91622 | ··· | 0.235837 | 2 |
| -7.47053 | -3.07935 | -4.23517 | -3.97877 | -5.34144 | ··· | 0.7614851 | 2 |
| -7.93019 | -3.6733951 | -6.648722 | -4.27818 | -2.01278 | ··· | 0.420708 | 1 |
| -16.2289 | -2.80187 | -0.41082 | 1.475467 | -1.89839 | ··· | 0.596002 | 1 |
| -15.1243 | -2.59871 | -0.35997 | 1.345838 | -0.34883 | ··· | 0.193837 | 2 |
| -7.00807 | -3.04678 | -4.24805 | -2.73724 | -3.68622 | ··· | 0.213977 | 1 |
| -14.8271 | -3.00109 | 1.520908 | 3.955348 | -1.09918 | ··· | 0.320708 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

TABLE V.
CHARACTERISTIC WEIGHT

| X1 | X2 | X3 | X4 | X5 | X6 |
|---|---|---|---|---|---|
| 0.3339 | 0.1346 | 0.1069 | 0.1553 | 0.1090 | 0.1332 |
| X7 | X8 | X19 | X10 | X11 | X12 |
| 0.1053 | 0.1481 | 0.1198 | 0.1154 | 0.0711 | 0.0880 |
| X13 | X14 | X15 | X16 | X17 | X18 |
| 0.1241 | 0.0749 | 0.0756 | 0.0995 | 0.0970 | 0.1012 |
| X19 | X20 | X21 | X22 | X23 | X24 |
| 0.1383 | 0.0705 | 0.1036 | 0.1150 | 0.0752 | 0.0940 |

