TELKOMNIKA, Vol. 11, No. 6, June 2013, pp. 3159 ~ 3164
e-ISSN: 2087-278X

Diagnosis of Hepatocellular Carcinoma Spectroscopy Based on the Feature Selection Approach of the Genetic Algorithm

Shao-qing Wang*[1], Qiang Liu[2], Dong-yue Yu[4], Guang-ju Liang[4]
[1,2]Shandong Medical Imaging Research Institute, Shandong Jinan 250021, China
[3]Shandong Zibo Prison Hospital, Shandong Zibo 255129, China
[4]Shandong Province XinTai City People's Hospital, Shandong Xintai 271200, China
Corresponding author, e-mail: 2006wsq@163.com, 2002md@163.com, fishlet@163.com, liangguangju1@163.com

Abstract

This paper aims to study the application of medical imaging technology with artificial intelligence technology on how to improve the diagnostic accuracy rate for hepatocellular carcinoma. The recognition method based on genetic algorithm (GA) and Neural Network are presented. GA was used to select 20 optimal features from the 401 initial features. BP (Back-propagation Neural Network, BP) and PNN (Probabilistic Neural Network, PNN) were used to classify tested samples based on these optimized features, and make comparison between results based on 20 optimal features and the all 401 features. The results of the experiment show that the method can improve the recognition rate.

Keywords: *hepatocellular carcinoma, back-propagation neural network, probabilistic neural network, genetic algorithm, 31P-hosphorus magnetic resonance spectroscopy*

1. Introduction

In the clinical work, 31P Magnetic Resonance Spectroscopy (31-Phosphorus and Magnetic Resonance Spectroscopy, 31P-MRS) technology [1, 2] can use slight changes of chemical shift in information collection todetermine the human energy metabolism and body chemicals. That is currently the only noninvasive approach in studying physiological pathology changes of emerging technologies in vivo. So the evaluation and 31P-MR spectrum of disease diagnosis and treatment are important clinical significance [3, 4].

Artificial neural network is an imitation of biological brain in the information processing method. This technique can be a very good deal with multivariable nonlinear relation. It can be usedfor identification and classification through the training of complex mode. At present, the neural network in 31P magnetic resonance spectroscopy (31P-MRS) study has been widely used. Among them, the reverse transmission Neural Net-work(Back prop-agation Neural Network, BP) model is a more important artificial Neural Net-work model. BP is the advantage of network optimization with accuracy. Proba-bilistic neural network (probabilistic neural network, PNN) is a variation form of the radial basis function, also with the characteristic of the simple structure, training quickly and so on.

2. Magnetic Resonance Phosphorus Spectrum

All cases were selected randomly from Shandong medical imaging research institute from Jan. 2008 to Jan. 2009. There are 130 sample data, including 45 cases for hepatocellular carcinoma, and 28 cases for liver cirrhosis, 57 cases for the normal. In the normal group, with the conventional check, no history of liver disease is recognized, all liver cirrho-sis and HCC patients in 31P-MRS after research all cases are confirmed by biopsy pathology. 31P-MRS can measure the seven formants (Figure 1): single phosph-ate ester(phosphomonoester, PME), inor-ganic phosphorus (inorganic phosphorus, Pi), phosphoric acid two fat (phospho-diester, PDE), phosphoric acid creatine (phosphocreatine, PCr), adenosine tripho-sphate (α-ATP, β-ATP, γ-ATP) [5]. 31P MRS curve describs the main index: chemical shift, wave integral area,

Received December 29, 2012; Revised April 2, 2013; Accepted April 17, 2013

seven parameters of the ratio of the value. Chemical shift spectroscopy curve in the abscissa denotes the compounds to the location, with the magnetic resonance frequency in parts per million (PPM). Wave integral area (line the following areas) represents the compound's concentration, and resonance is proportional to the number of the nucleus; Peak refers to the height of the curve y-coordinate, representatives of the compounds signal strength unit.

