Feature Selection Algorithm for Palm Bio-impedance Spectroscopy based on Immune Clone

Lü Lin-tao

College of Computer Science and Engineering
Xi'an University of Technology
Xi'an ,China
lvlintao@xaut.edu.cn

Li-peng, YANG Yu-xiang,Tan-fang

College of Computer Science and Engineering , School
of Mechanical and Precision Instrument Engineering
Xi'an University of Technology
Xi'an ,China
lp19911126@163.com, yangyuxiang@xaut.edu.cn

Abstract—**According to the features of Palm bio-impedance spectroscopy (BIS) data, this paper suggests a kind of effective feature model of palm BIS data— elliptical model. The model combines immune clone algorithm and least squares method, establishes a palm BIS feature selection algorithm, and uses the algorithm to obtain the optimal feature subset that can completely represent the palm BIS data, and then use several classification algorithms for classification and comparison. The experimental results show that accuracy of the feature subset obtained through the algorithm in SVM classification algorithm test can reach 93.2, thereby verifying the algorithm is a valid and reliable palm BIS feature selection algorithm.**

Keywords- palm BIS; immune clone; feature extraction; feature selection; pattern classification;

I. INTRODUCTION

Bio-impedance Spectroscopy (BIS) [1] is a kind of biological tissue testing technology which is developed with measurement technology and based on multi-frequency and complex impedance measurement. Due to varieties of defects in the current various biometric identification technology [2-4]. The author proposes that taking the BIS characteristics as a representation of biological characteristics of individual differences for identification for the first time, and developed a special palm BIS measuring instrument, measuring the arbitrary two fingers of palm except thumb as electrode consisting six loops of internal and external electrical impedance under multi-frequency condition. Research shows that BIS, which is used for identity recognition, possesses characteristics of uniqueness, universality, stability and collectability, and is completely in line with characteristic condition of data which identification requires.

In pattern recognition, there are more samples and data features in data measuring, and data which has no effect on the classification exists, and in order to improve recognition accuracy, measured data shall be processed before classification, generating the most exceptional excellent subset.

The formation of optimal characteristics is the problem of feature extraction and feature selection, Feature extraction is to find the most representative sample data, which can completely represent characteristics of this sample. The process of feature selection is essentially a process of solving, combining, optimizing problems, i.e. solving, combining,

optimizing problem. Immune clonal selection algorithm (ICSA)[5-7] introduces affinity maturation, clone, and memory mechanism based on traditional evolutionary algorithm, it also takes into account of local optimizing and global search, thereby showing good performance in solving optimization problems.

Extract and select the ellipse parameters characteristic of the BIS data with the model of oval, and use different classification algorithm for classification and comparison via combining immune clonal selection algorithm with the method of least squares, using ICSA rapid convergence to the performance of the global optimal solution. The results show that ellipse model is an effective BIS feature model, ICSA possesses better performance of feature selection, and the generated optimal feature subset possesses better performance of classification.

II. EXTRACTION AND SELECTION OF PALM BIS FEATURES

In order to extract the typical characteristics of BIS data, first establish a model of it. After analyzing the BIS data, it is found that BIS data on the coordinate system with impedance real part and imaginary part for coordinates presents special geometric model. So with the model for the feature model of BIS for feature extraction and feature selection.

A. Establishment of Palm BIS data elliptical model

In the biomedical field, using circuit model is a most commonly method in bioelectrical impedance analysis. In 1940, Kenneth S Cole puts forward Cole model[8], after measuring different frequency complex impedance, least squares method can be used to estimate model parameters Cole. It is found in the experiment that some measurement data fitting the parameters is imaginary or negative, leading to fitting the data and original data error is bigger, so its biggest defect is not perfectly fitting all the BIS data. At the same time because Cole model is not special for palm the structure of biological tissue, it can not completely describe the properties of the palm BIS, thereby leading to a higher false rejection rate.

Observation and experimental research on the palm BIS data, it is found in the measured frequency range, showing a semi-elliptical geometry model, as shown in Fig. 1, measurement data shown in the figure is two different

people 's measurement for three times. As it can be seen, elliptical model of the BIS data of each person has different parameter information, such as location, size, direction, and other information.

