Research on liver disease diagnosis basd on RS_LMBP neural network

Dong Xu, Hai-yan Fu* ,Wen-Juan Jiang
School of Information Science and Technology Hainan Normal University
Haikou, China
59125057@qq.com, 48318159@qq.com , may_jwj@163.com

Abstract— **For the traditional way of Liver disease diagnosis, there exists a certain subjectivity, and easily missed diagnosis and misdiagnosis. The use of a single neural network can not eliminate the redundant information among various indicators, resulting in diagnostic accuracy is not high. In order to improve the correctness of the early diagnosis of liver lesions, This paper proposes a liver disease diagnosis based on a combination model of rough set theory (RS) and LMBP neural network (RS_LMBPNN). The model first, the use of rough set theory powerful numerical analysis capabilities of each clinical diagnostic criteria attribute reduction, thereby reducing the LMBP neural network training data, simplifies the network structure, and then after the reduction of data import LMBP neural network training, verification experiment was carried out in the end. The simulation results show that compared with the single LMBP neural network model, the combined model can speed up the training speed of the network, have stronger learning ability, and further improve the diagnostic accuracy.**

Keywords—rough set theory; neural network; Liver lesions

I. OVERVIEW

Various diseases threat to human health seriously, the incidence of malignant lesions mortality rates are rising rapidly, at present, the liver pathological changes has become one of the main diseases of harm human health, the early diagnosis is the key to the treatment of liver lesions. Through the study found that the doctor to the patient's disease diagnosis is often through many years of clinical experience, including reading CT slices, but the doctor individual clinical experience is limited, so the knowledge hidden in the cases of data to help doctors to diagnose the disease has a significant meaning. However, how to extract useful knowledge from the data of these cases is a problem, because the original cases include hundreds of attributes of the patient.

Traditional diagnostic methods are based on the doctor's examination of the patient's CT images, data, previous cases, family cases and doctor's experience of the patient's diagnosis. There are some subjectivity in the process of diagnosis, prone to misdiagnosis and missed diagnosis. Application of computer with the help of the method of data mining auxiliary diagnosis model is established and effective improve the accuracy of disease diagnosis expert system is a very promising direction. Neural network can help diagnosing liver lesions, however, due to the existence of redundant test indicators, on the one hand to increase the inspection items, increase the burden on the patient's cost; on the other hand, interference with the clinical diagnosis of the disease, affect the diagnostic results .The neural network can not effectively eliminate the redundant information of the test index, and also make the diagnosis result is affected.

This paper choose the BP neural network as liver lesions early diagnostic methods, but as a result of the BP neural network algorithm with gradient structure, network objective function decline will gradually slow, slow convergence speed, easy to appear the defects such as local minimum, so often leads to error diagnosis. Therefore, based on LM algorithm (Levenberg - Marquardt) improves the BP neural network design, LM algorithm is an improved algorithm based on gauss - Newton method, this algorithm not only has the local characteristics of the gauss - Newton and has the global features of gradient method, it is more stable than the gradient method, and its convergence rate is fast.

Aiming at the redundant information between the indexes of clinical examination of liver lesions, a diagnosis algorithm of liver lesion based on rough set and LMBP neural network (LMBPNN) is presented in this paper. The rough set theory is used as the preprocessing system of LMBPNN, the reduction rules and rules are extracted from the clinical test indexes, the reduced rule set is input into LMBPNN for training, then the samples are predicted, The diagnosis model of RS_LMBPNN liver lesion was established, and the simulation experiment was carried out. The results show that the rough set attribute reduction can not only eliminate redundant information among clinical indexes, but also simplify the structure of LMBPNN, improve the convergence speed of the network and the generalization ability, meanwhile the diagnostic accuracy is greatly improved.

II. THE LIVER DISEASE DIAGNOSIS ALGORITHM BASED ON RS_LMBPNN

A. RS - LMBPNN Model

Based on the advantages of rough set and LMBPNN, RS_LMBPNN model is put forward, for clinical diagnosis of liver disease cases. Because LMBPNN can not eliminate the noise of clinical test indicators, the choice of the number of hidden layer nodes has not yet theoretical guidance, often due to network structure redundancy caused by slower convergence, easy to fall into local minimum, different accuracy and other issues, Leading to clinical diagnosis is wrong. In order to solve these problems effectively, the following corrections are made during the research:

- Rough set theory was used to discover the redundant information between the clinical test indexes, to reduce

978-1-5090-4840-3/16 $31.00 © 2016 IEEE
DOI 10.1109/CIS.2016.155

the parameters, to select the main index to affect the diagnosis of liver lesions, to eliminate the noise.

- To consider in the clinical test sample is limited, affect LMBPNN convergence speed, and LM algorithm is faster than gradient algorithm convergence speed, and the algorithm is more robust, Therefore, the BP neural network is trained and learned by the LM algorithm in the study.