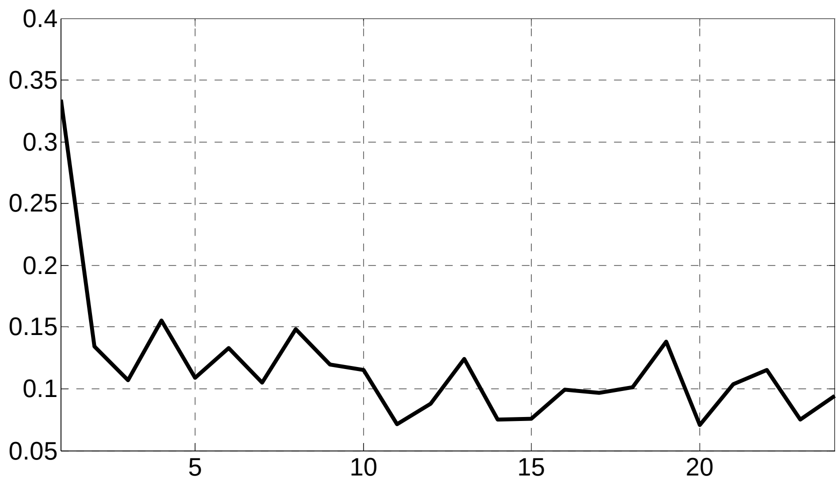


Figure 1. Feature weighting curve

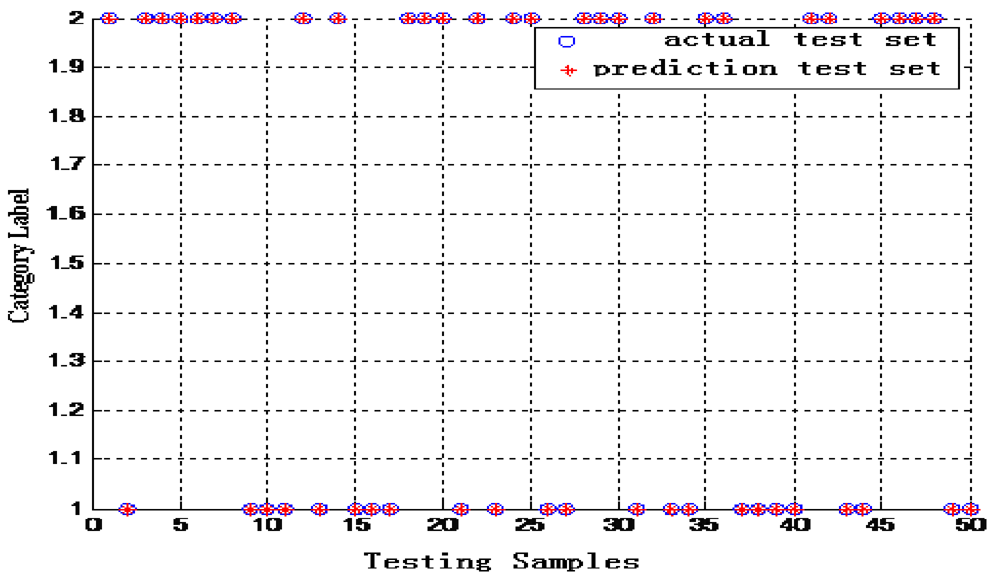


Figure 2. Classification on Voice Data before Characteristic Selection

characteristic X11, X14, X15, X20 and X23 is smaller, therefore, these characteristics with lesser weight in the category may be rejected in the category to save the category time.

### *C. Selection, Classification and Comparison for Voice Signal Characteristic*

In order to explain the effectiveness of Relief Algorithm, the classification accuracy rate for SVM is

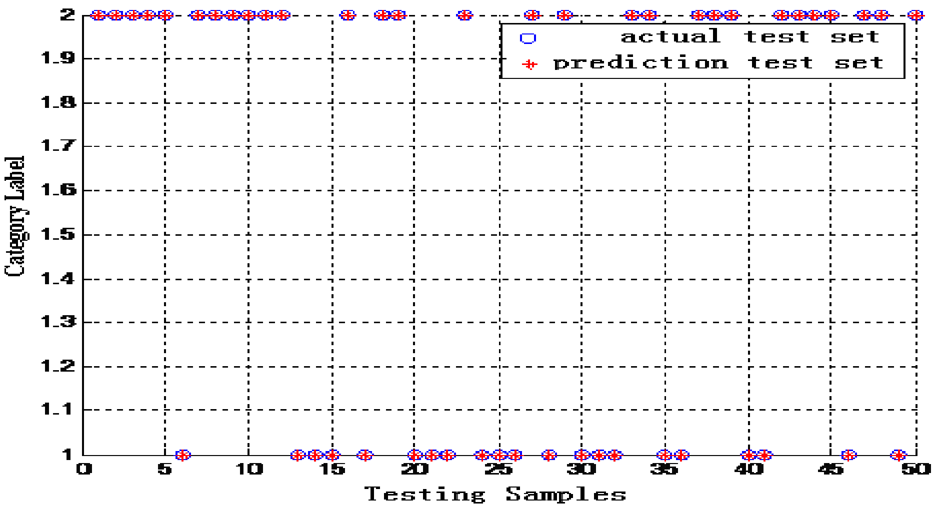


Figure 3. Classification on Voice Data after Characteristic Selection

TABLE VI.
CLASSIFICATION PERFORMANCE FOR RELIEF IN SVM

| Data Collection | Classification Accuracy Rate (%) | Classification Time (s) | Data Dimensionality |
|---|---|---|---|
| Voice Data | 96.7 | 3.392571 | 24 |
| Voice Data | 95.6 | 2.289091 | 19 |

employed to explain the effectiveness for selection of voice signal characteristic in this section, the classification result is shown as in Fig. 2 and Fig.3.

According to the Fig.2, the result can be seen that the classification accuracy rate is of 96.7%, and the classification time is of 3.392571s.

According to the Fig.3, the result can be found that The classification accuracy rate is of 95.6%, and the classification time is of 2.289091s.the classification accuracy rate is of 96.7%, and the classification time is of 3.392571s

The results can be seen from the above comparison that the Relief algorithm is effective to choose the phonetic characteristics and greatly reduce the complexity of feature selection for speech signals.

