Application of Decision Tree and Neural Network Fusion Model

Liping Zhao [1], Qilianag Shu [2], Bingxiang Liu [1]

[1]School of information engineering, Jingdezhen ceramic University

[2]College of science and technology, Jingdezhen ceramic University

Abstract

Based on the analysis of the complementary advantages and disadvantages between the decision tree and neural network and on the basis of the established neural network fusion model of decision tree, pointed out the feasibility of the application of CRM system of data mining in the banking industry in the development and effectiveness based on specific application of the fusion model in bank customer relationship in the management, and gives a case analysis of bank customer relationship management. From the results of the case analysis, provide routine reports for decision makers, marketing plans and customer contact management strategies in order to develop differentiated, a new research method and analysis methods of customer relationship management for the bank. The method used in this paper is to classify and forecast the customer information pretreatment and BP neural network by the decision tree classification analysis.

Keywords

Data mining, CRM, Decision Tree, Neural Network.

1. Customer relation management and data mining

1.1 Customer relation management

Customer relation management (CRM) [1] is a new management method to improving the relationship between the enterprise and the customer's. Its core idea is "Customer-Centric Management". It is apply to enterprise marketing, sale, service and technique etc. On the other hand, CRM is a kind of concept, and is a set of management software and technique, make use of the data of CRM system excavation tool, the enterprise can collect, track and analyze the information of each customer by the data mining tool of CRM.

1.2 Data mining

Data Mining is a process of collecting unknown and valuable information and knowledge from a lot of incomplete, noise, and random, fuzzy Data. But these information and knowledge is cannot achieve depend on a pure data search.

2. Decision tree classifier based on BP neural network model

Decision tree study is a kind of concluding study method on the base of examples, which can be used to form classification and predictive model. It can be infer classification rules from a set of no order, no rules of case. It uses a self-downward recursive method,in the internal node, the attribute values are compared and the branches are judged from the node to the nodes, and finally the conclusion is obtained. Therefore,a path from the root to the leaf node corresponds to a conjunctive rule, and a decision tree corresponding to a group of disjunctive expression rules.

2.1 The establishment of decision tree classifier algorithm

For each attribute of examples concentration, we first compute the influence degree of all attribute values and then compare them, we can know that the influence degree more big attribute provide to the classified more information .Comparison in turn, we can make sure an attribute to classified influence degree size and construct a decision tree based on the size of the decision tree.

2.1.1 Specific steps of the main algorithm [2]

Suppose there are only two categories, respectively P, N:

(1) Random choose a subset contain positive example and anti-example from train concentration (called "window ");

(2) Form a decision tree to the current window use “decision tree algorithm";

(3) On the training set (except window) using the example of the decision tree classification, find out the wrong example;

(4) If exist wrong judge of example, insert them into the window, repeat step 2, otherwise be over.

The main algorithm flow as shown below. The PE and NE is the positive example set and the anti-example set respectively and constitute to training set together. PE', PE'' and NE', NE'' are subset of the positive example set and the anti- example set respectively.

In the main algorithm, the decision tree will be different in each iteration cycle, As shown in Fig. 1.