B. The Rough Set Theory

Rough set (RS) theory is used to dealing with imprecise and incomplete data of a mathematical method, the RS theory in processing uncertain information, do not need a priori knowledge condition constraints, can effectively analyze the incomplete data and reasoning, and thus to discover the internal relation between data, is used to extract useful features to simplify the processing of information.

In the RS theory, Knowledge is described by the basic feature and the characteristic value of the specified object. knowledge system S is defined as:

$$S = (U, R, V, f, g) \quad (1)$$

Among $U = (x_1, x_2, \cdots x_n)$ is domain of discourse, Said all the collections of objects, $C \bigcup D = R$ for the non-empty finite set of attributes, subset C is attributes, D as the decision attribute set, $V = \bigcup_{r \in A} V_r$ is a collection of attribute values, Range of values of the V_r for attributes , $f : U \times R \to V$ is an information function , used to determine the attribute values of the each object x_i in the U, When there is any $x_i \in U, r \in R$ then $f(x_i, r) = V_r$, And g by condition attribute to decision attribute set of a map:

$$g : C \to D \quad (2)$$

There is any attribute subset $P \subseteq R$, if $x_i, x_j \in U, \forall_r \in P$, if and only if $f(x_i, r) = f(x_j, r)$, x_i , x_j cannot distinguish , can be marked as $Ind(P)$, namely, said x_i , x_j as the equivalence relation. If $r \in P$, when the formula (3) set up,

$$Ind(P) = Ind(P - \{r\}) \quad (3)$$

By type (3), the effect of Properties r is not big , belongs to the redundant attributes and omission will not affect the characteristics of object description after. Using this feature, when liver disease clinical diagnosis, can from several test indicators, extract to really play a key role for disease diagnosis index, simplifies the BP neural network input, eliminates the redundant information.

C. The BP Neural Network

BP network is a kind of multilayer feed forward neural networks, generally composed of input layer, hidden layer and output layer, with all the interconnection between layer and layer, there is no mutual connection between the units the same layer. Hidden layer for a layer or layers (see figure 1), the basic processing unit of the BP network (except for the output layer unit) for nonlinear input-output relationship, the transfer function of neurons generally adopts s-shaped transformation function. After determine the structure of the network, the use of input and output sample set for training, namely the network weights and threshold of learning and training, so that the network to achieve the given input/output mapping relation.

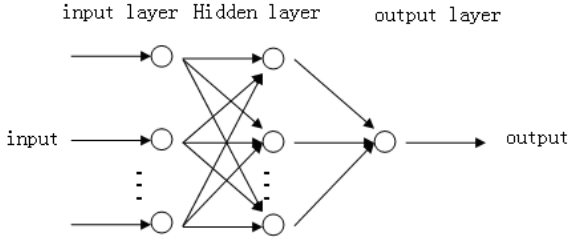

Figure 1. the BP neural network structure

D. Used for LM Algorithm of BP Neural Network Learning

In the forward neural network type, set as the first k iteration of weights and threshold vector, for the new weights and threshold of vector, the specific calculation steps are as follows:

1) For a given initial vector, set the search step length = 0.0011, the coefficient = 10, precision =10-17.

2) To calculate the error vector:

$$e(x^{(k)}) = [e_1(x^{(k)}), e_2(x^{(k)}), \cdots, e_M(x^{(k)})]^T \quad (4)$$

Among them, $e(x^{(k)})$ is error, $i = 1, 2, \ldots, N$, Jacobi matrix can be obtained:

$$J(x^{(k)}) = J_{ij}(x^{(k)}) = \frac{\partial_{e_i}(x^{(k)})}{\partial_{x_j}} \quad (5)$$

3)By solving the linear equations, and the search direction $\Delta x^{(k)}$:

$$\Delta x^{(k)} = -[J(x^{(k)})^T J(x^{(k)}) + u_k I]^{-1} J(x^{(k)})^T f(x^{(k)}) \quad (6)$$

In the process of solution of equations, if found the rank of matrix $J(x^{(k)})^T f(x^{(k)})$ is not N, Directly take $\Delta x^{(k)}$ is negative gradient direction :

$$\Delta x^{(k)} = -\frac{1}{2} \nabla E(\Delta x^{(k)}) = -J(\Delta x^{(k)})^T (\Delta x^{(k)}) \quad (7)$$

∇ said gradient, $E(x)$ said error index function :

$$E(x) = \frac{1}{2} \sum_{i=1}^{N} e_i^2(x) \quad (8)$$

4) For linear search:

$$\Delta x^{(k+1)} = \Delta x^{(k)} + \lambda_k \Delta x^{(k)} \quad (9)$$

Which λ_k satisfy the type:

$$E(\Delta x^{(k)} + \lambda_k \Delta x^{(k)}) = \min E(\Delta x^{(k)} + \lambda \Delta x^{(k)}) \quad (10)$$

5) If meet $\left\|\Delta x^{(k+1)} - \Delta x^{(k)}\right\| < \varepsilon$, then get the solution x_{opt} , stop counting; Otherwise to (6).