The characteristic selection for 24-dimensional MFCC parameters is performed via Relief Algorithm in this paper to find out the most important characteristic in 24-dimensional MFCC parameters, and the characteristics with lesser importance is rejected to effectively classify the voice signal. Although there is certain distance on actual application in the voice recognition for Relief characteristic selection Algorithm, the better classification performance in the test proposes a new direction for the characteristics of voice recognition.

## V. CONCLUSION

Feature selection technique not only can reduce the computational complexity, and improve the classification accuracy, but also help to find more streamlined algorithm model. The paper studies on the relief algorithm and its application in speech recognition, and finds out the most important features in the 24-dimensional MFCC parameters. What' more, it excluds some characteristics which have less importance. Finaly, speech signals are classified by using the SVM classifier. Meanwhile, the classifier obtains a better classification effect. Therefor, the study provides a new solution for the feature selection of speech recognition.

In future studies, the ReliefF and RreliefF will be explored whether have good select properties for speech recognition and evaluation index

## ACKNOWLEDGEMENTS

The authors are grateful to the referees for their valuable comments and suggestions to improve the presentation of this paper. The work is supported by National Natural Science Foundation of China (No. 51075418), Application Development Major Projects of Chongqing (No.cstc2013yykfC0034), Natural Science Foundation Key Project (No.CQ CSTC2013jjB40007), Chongqing Educational Committee Science and Technology Research Project (No.KJ111417 and No.KJ091402).

## REFERENCES

[1] P. C. Pendharkar, "A maximum-margin genetic algorithm for misclassification cost minimizing feature selection problem," Expert Systems with Applications, Elsevier, vol.40, no.10, pp. 3918-3925, 2013.

[2] Y. J. Li, D. F. Hsu, S. M. Chung, "Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic," International jouranl on artificial intelligence tools, Singarore, vol.22, no.2, 2013. [Online]. Available: DOI: 10.1142/S0218213013500012

[3] N. Garcia-Pedrajas, A. de Haro-Garcia, J. Perez-Rodriguez, "A scalable approach to simultaneous evolutionary instance and feature selection," Information sciences, Elsevier, vol.228, no.1, pp. 150-174, 2013.

[4] W. J. Hu, K. S. Choi, Y. G. Gu, S. T. Wang, "Minimum-maximum local structure information for feature selection," Pattern Recognition Letters, Elsevier, vol.34, no.50, pp. 527-535, 2013.

[5] I. Santos, X. Ugarte-Pedrero, F. Brezo, P. G. Bringas, J. M. Gomez-Hidalgo, "Noa: An information retrieval based malware detection system," Computing and informatics, Slovakia, vol.31no.1, pp. 145-174, 2013.

[6] M. Bozic, M. Stojanovic, Z. Stajic, N. Floranovic, "Mutual Information-Based Inputs Selection for Electric Load Time Series Forecasting," Entropy, Switzerland, vol.15, no.3, pp. 926-942, 2013.

[7] S. Baccianella, A. Esuli, F. Sebastiani, "Using micro-documents for feature selection: The case of ordinal text classification," Expert Systems with Applications, Elsevier, vol.40, no.11, pp. 4687-4696, 2013.

[8] S. Shilaskar, A. Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases," Expert Systems with Applications, Elsevier, vol.40, no.10, pp. 4146-4153, 2013.

[9] G. Stegmayer, D. H. Milone, S. Garran, L Burdyn, "Automatic recognition of quarantine citrus diseases," Expert Systems with Applications, Elsevier, vol.40, no.9, pp. 3512-3517, 2013.

[10] A. K. Farahat, A. Ghodsi, M. S. Kamel, "Efficient greedy feature selection for unsupervised learning," Knowledge and information systems, Springer, vol.35, no.2, pp. 285-310, 2013.

[11] A. Kaban, "Non-parametric detection of meaningless distances in high dimensional data," Statistics and Computing, Springer, vol.22, no.2, pp. 375-385, 2012.

[12] H. X. Guo, X. W. Liao, K. J. Zhu, C. Ding, Y. H. Gao, "Optimizing reservoir features in oil exploration management based on fusion of soft computing," Applied soft computing, Elsevier, vol.11, no.1, pp. 1144-1155, 2011.