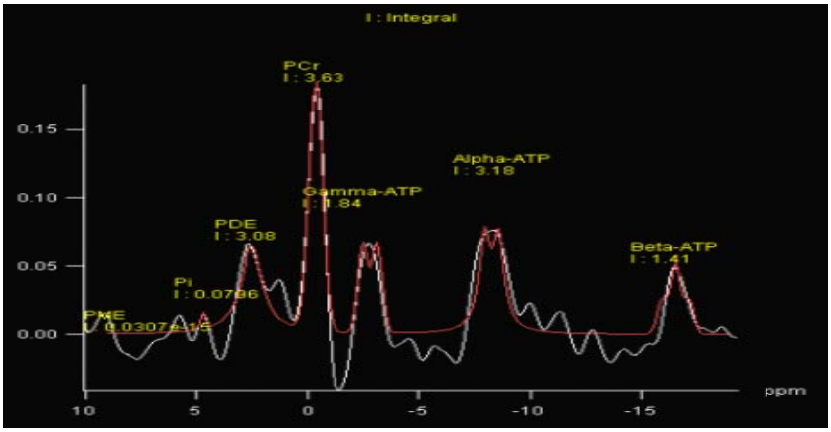

Figure 1. 31P Spectrum Diagram of the Hepatocellular Carcinoma Organization

3. Reverse Transmission Neural Network (BP)

BP network [6, 7] is a layered structure of feedforward neural network, the network is divided into input layer, hidden layer and out-put layer, as shown in Figure 2. Among them, $x_1, x_2, ..., x_i$ in-dicate the input of the network, $y_1, y_2, ..., y_j$ indicate output vector, corresponding to three experiments of output: hepato-cellular carcinoma, liver cirrhosis and normal. Here hidden nodes choose number 25, take Sigmoid function as neurons transfer function.

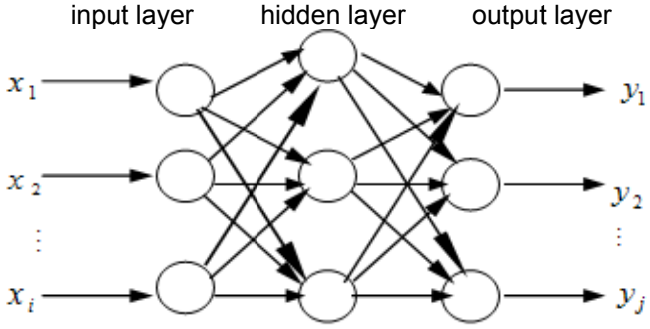

Figure 2. BP Neural Network Structure

Hypothesis the first k-1 layer in the first i neurons $y_i^{(k-1)}$ as the input, y_j^k as output, the relationship between input and output is:

$$y_j^{(k)} = f[\sum_{i=1}^{N_{k-1}} w_{ij}^{(k-1)} \cdot y_i^{(k-1)} + \theta_j^{(k)}], \quad (1)$$

$$f = \frac{1}{1+\exp(-x)}, \quad (2)$$

The $w_{ij}^{(k-1)}$ in type (1) is the link weight between the first k-1 layer for the first i neurons to the first k layer in first j neurons; N_k is the number of neurons in the first k layer; F (x) is the transfer function of neurons.

4. Probabilistic Neural Network (PNN)

Probabilistic neural network[8] is made up of the input layer, radial base and competition layers. The network structure as shown in Figure 3. $LW_{1,1}$ is the weight matrix connecting input layer and radial basis function (RBF) layer, $I_{1,1} = Q_R$, Q is the number of target for input , that is the number of neurons in hidden layer ; R indicates the input vector dimension; P is for waiting characteristic vectors (R x 1); b^1 is the threshold of the radial basis fun-ction layer, which belongs to the threshold vector (Q×1)); a^1 is the output vector of the radial transfer function in the radial basis; $LW_{2,1}$ is the weight matrix connecting radial base and competition layer; C is a transfer function for the competition.