General algebraic form of the ellipse in the plane coordinate system can be expressed as:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \qquad (1)$$

Another more intuitional representation is the geometric parameters of the plane coordinate system, e.g. the center of the ellipse (x_c, y_c), long axis and short axis (a, b), the intersection angle H of the long axis, which is intersection angle between the long axis and the abscissa axis, and any ellipse in a two-dimensional plane can be uniquely determined by these five parameters. So the BIS information of each person can uniquely be expressed by the parameters of the model, thereby extracting parameter information for serving as feature vectors of classification and identification. As the parameter information in the general formula (A, B, C, D, E, F) entirely contains the geometric parameter information of the ellipse, which is easily to be extracted, and computational complexity is small, so select the parameter information of the general form as the data characteristic.
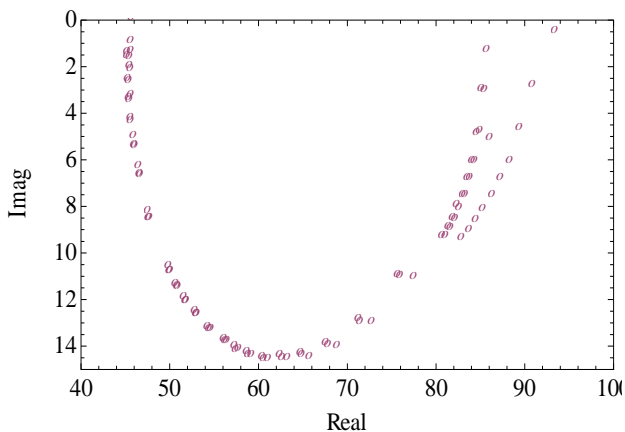

Figure 1. two people's palm BIS Wessel figure

B. Parameter feature extraction of BIS elliptical model

In order to get the parameter information of ellipse, fitellipse is necessary, and the least squares fitting is a commonly used method for ellipse fitting[9]. The general formula of the ellipse is expressed as formula （1）, as shown in Fig. 1, there is hardly any relationship between BIS elliptical model and elliptical angle, so it can expressed as:

$$Ax^2 + By^2 + Cx + Dy + E = 0 \qquad (2)$$

In order to avoid zero solution, suppose $E \neq 0$, then it becomes :

$$a \cdot x^2 + b \cdot y^2 + c \cdot x + d \cdot y + 1 = 0 \qquad (3)$$

And here $a = \dfrac{A}{E}, b = \dfrac{A}{E}, c = \dfrac{A}{E}, d = \dfrac{A}{E}$ .namely,the objective function is :

$$F(a,b,c,d) = \sum_{k=1}^{G} (a \cdot x_k^2 + b \cdot y_k^2 + c \cdot x_k + d \cdot y_k + 1)^2 \qquad (4)$$

G is the number of samples, minimum value of the objective function to determine the coefficients, to obtain minimum value of $F(a,b,c,d)$ in accordance with extremum principle ,there will be

$$\frac{\partial F(a,b,c,d)}{\partial a} = \frac{\partial F(a,b,c,d)}{\partial b}$$
$$= \frac{\partial F(a,b,c,d)}{\partial c} = \frac{\partial F(a,b,c,d)}{\partial d} = 0 \qquad (5)$$

Thus a linear equations can be obtained, then, solve the equations equation to obtain coefficient (a,b,c,d) values.

C. optimal feature selection

Directly using the least squares method does not consider the differences of each sample point error difference, it is assumed that the sample data is zero mean with a common covariance matrix, which is inconsistent with the actual situation and will lead to biased estimate for parameters, so the results of the algorithm are often unsatisfactory.

ICSA induces the whole evolution process via antibody affinity, intuitively, the higher of the fitness function the value is, the better antibody in the population is.

Define the error function as following:

$$F = \sum_{k=1}^{L} (a_i x_k^2 + b_i y_k^2 + c_i x_k + d_i y_k + 1)^2, i = 1, 2, ..., L \qquad (6)$$

L is the encode length, then the fitness function is:

$$f = 1 / F \qquad (7)$$

The main steps of the algorithm are:

- *Step1* Initialize the antibody population: Randomly generate N antibody populations, each antibody represents a combination of characteristics, the gene string length is L, which is feature vector length, encoded as $(a_1, \ a_2, ..., \ a_{L-1}, \ a_L)$, using the binary encoding, if the gene bit a_i is 1, it indicates that the corresponding feature component is selected,;if contrary is 0, it indicates is the corresponding feature component selected.