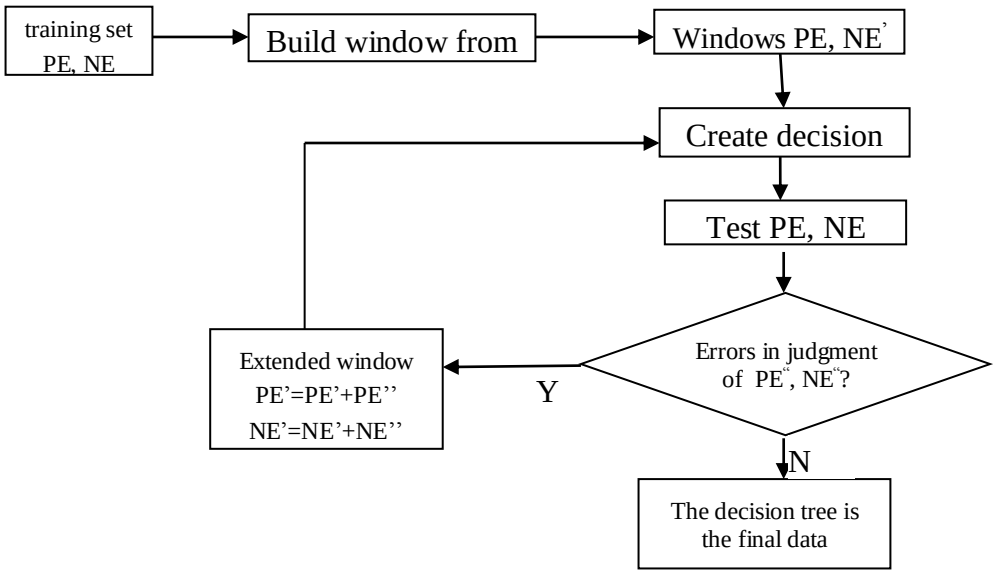

Fig. 1 the decision tree algorithm flow chart

2.1.2 The decision tree algorithm

(1) To current example subset, compute the conditional probability of each characteristic;

(2) Choose the characteristic A_k has biggest conditional probability;

(3) The examples are classified into a same subset if their values are equal at A_k and several values are obtained from A_k.

(4) Recursive call the contribution algorithm to the subset not only contain a positive example but contain anti- example;

(5) If the subset only contain positive example or anti- example, mark P or N on the corresponding branch and return.

2.2 The establishment of BP Neural Network

The basic principle to process information in BP network model is : input signal X_i through the middle node (hidden layer) applied to the output nodes and through a nonlinear transformation, get output signal Y_k. Each sample of the network training includes the input vector X and the expected output t, the deviation between the network output value Y and the expected output value t, By adjusting the input nodes and hidden nodes of the connection strength value of W_{ij} and the hidden nodes and the output nodes connecting strength T_{jk} and threshold, the error decreases along the gradient direction, after repeated training, determine the network parameters corresponding to the

minimum error (weights and thresholds), the training is stopped. At this time, the trained neural network can be able to input information to the similar samples, and the output error is minimized.

In this paper, the existing neural network model is improved to improve the iterative speed of BP learning algorithm.

(1) Jump learning: Some samples may lead to a very small error in the learning process, if they are still in the forward and backward propagation calculation, then the time wasted, and then affect the network learning speed. In order to prevent this phenomenon, you can set a minimum error. If the output error of a sample is smaller than that of the minimum error, then the calculation of the inverse propagation is not carried out. Considering the continuity of network changes, the next study even prior to the propagation calculation are omitted. Jump learning until the error is more than the minimum error, once again to the weight correction.

(2) Dynamic adjustment of learning rate and inertia factor : On the basis the symbol of the t iteration s samples the total error of E (t) and t-1 iteration s samples the total error of E (t-1) changing to decision t times iteration properties, and determine the first t iteration is effective iterative or invalid iteration.Then adopt different rules dynamic adjustment η and α ,then adjust the neural node connection weights and threshold.

The dynamic adjustment of learning factor and inertia factor are as follows:

make E(0)=0.

(1) If $\Delta E(t) = E(t) - E(t-1) \langle 0$, it is indicated that the t times iteration is effectively.

(2) At this time, according to formula (2-1) and type(2-2) adjust η and α :

$$\eta(t+1) = \eta(t) + \lambda\eta(t) \quad \lambda \in (0,1) \tag{2-1}$$

$$\alpha(t+1) = \alpha(t) + \lambda\alpha(t) \quad \lambda \in (0,1) \tag{2-2}$$

At the same time the connection weights to do the following adjustment:

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t+1) \bullet \delta_{pj} \bullet O_{pj} + \alpha(t+1) \bullet (W_{ij}(t) - W_{ij}(t-1)) \tag{2-3}$$

(3) If $E(t) - E(t-1) \geq 0$, indicates that the t times iteration is an invalid iteration. At this time, according to formula (2-4) and type(2-5) adjust η and α :

$$\eta(t+1) = \eta(t) - \lambda\eta(t) \quad \lambda \in (0,1) \tag{2-4}$$

$$\alpha(t+1) = \alpha(t) - \lambda\alpha(t) \quad \lambda \in (0,1) \tag{2-5}$$

At the same time the connection weights to do the following adjustment:

$$W_{ij}(t+1) = W_{ij}(t-1) + \eta(t+1) \bullet \delta_{pj} \bullet O_{pj} + \alpha(t+1) \bullet W_{ij}(t-1) - W_{ij}(t-2)) \tag{2-6}$$

Where δ_{pj} is difference from formula (2-3), and the former takes the results of the t-1 times iteration, but the latter takes the results of t times iteration.