When the network provides an input mode vector, radial grassroots compute the input vector and the input vector of the samples, between two vectors the distance is:

$$D=(\| P-I_1 \|,\| P-I_2 \|,\cdots,\| P-I_Q \|)^T \tag{3}$$

Generate vector D multiply by b^1 and multiplication, get n^1, that is the input vector of the radial basis transfer function:

$$n^1=(b^1 \| P-I_1 \|,b^1 \| P-I_2 \|,\cdots,b^1 \| P-I_Q \|)^T \tag{4}$$

Then, the radial basis function for the output:

$$a^1 = Radbas(n^1) = (a_{1,1}, a_{1,2}, \cdots, a_{1,Q})^T \tag{5}$$

$0 \le a_{1,i} \le i(i = 1, 2, \cdots)$, In the weight of the radial grassroots according to the model category sum a^1, get probability vector:

$$n^2 = (n_{2,1}, n_{2,2}, \cdots, n_{2,K})^T = T \times a^1 \tag{6}$$

The dimension of n^2 is K, each component corresponding to a model category.

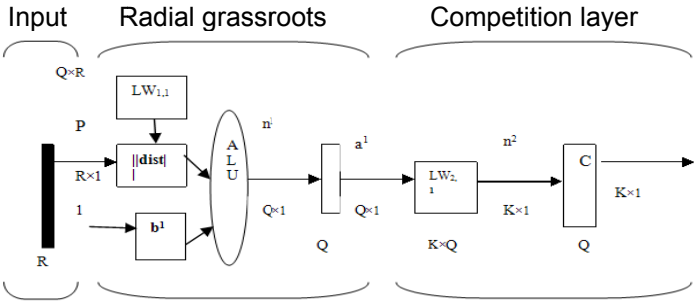

Figure 3. Probabilistic Neural Network Structure

Finally, the vector passed through a competitive transfer function C, calculate the probability of each pattern appears, through the competition for the probability of the largest transfer function elements corresponding out-put 1, otherwise the output is 0. Here, we adopt

Euclidean norm to measure weights of the distance between the input vector and the weights vector. At the same time, using reflect probability density of gaussian function as the transfer function of hidden layer (Be e^{-n^2}, including n is the input value of the radial basis function neurons).

5. Results

This experiment is based on three data set: I. the medical feature set collecting from the liver 31P-MRS of the 20 features out, 20 characteristic separately referred to PME, Pi, PDE, PCr, ATP(α, β, γ) chemical shift and the area under the peak and PME/PDE, Pi/PDE, PCr/PDE, α-ATP/PDE, β-ATP/PDE, γ-ATP/ PDE; II. 31P-MRS data of 401 all the spectrum characteristics; III. Using GA algorithm to select the best out of the 20 features.

5.1. Based on BP Neural Network Experiment

Table 1. 3-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	84	73.6	91.2	56.8
II	75.1	64.3	92.9	37.5
III	82.2	87.1	95.1	28.9

Table 2. 5-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	80.4	73.6	89.8	125.48
II	77.8	70	89.8	78.2
III	82.2	80.7	90.9	62.8

Table 3. 10-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	81.5	79.3	90.2	172
II	74.2	72.1	93.7	133
III	84.4	87.1	95.8	89

5.2. PNN Neural Network based on Experiments

Table 4. 3-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	71.1	70.7	92.6	3.14
II	77.8	71.4	91.2	3.69
III	78.2	67.8	92.9	2.83

Table 5. 5-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	68.6	70	92.9	5.48
II	78.2	70.7	90.5	5.73
III	78.7	72.1	90.9	4.14

To test and verify the validity of the method of GA-BP, this paper respectively uses the 3-fold, fold 5-10-fold cross validation methods, to randomly divide the feature setsinto training set and test set, and to count the test set an average of 10 times identification accuracy. The results shown in table 1, 2, 3, and the experiment in three different feature set the results were compared.

Table 6. 10-fold Experimental Results

Feature Set	Carcinoma(%)	Liver cirrhosis(%)	Normal(%)	Running time(S)
I	80	75	89.5	25.48
II	80	92.8	80.7	11.61
III	82.2	82.1	87.7	7.23

6. Discussion

Genetic algorithm (Genetic Algori-thms, the GA) [9, 10, 11] was proposed firstly inthe university of Michigan by John Holland in 1975, it is a kind of biological by natural selection and natural Genetic mechanism of random search algorithm. It is a kind of group, operation objects are all individuals in the group. Through the choice, a new generation of groups is produced by cross and variation operation. As a kind of efficient parallel, its main characteristic is the searching strategy in group and individual information exchange, automatic acquisition and accumulation of the knowledge of the search for the space can be achieved in the search process, and the optimal solution can be got in the adaptive control search process.

