Systems Engineering - Theory & Practice
Volume 27, Issue 12, December 2007
Online English edition of the Chinese language journal

Available online at www.sciencedirect.
ScienceDirect

Cite this article as: SETP, 2007, 27(12): 49–55

RESEARCH PAPER

Customers' Clustering Analysis and Corresponding Marketing Strategies based on Improved SOFM ANN in e-Supply Chain

HUANG Li-juan[1,*], **GAN Xiao-qing**[2]

1. Jiangxi University of Finance and Economics, Nanchang 330013, China
2. School of Business, Jiujiang University, Jiujiang 332005, China

Abstract: According to the similarity of data, the huge number of customers' data in e-supply chain can be clustered objectively and scientifically, and those customers are be clustered into corresponding groups based on SOFM ANN (Self-Organizing Feature Map Artificial Neural Network). Through recognizing and analyzing the different features of these different groups, adopting corresponding marketing strategies can enhance customers' satisfaction, and moreover, can realize the e-supply chain's benefit maximization. In this article, three aspects of improvement are made in the SOFM ANN that applied in customers' clustering analysis; the sample data comes from the Google group. The result shows that the improved SOFM ANN's performance is considerably better than the traditional one's performance. Customers' clustering analysis and corresponding marketing strategies based on the SOFM ANN is a comparatively new topic. Therefore, the result of the research in the article is only for reference.

Key Words: SOFM; e-supply chain; clustering analysis; customer; marketing strategies

1 Introduction

Artificial Neural Network (ANN) is a new artificial intelligence technology. Since famous psychological dissector McCulloch and talent mathematician Pittst[1] have proposed the 1st nerve model in 1943, and has experienced 1980s ANN development golden age, until today, the humanity already had more than 60 years development histories to the ANN research. However, until now, the research and application of ANN is still rising. ANN technology has been widely used in areas such as pattern recognition, Signal Processing, Auto control, Fault Diagnosis, Communications, Electronics, Financial Forecasts, and Knowledge Engineering Management. Now, there are at least 30 kinds of mature ANN models, well-known as BP neural networks, Fukushima neural networks, Self-Organizing Feature Map (SOFM ANN), Hopfield neural network, Boltzmann neural networks, Elman neural networks, and so on. Among them, SOFM ANN is of main concern to people and is used widely.

From the view of neurobiological basis, SOFM ANN can be divided into two category of models. The first category is proposed by Willshaw[2] that the dimension of input is equal to the output ANN model, and the other is from Finland Helsink professor at the University of Kohonem[3] of the high-dimensional input mapping to the low-dimensional output ANN model. As the latter does not lie in the details of neural biology, but rather seizes the substantive characteristics of calculate mapping in the human brain and retains the ease of the calculation, the Kohonen model is more widely applied. If unspecific, the SOFM ANN model usually refers to the Kohonen model. SOFM ANNs are more widely used in data compression and image recognition, but application to customer relationship management, such as clustering analysis and strategic studies in e-supply chain, is a new research field.

In the clustering analysis and strategic studies based on the SOFM ANN, according to the similarity of measured customer data, we can classify the similar characteristics of the customers into the same group using the characteristics of the SOFM ANN. Realizing the identical group's customer difference minimum and the different group's customer difference maximization, is advantageous for adopting the different marketing strategy to the different group's customer and adopting the similar marketing strategy to the identical group's customer. This finally realizes the e- supply chain benefit maximization.

In the virtual Internet world of the e-supply chain, there are magnanimous customer data. The clustering analysis and strategic studies based on the SOFM ANN can effectively reduce the huge management cost with fully personalized marketing, and avoid the enormous risk of losing the 20% potential VIP clients because of taking the same marketing strategy for all customers. Thus, we acquire a Pareto optimal equilibrium between the cost of customer service and the profit of e-supply chain channel. In the traditional classification way for clustering analysis of the e-supply chain, as the limitations of personal experience, we will artificially divide the classification standards in prior, and it is difficult to deal with the huge customer data with limitation of mental, all of which can led to the emergence of errors in classifying the customer groups, resulting in the wrong marketing strategy. The customer data of e-supply chain are not only

Received date: September 26, 2006
* **Corresponding author:** E-mail: huanglijuan66s@126.com
Foundation item: Supported by the National Nature Science Foundation (No.70432001), and the National Postdoctor Foundation (No.20080430179)

large and complex, and are usually non-structural or semi-structured, and the traditional database technology can only handle structured data; thus customer groups cannot be divided scientifically. It is difficult to form a scientific customer marketing strategy using the traditional database technology. In the e-supply chain customers' clustering analysis of this article, the ANN technology, the modern computer technology is used to achieve a fully automated cluster group according to the similarity of customer data. The objective of this study is to make the complex nonlinear of customer data packet processing more objective, scientific, and fair and reasonable than the traditional methods.