The learning process of BP neural network algorithm is composed of the forward propagation and the back propagation. In the forward propagation process, the input information is processed by the input layer from the input layer to the layer by layer, and the output layer is transmitted to the output layer. The state of the next layer of neurons is affected by the state of each layer of neurons. If the output layer can not get the desired output, then transferred to the reverse propagation, the error signal along the original connection channel to return, by modifying the weights of each layer of neurons, making the error signal is minimal.

2.3 Fusion of decision tree and neural network algorithm

This paper uses decision tree to determine the neural network design of neural network structure, the initial weights and threshold value. It uses the decision tree algorithm to give the initial value of the initial neural network model, the number of hidden units and the initial value of the network parameters, and improve the stability of the neural network training efficiency and results.

In general, the transfer and activation function of neurons using the S function, commonly used as (2-7):

$$f(x)=\frac{1}{1+e^{-\beta x}}, \tag{2-7}$$

$$f(x)=\frac{2}{1+e^{-\alpha_i x}}-1, \tag{2-8}$$

Where x represents the input connection weights and the input values of some kind of mapping results, β parameters called gain .

Property of numeric type, the general form of rule is an IF A>A ∧ … ∧ B ≤ b THEN Class is C, This type of rule is not fully applicable to the traditional decision tree construction method. Nguyen Hong son is proposed for the design of neural network method with continuous attribute data, the first step is to construct a hyperplane, then rule structure of neural network formed by conjunctive hyperplane to classify, its essence is the former layer to super plane layer. In this paper, we make some improvements on the structure of neural network. The former part of the rules generated by the decision tree can be rewritten as (A-a)>0 ∧ … ∧ (B-b) ≤ 0, obviously, (A-a) for a simple hyper plane expression, namely the initial super plane. Then, based on the generated rules, the four layers of the feed forward and the fully connected neural networks are constructed.

If all the former hyperplane of the conjunctive formula set is H, all the neurons of the first hidden layer represents the simple hyperplane, the I neurons using shape (2-8) activation function type.

The initial connection weights of the neurons in the input layer and the first hidden layer are the coefficients corresponding to the hyper plane in the property, therefore not 1 or 0. The initial threshold for Cain neurons correspond to hyperplanes of the simple constant. The second hidden layer neurons rule, each neuron of the layer using the activation function (2-7) function provided by Cain layer of all neurons threshold initialized to 0. If the rule m is represented as (2-9), the first hidden layer and the first connection between the second hidden layer m neurons are set as (2-10):

$$\left(\left(H_i(u)\le 0\right)\wedge\left(H_j(u)>0\right)\right)\Rightarrow D(u)=v_m \tag{2-9}$$

$$w_{nm}=\begin{cases}-1/2 & for \quad n=i\\ 1/2 & for \quad n=j\\ 0 & otherwise\end{cases} \tag{2-10}$$

The neurons in the output layer corresponds to the decision category, the activation function of each neuron in the layer is also used (2-7) provides the function of all the neurons in the output layer 0 threshold, the threshold is not adjusted during the learning process. Second the initial connection weights between the hidden layer and the output layer neuron are 1, which indicates that the corresponding rule output corresponds to the class, otherwise it is 0.

Basic steps are as follows:

(1) According to the transformation rules, the initial number of layers and the number of hidden units of the neural network are determined.;

(2) Set the initial value of each parameter according to certain rules;

(3) Add the required neurons and connections;

(4) Training the neural network.

2.4 Determination of parameters

(1) The connection weights of the input layer and the external input can be initialized by the coefficients of the parameters of each node in the decision tree.

(2) To connect the right of the second layer and the input layer can be initialized using the following rules: in the decision tree for the premise of positive judgment, the connection weights are initialized to W; the premise is a negative judgment is the connection weights are initialized to -w; the threshold for initialization of neurons - (2M-1) w/2 (where M is the number of fixed premises to)

(3) To connect the right initial output layer and the second layer is w; threshold neurons for -w/2 initialization, As shown in Fig. 2.