From Table 1, which can be seen after GA feature selection of the classified accuracy, all spectrum is significantly higher than the original one, liver cirrhosis recognition rate increased from the original 64.3% to 87.1. Table 3 based on data set III in the recognition rate of normal as high as 95.8%. Compared to medical 20 characteristics, using GA 20 feature selection, running time also greatly reduced. Table 3, the use of medical 20 feature, run 10 times need 172 seconds, and use the feature extraction GA 20, run 10 times only 89 seconds.

From Table 4, 5, 6, which can be seen PNN neural network based on the hepato-cellular carcinoma diagnosis also achieve high accuracy. And medical feature 20, compared the choice of GA 20 characteri-stics, we can draw the higher recognition rate. In table 6, based on GA choice of 20 characteristics, the recognition rate of cancer by more than 2.2% out.

A combination of the experimental results, the experiment based on BP is slightly below the accuracy of PNN, but from running time to see, the cost of the experiment PNN time be much less. To contrast Table 1 and table 4, the cost of the experiment time based on BP is 15-20 times of PNN.

This neural network particularly is much more suitable for solving pattern classification problems, which can realize fault detection and diagnosis. In the model classification, its advantage is that it can use linear learning algorithm to complete before nonlinear algorithm work, mean-while, it can keep the characteristics of high precision nonlinear algorithm.

7. Conclusion

Through the above analysis, to use the genetic algorithm in feature selection and features deletion have small contribution to the correct classification. Thus, it influence the characteristic classification, and can find out the problem space to represent the optimal feature.

The experiment proves that the genetic algorithm can overcome some of the pitfalls of neural network in different extent, which means it can make use of medical imaging technology and artificial intelligence technology to improve the combination of sample classification accuracy that is the diagnosis accuracy rate.

Acknowledgments

This work was supported in part by Shandong province natural science fund projects under Grant Nos. Y2006C96, ZR2010CM051, Y2008G30, and Shan-dong medical health science and technology development plan project No. 2009 HZ081.

References

[1] Baopeng Li, Qiang Liu. The present research situation and development trend of Magnetic resonance spectroscopy technology in the liver tumor. *Journal of Medical Imaging*. 2008; 18: 1461-1463.
[2] Yubao liu, Changhong Liang, Zhonglin Zhang. Preliminary clinical application research of the ^{31}P MR spectral imaging in the liver. *Journal of Practical Radiology*. 2006; 22: 417-420.
[3] Jiang Lin, Kangrong Zhou, Phospho-rus-31 magnetic resonance spectrum analysis in the liver, *J. of Fudan (medical sciences*). 2006; 33: 283-285.
[4] Wu Bi, Bin Song, Yang Yang. The clinical application of liver ^{31}P-MR spectral imaging. *China general surgery basic and clinical magazine.* 2005; 12: 624-627.
[5] Shah N, Sattar A, Benanti M, Magnetic resonance spectroscopy as an imaging tool for cancer. A Review of the Literature. *Journal Am Osteopath Assoc*. 2006; 106: 23-27.
[6] Juanjuan Lu, Hong Chen, The research progress of BP neural network. *Control engineering*. 2006; 13: 449-451.
[7] Negnevitsky M. *Artificial Intelligence: A Guide to Intelligent Systems,* 2nd Ed., Addison Wesley, England. 2003.
[8] Banghua Yang, Guozheng Yan. To improve brain computer interface the midbrain electrical signal recognition rate based on genetic algorithm and probabilistic neural network. *Journal of Shanghai Jiaotong university*. 2005; 1: 1689-1692.
[9] Xiaoping Wang, Liming Cao. Genetic algorithm theory and software realization, Xian: Xian Jiaotong university Press, 2002
[10] Minqiang Li, Jisong Kou, Dan Lin, Genetic algorithm the basic theory and application, Beijing: science press. 2002: 9-11.
[11] Zixing Cai, Guangyou Xu. *Artificial intelligence and its application.* Beijing: China university press away. 2000. I17.