2 Fundamental principle of improved SOFM ANN

SOFM ANN is a self-learning self-organizing network without teachers composed by whole connected neurons array, and is the simulation of cerebral cortex function and structure. Physiologists found that neural cortex has various regions, nerve cells with the same function will gather in the same region, and different regions have different and specific functions. For example, the various regions in the brain take charge of the different functions such as language, vision, movement and others. The SOFM ANN precisely simulates the cerebral cortex and makes the topological orderly characteristic region reflection, which to the outside different sense organ information (vision, sense of hearing, sense of taste, sense of touch and so on) inputs, will have the same/similar characteristic data automatically clustered in the identical region.

Three aspect improvements have been made in the article's SOFM ANN relative to the traditional SOFM ANN. First, to access the faster network convergence rate, taking the logarithmic function with Euclid distance Equation to access the generalized d_j . Second, to avoid the disruption and neighborhood invalid mapping, using a spherical topological structure of the neurons, rather than the traditional flat neurons topology; Third, to obtain more precise clustering effect, the PAM clustering barycenter of sample data will be the initial value of weight, and the traditional stochastic value method is not used. These improvements will be reflected in the article. SOFM ANN is a self-organized learning based on competitive learning. Its learning algorithm is described as follows:

2.1 Set initial value

Establish the neighborhood function initial value $S_j(0)$, the learning rate initial value $\eta(0)$, and the weight initial value $w_{ij}(0)$ separately. To speed up the computation speed and the convergence rate of the SOFM ANN, this article takes PAM cluster barycenter value of the sample data as the weight initial value $w_{ij}(0)$ shown as Eq.(1).

$$w_{ij}(0) = \frac{1}{n} \sum_{X \in S(k)} X_{PAM} \tag{1}$$

2.2 Enter new input mode

Random sample X comes from the input space, and the sample data are normal treatment as Eq.(2).

$$X = (x_1, x_2, \cdots, x_n) \tag{2}$$

2.3 Calculate distance d_j

In the traditional method, d_j is calculated by the Euclid distance Equation, and the convergence is slower. The article takes the logarithmic function to access the generalized d_j shown as Eq.(3).

$$d_j = \ln(\|X - w_{ij}\|) = \ln \sqrt{\sum_{i=1}^{n} [x_i(t) - w_{ij}(t)]^2} \tag{3}$$

Where, $i \in \{1, 2, \cdots, n\}$, $j \in \{1, 2, \cdots, m\}$, n is the number of the sub variables of input vector X, and m is the number of output nodes.

2.4 Offer neighborhood function $S_j(t)$

$S_j(t)$ is the set of "closed neuron" at time; it reduces unceasingly along with the time growth. The typical neighborhood function is the Gauss function, shown as Eq.(4).

$$s_j(t) = s_j(0) \exp\left(-\frac{d_j}{2\sigma^2}\right) \tag{4}$$

2.5 Get winning neuron N_k^*

Identify a unit k, in a given surrounding neighborhood $S_j(t)$; for arbitrary unit j, there are:

$$d_k = \arg\min_i(d_j), j \in \{1, 2, \cdots, m\} \tag{5}$$

2.6 Working out learning rate $\eta(t)$

From Eq.(6), we can see clearly that the learning rate is descending with the increase of time. Where, τ is a time constant in SOFM ANN algorithm.

$$\eta(t) = \eta(0) \exp\left(-\frac{t}{\tau}\right), i \in \{1, 2, \cdots, n\} \tag{6}$$

2.7 Modify weight $w_{ij}(t)$

In a given surrounding $S_j(t)$ neighborhood, modify output neuron N_k^* and "abutting neuron" weight according to the following equation:

$$\Delta w_{ij}(t) = \begin{cases} \eta(t)[x_i(t) - w_{ij}(t)], & X \in S(k) \\ 0, & X \notin S(k) \end{cases} \tag{7}$$

and let $t \leftarrow t + 1$; continue to repeat the above learning steps, which start from step 2.2, until the learning rate weakens to the zero or some predetermined positive infinitesimal, namely, until the change of characteristic maps is not obvious[4−5].

3 Concrete application example of customers' clustering based on improved SOFM ANN

3.1 Clustering variables table

The sample data pooled from the Google group annual consuming research report was issued by the commercial consultant firm Marton at the year end of 2005. This article extracts the data of books, which expend in network from the report. The client is the terminal consumer of e supply chain. These expense data can be divided into two kinds: the basic attribute data and the transaction business data. Consumer basic attribute data mainly includes name, gender, age, income, educational level, occupation, living city, marital status, membership time, home address, preferences, etc. The