6)if $E(x^{(k+1)}) < E(x^{(k)})$ ， then $u_k = u_k / 2$ ， $k = k+1$ ， to（2）； otherwise $u_k = 2u_k$ 。to（3）。

E. RS_LMBPNN Liver Disease Diagnosis Process

Based on rough set, this paper uses the BP neural network LM algorithm based on the diagnosis of liver lesions. Because LMBPNN do not distinguish good clinical examination index of the noise information, so as to make the network structure is more complex, reduces the network generalization ability and diagnosis of the correctness. And rough set theory can well remove the redundant attributes and knowledge can simplify the expression of space dimension, thus reduce the input dimension of LMBPNN, therefore, combines rough set with LMBPNN can better use of their advantages. We collected data from hospital patients with liver disease, select clinical indicators from these data, to deal with various indicators, using rough set theory to select one of the most representative test indicators, eliminate redundant information between the indicators. In this paper, we design three layers of LMBPNN, the number of hidden layer nodes of LMBPNN is selected using Golden section optimization method, and then to learn LMBPNN training, until the network meets the set accuracy requirements or to achieve the prescribed number of cycles, and store the network weights and threshold. With the training of good LMBPNN, the test indicators of the patient to be diagnosed will be input to the training of good LMBPNN, diagnosis, and output the results. Algorithm flow chart is shown in Figure 2.

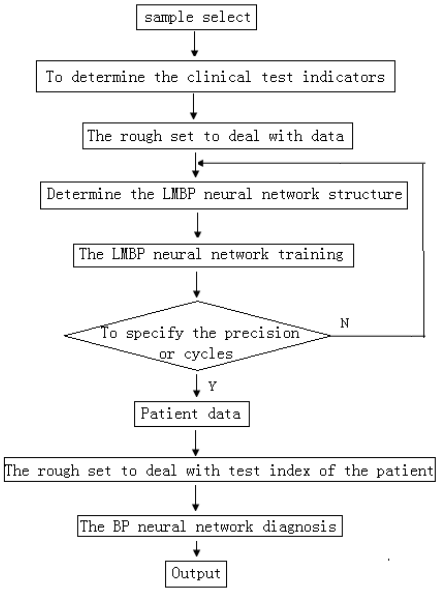

Figure 2. RS_LMBPNN liver Disease Diagnosis Process

III. EXPERIMENTAL SIMULATION

A. Clinical Inspection Index of Liver Lesions Build

According to the related research of liver lesion, this paper collected a large number of medical CT liver images, search and study of relevant literature at home and abroad, analysis of medical CT images, and ultimately determine the seven characteristics of the CT image to be extracted, add these seven features together with the patient's existing features, which constitute the conditional attribute set of decision table C, the characteristics of patients have lesions including age, gender, position in the CT image (left or right) and other clinical parameters of liver lesion. C = {Gray characteristics, energy, entropy, maximum aim-listed probability, contrast, deficit, the correlation coefficient, the Age, sex, left or right}, whether patients with liver lesions as decision set D, D = {0, 1}, where 0 means liver lesions, 1 means non liver lesions. Then, 270 of the 300 cases were collected as training set, and 30 were treated as the prediction set.

B. Different Algorithm Model of Learning Training Process

The collected example data import ROSETTA software of rough set, parameter reduction, D reduction after computing the multiple C, selects 8 indicators of liver disease diagnosis plays a major role in 10 clinical test indexes, namely {Gray characteristics, energy, entropy, maximum aim-listed probability, contrast, deficit, the correlation coefficient}, then the first 270 cases as a training set, LMBPNN structure for 8 x 8 x 2, namely after rough intensive Jane 8 index as input variables, the number of neurons in hidden layer for 8. In order to better illustrate the advantages of coarse RS_LMBPNN model, this paper also used without reduction of LMBPNN contrast experiment was carried out, the network structure of 13 x 8 x 2. The algorithm program is implemented on Matlab7.01 platform, two algorithm models are trained for 500 times. From the Figure 3 and Figure 4, we can find that the convergence rate of LMBPNN is obviously faster than that of LMBPNN after rough set reduction, and the former is much higher than that of the latter.