- *Step2* decoding of antibody: Each antibody is decoded into the corresponding feature combination, getting the training sample set.

- *Step3* Calculate the affinity: Use formula (7) to solve antibody affinity , and rank according to the affinity size sort.

- *Step4* Judge whether iteration termination conditions are met: the termination condition can be supposed to the fitness functions which can reach the threshold value or the number of iterations, if the condition is met, then terminate the iteration as the algorithm eventually to find the solution to determine the best

individuals in the current population; otherwise, continue.

- *Step5* Cloning operation: the essence of the clone is produce a sub-groups according to the size of the affinity in generation evolution near the candidate solutions, thus expanding the search range. clone the current k generation population, and obtain $A'(k) = \{A'_1(k), A'_2(k), A'_3(k),...\}$ Each antibody clone scale can be distributed proportionally according to antibody and original affinity size, also a fixed integer can be simply supposed.

- *Step6* Mutation operation: high frequency of mutation operation is conducted to $A'(k)$ with mutation probability pm = 1/ L for, then obtain $A''(k)$.

- *Step7* Calculate affinity: Decode each antibody in the current population $A''(k)$ for corresponding features combination , and get a new training sample, calculate affinity $\{f(A''(k))\}$ of each antibody according to (7) formula.

- *Step8* Clonal selection: In the progeny population, if there exists largest individual whose affinity is greater than the affinity of father generation, then, use this individual to replace the corresponding position of the father generation antibody, and it is added to $A(k+1)$, this process realizes the compression through the local optimal solution.

- *Step9* Calculate affinity: According to the new-obtained population, calculate affinity degree of antibody.

- *Step10* k=k+1, back to Step4.

III. RESULTS AND ANALYSIS

The palm BIS data measurement frequency range is 1 kHz-100 kHz, divided into 32 frequency point, 50 people each of them have eight times measurement.

In ICSA feature selection, the population size is 100, individual encoding length is supposed to be 32 according to measurement frequency point number, clone size is 0.2 times of the population , termination condition is combination of evolution algebra and threshold, evolution times for 500, threshold value chosen as 0.001.

Fig. 2 is composed of a set of sample parameters obtained from the fitting ellipses. Fig. 3 shows all samples of different times measurement of the same person (Interclass, intra-class), and between different people (Interclass, inter-class) of measurement value of Euclidean distance distribution. As can be seen from the Fig. 3, Elliptical model parameters possess better distinguish performance, there are still more intra-class data distribution at the position where the distance is greater than 8 , pulling down overall classification accuracy.
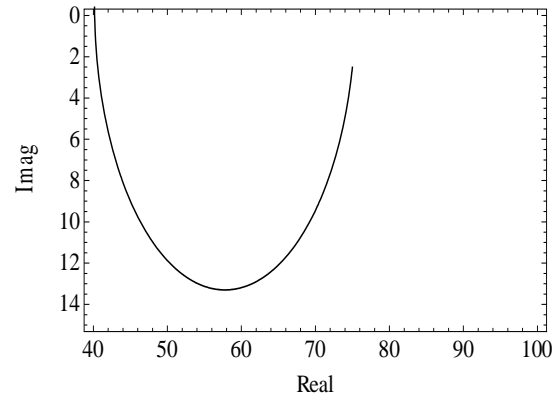

Figure 2. A set of sample data fitting results

Use all the sample data to determine classification threshold，taking false accept rate (FAR) and false reject rate (FRR) which equal to the threshold value as final classification threshold value. Use each person's of measurement data as a template, and Use Euclidean distance for classification, Receive Operating Characteristic(ROC) curve as Fig. 4, due to the above Fig. 3 problems, leading to larger FRR.
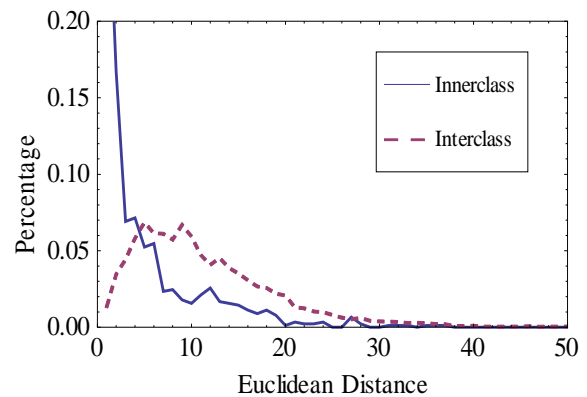