[13] J. Yan, N. Liu, S. C. Yan, Q. Yang, W. G. Fan, W. Wei,etc, "Trace-Oriented Feature Analysis for Large-Scale Text Data Dimension Reduction," Knowledge and data engineering, IEEE Transactions on, vol.23, no.7, pp. 1103-1117, 2010.

[14] A. Ozcift, A. Gulten, "Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases," Digital signal processing, Elsevier, vol.23no.1 pp. 230-237, 2013.

[15] X. Y. Liu, H. Q. Wang, D. M. Xu, "The application of adaptive partitioned random search in feature selection problem,"in Advanced Data Mining and Applications 2005, 2005 1st International Conference on vol.3584, 2005,pp. 307-314.

[16] L. Xu, J. C. Sun, Y. Tian, "Feature selection using rough entropy-based uncertainty measures in incomplete decision systems," Knowledge-based systems, Elsevier, vol.36, no.1, pp. 206-216, 2012.

[17] A. K. Mohanty, M. R. Senapati, S. K. Lenka, "A novel image mining technique for classification of mammograms using hybrid feature selection," Neural computing & applications, Springer, vol.22, no.6, pp. 1151-1161, 2013.

[18] B. D. Bai, Y. Y. Wang, C.W. Yang, "Predicting atrial fibrillation inducibility in a canine model by multi-threshold spectra of the recurrence complex network," Medical engineering & physics, Elsevier, vol.35, no.5, pp. 668-675, 2013.

[19] M. M. Han, D. H. Li, "Greedy Binary Search and Feature Subset Selection," Information-an international interdisciplinary jouranl, Hosei Univ, vol.12, no.6, pp. 379-1395, 2009.

[20] X. Cheng, X. Xiao, Z. C. Wu. P. Wang, W. Z. Lin, "Swfoldrate: Predicting protein folding rates from amino acid sequence with sliding window method," Proteins-structure function and bioinformatics, Wiley-blackwell, vol.81, no.1, pp. 140-148, 2013.

[21] R. M. Luque-Baena, D. Elizondo, E. Lopez-Rubio, E. J. Palomo, T. Watson, "Assessment of geometric features for individual identification and verification in biometric hand systems," Expert Systems with Applications, Elsevier, vol.40, no.9, pp. 3580-3594, 2013.

[22] C.V. Bolon, M. N. Sanchez, B. A. Alonso, "A review of feature selection methods on synthetic data," Knowledge and information systems, Springer, vol.34, no.3, pp. 483-519, 2013.

[23] A. Ghosh, A. Datta, S. Ghosh, "Self-adaptive differential evolution for feature selection in hyperspectral image data," Applied soft computing, Elsevier, vol.13, no.4, pp. 1969-1977, 2013.

[24] K. Kira, L. A. Rendell, "Feature selection problem: traditional methods and a new algorithm,"in Artificial Intelligence - AAAI-92,1992 Proceedings Tenth National Conference on, vol.1992,1992, pp. 129-134.

[25] I. Kononenko, "Estimating attributes. Analysis and extensions of RELIEF,"in Machine Learning 1994,1994 Proceedings of the European Conference on, vol.784,1994, pp. 171-182.

**Lun Gao** is a lecturer in Chongqing University of Science and Technology, China in 2013. His main areas of interest and research are intelligent control and soft-sensing.

**Taifu Li** received his B.Sc, M.Sc and Ph.D all from Chongqing University in 1996, 2000 and 2004 respectively. Now he is a professor and supervisor for M.Sc candidate in Chongqing University of Science and Technology. His main areas of interest and research are intelligent control and soft-sensing.

**Lizhong Yao** received his B.S. degree in automation from Chongqing University of Science and Technology, China in June 2009 and his M.S. degree in Detection Technology and Automatic Equipment from Xi'an Shiyou University, China in June 2013. His current research interest includes modeling and optimization of complex systems.

**Feng Wen** received his B.S. degree in automation from Chongqing University of Science and Technology, China in June 2012. His current research interest includes modeling and optimization of complex systems.