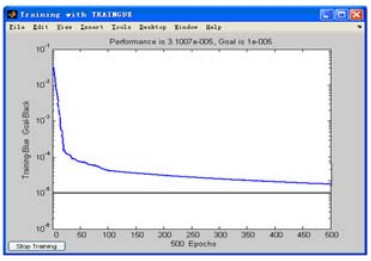

Figure 3. RS_LMBPNN Training Process

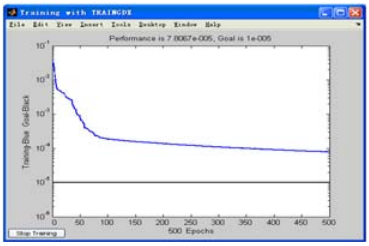

Figure 4. LMBPNN Training Process

C. Different Algorithm Model Results

After the network is trained to meet the requirements, the 30 prediction set data will be entered into LMBPNN according to the requirements. The two algorithms are tested and the test results are shown in Table 1.

Table 1. Two Prediction Results of The Algorithm.

Serial number	Actual value	LM BPNN	RS_LMBPNN	Serial number	Actual value	LM BPNN	RS_LMBPNN
1	0	0	0	16	0	1	0
2	0	0	0	17	0	0	0
3	1	0	1	18	0	0	0
4	0	0	0	19	1	1	1
5	1	1	1	20	1	1	1
6	1	1	1	21	0	0	0
7	0	0	0	22	1	0	0
8	0	0	0	23	0	0	0
9	1	1	1	24	1	1	1
10	0	0	0	25	0	0	0
11	0	0	0	26	1	1	1
12	1	1	1	27	0	0	0
13	0	0	0	28	0	0	0
14	0	0	0	29	1	1	1
15	1	1	1	30	0	0	0

The Table 1 shows, in RS_LMBPNN , using rough set to eliminate noise of the 30 prediction results, only 22 sample prediction errors, whereas untreated rough set LMBPNN, the predictive results of the 3, 16, 22 samples appear a mistake, The correct rates of the two kinds of LMBPNN are shown in Table 2.

Table 2. Forecast Accuracy

	LMBPNN	RS_LMBPNN
condition attribute	C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}	C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_{10}
Prediction accuracy	90%	96.67%

From Table 2 can get the following conclusion:

1)LMBPNN predicted results significantly below RS_LMBPNN, the forecast accuracy of 90%, while the latter achieved 96.67% accuracy.

2)RS_LMBPNN by clinical examination index to eliminate redundancy, makes the condition attribute reduced from 10 to 8, LMBPNN can reduce amount of calculation and further simplify network structure.

3) Rough set is sensitive to the noise of the data, can effectively eliminate the redundant information test indicators, LMBPNN has strong fault tolerance, but does not have the function of eliminating redundant information. RS_LMBPNN algorithm combines rough set theory and LMBPNN integration, which can give full play to the advantages of the two algorithms, so the prediction accuracy is greatly improved.

IV. Conclusion

The liver disease has become one of the main diseases of harm to people's health, due to the influence the indicators of liver disease diagnosis has the redundant information, there have been some misdiagnosis phenomenon. In this paper, according to the characteristics of liver lesion of CT image, proposes a diagnostic method of rough set combined with LMBPNN liver disease. By rough sets to eliminate redundant information between the clinical test indicators, effectively simplify the neural network structure, accelerate the network training speed. The BP neural network is improved by using LM algorithm, which makes the BP neural network more robust, and the convergence speed is faster. Through the simulation test, test and analyze the effectiveness and superiority of RS_LMBPNN. The results showed that compared with the single LMBPNN RS_LMBPNN, the prediction accuracy is greatly increased, obtained more satisfactory results, is a kind of effective method for diagnosis of liver lesions.

ACKNOWLEDGMENT

Thanks to the support by National Natural Science Foundation of China (No. 71461008).

References

[1] Ming-quan Ye, Chang-rong Wu, Xue-gang Hu, “Research and application of medical data mining based on rough sets”. Computer engineering and application, 2010 46-48 (21), pp. 232-234.

[2] LI Qing-meng, NIE Sheng-dong, “Neural network technology and its application in medical image processing”. Chin J Med Imaging Technol, vol27,No6,2011, pp. 1291-1293.

[3] Chen Yan, “The application of medical diagnosis based on rough set theory”. Science and technology in western China, vol14,No12,2015, pp. 108-110.

[4] Su Jing,Yang Wujun “Image Segmentation Algorithm Based on BP Neural Network”. Industrial control computer, vol28,No12,2015, pp. 29-31.

[5] De Santis A, Iacoviello D. “Discrete image modelling for piecewise constant segmentation by artificial neural networks”. Image Processing, 2008,2(1), pp.37-47..

[6] Verma B, McLeod P, Klevansky A,”A novel soft cluster neural network for the classification of suspicious areas in digital mammograms”. Pattern Recognition, 2009,42(9), pp.1845-1852.