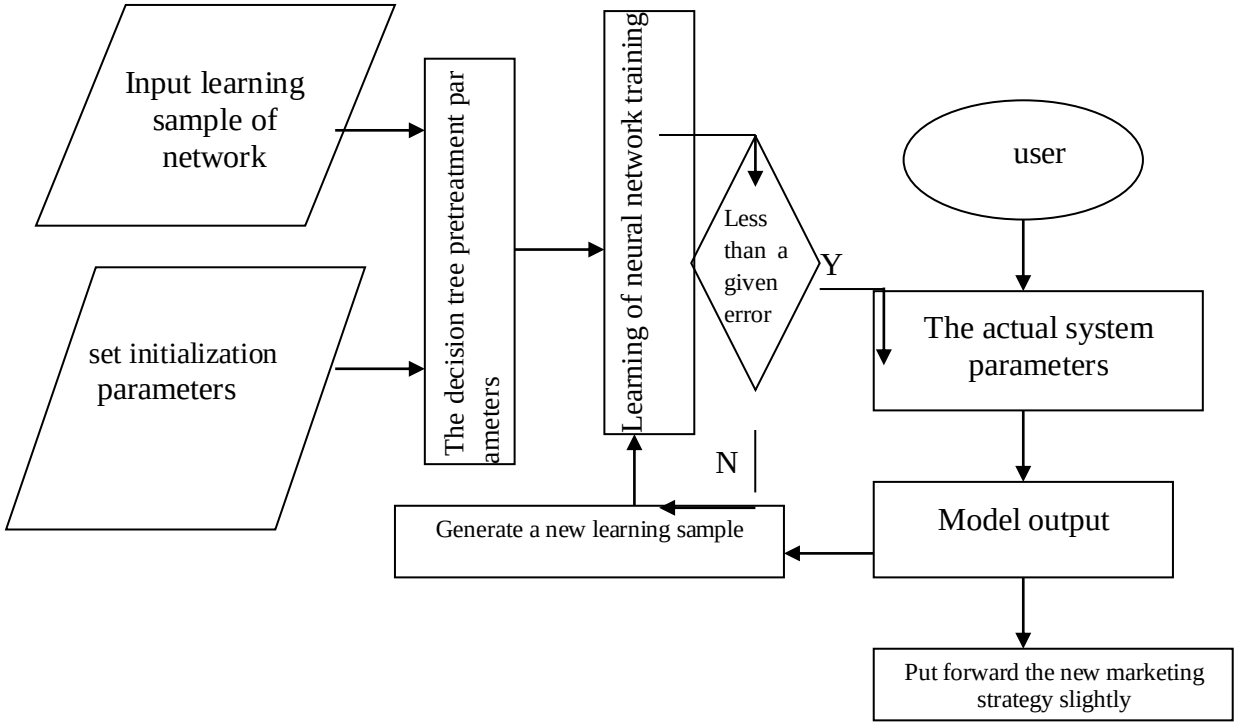

Fig. 2 The fusion of the decision tree and neural network

The essence of decision tree classification is to select the attribute feature which can produce the maximum information gain in the learning process to segment the input field, and the corresponding value forms the decision boundary of each category. The decision boundary in the input space data set (feature space or attribute space) according to the characteristics of the value of division into mutually exclusive decision regions, each region is endowed with a logo, and then integrate the decision region of each category by conjunctive and disjunctive function.

The essence of decision tree classification is to select the attribute feature which can produce the maximum information gain in the learning process to segment the input field, and the corresponding value forms the decision boundary of each category. The decision boundary in the input space data set (feature space or attribute space) according to the characteristics of the value of division into mutually exclusive decision regions, each region is endowed with a logo, and then integrate the decision region of each category by conjunctive and disjunctive function.

In fact, the classification of neural network and decision tree is equivalent to [15], for each neuron of input layer is directly connected with the external input, is the classification hyperplane decision surface input form, and determine the form of each node in the decision tree form classification decision, therefore, the decision of the parameters of the decision surface parameters as neural network formation, classification and decision surface of the neural network is initialized with these parameters is reasonable.

Reference

[1] http://www.empiresoft.net/plex/view.asp?mdocID=7134

[2] Lining, Xie Zhenhua, Shang Lin. Construction method of neural network and decision tree based applications. Journal of Fudan University (NATURAL SCIENCE EDITION). Vol.43 No 5 Oct. 2004
[3] Li Aijun. Neural network based on decision tree. College of computer and information technology, Beijing Jiaotong University
[4] Han Wei. Data mining: concepts and techniques
[5] Pang Shuying et al. The extraction of construction enterprises in the construction industry by using the combination method of decision tree induction and attribute oriented induction. Computer engineering and application. 2003, 208-211 16:
[6] left Du wave. Love people, the application of data mining in customer relationship management system of banking. Journal of Wuhan Polytechnic University twenty-fifth. Volume third. September 2006
[7] Zhang Ying. Application of data mining technology in bank CRM. Guangxi financial research. 2004.2
[8] Shan Siqing, Chen Yin, Yan Cheng. The concept of data mining, models, methods and algorithms. Tsinghua University press