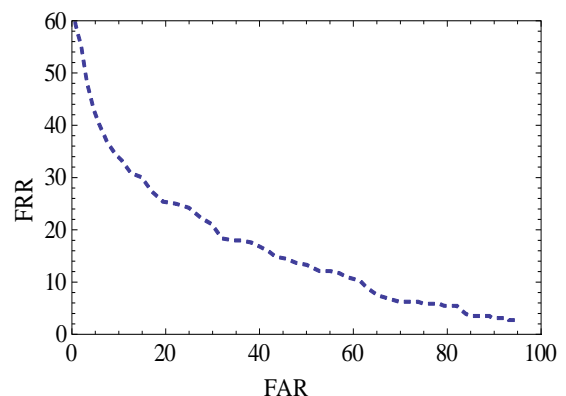Figure 3. Euclidean distance distribution of data

Figure 4. ROC curve of Euclidean distance classifier

Several results of classification methods using elliptical model parameters are listed in Tab. 1, it can be seen, using Euclidean distance for classification is not ideal, and the result is very good by using SVM classification, because SVM uses vector information in the characteristics data , compared with Euclidean distance serving as a scalar, vector contains more rich classification information. The Fig. 3 also proves the results.

TABLE I. ERROR RATES OF DIFFERENT CLASSIFIERS

Classification method	Training set error rate（%）	Testing set error rate(%)
Euclidean distance	--	24.2
k-NN (k=4)	--	22.5
Bayes Classifier	11.3	19.5
SVM	5.2	6.8

IV. CONCLUSION

From the experimental results, it can be seen that the elliptical model proposed in this paper is a kind of effective characteristic model which can characterize BIS data. Through the ICSA, select sample the optimal feature subset, from identity recognition algorithm based on the evaluation criterion of the performance, use 4 kinds of classification algorithm to generate the optimal feature subset are classified comparison and evaluation, the results show that SVM, which uses the feature vector of the vector information, has a good classification effect because the Euclidean distance is a scalar which makes the classification of the effect not good. This paper proves that the algorithm is an effective and reliable palm BIS feature selection algorithm.

REFERENCES

[1] KYLE U G, BOSAEUS I, DE LORENZO A D, et al. Bioelectrical impedance analysis--part i: Review of principles and methods. Clinical Nutrition. 2004, 23 (5): 1226-1243

[2] T. v. d. Putte and J. Keuning, "Biometrical fingerprint recognition: don't get your fingers burned," presented at the Proceedings of the fourth working conference on smart card research and advanced applications on Smart card research and advanced applications, Bristol, United Kingdom, 2001

[3] PUTTE T V D, KEUNING J. Biometrical fingerprint recognition: Don't get your fingers burned // Proceedings of the fourth working conference on smart card research and advanced applications. Bristol, United Kingdom, 2001: 289-303

[4] FERNANDO A-F. Robustness of signature verification systems to imitators with increasing skills // JULIAN F, ALMUDENA G, JAVIER G, JAVIER O-G. International Conference on Document Analysis and Recognition. 2009: 728-732.

[5] Tian Wenjie, Ai Lan, Geng Yu. A New Fuzzy Identification Approach Using Support Vector Regression and Immune Clone Selection Algorithm Proc.of CCDC'09.Guilin, China:[s.n.],2009.

[6] Berna Haktanirlar Ulutas . Sadan Kulturel-Konak. A review of clonal selection algorithm and its applications. Artif Intell Rev (2011) 36:117–138.

[7] ZHANG Xiang-rong,J IAO Li-cheng. Feature Selection Based on Immune Clonal Selection Algorithm. Journal of Fudan University (Natural Science).2004.43(5).926-929.

[8] GRIMNES S, MARTINSEN Ø G. Bioimpedance and bioelectricity basics. London: Academic press, 2008.

[9] Yan Bei,Wang Bin,Li Yuan,. Optmial ellipse fitting method based on least 2 square principle. Beijing University of Aeronautics and Astronautics. 2008.34(3):295-298.