Expert Systems with Applications 36 (2009) 5475–5480

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The Mahalanobis–Taguchi system – Neural network algorithm for data-mining in dynamic environments

Ching-Lien Huang [a,b,*], Tsung-Shin Hsu [c], Chih-Ming Liu [a]

[a] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, No. 101, Kuang Fu Road, Section 2, Hsinchu, Taiwan, ROC
[b] Department of Industrial Management, Lunghwa University of Science and Technology, No. 300, Section 1, Wanshou Road, Guishan, Taoyuan County 33306, Taiwan, ROC
[c] Department of Industrial Management, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 106, Taiwan, ROC

ARTICLE INFO

Keywords:
Mahalanobis–Taguchi system (MTS)
Data-mining (DM)
Artificial neural network (ANN)

ABSTRACT

Data-mining analysis has two important processes: searching for patterns and model construction. From previous works finding that the Mahalanobis–Taguchi System (MTS) algorithm is successful and effective for data-mining. Conventional research in searching for patterns and modeling in data-mining is typically in a static state. Studies using a dynamic environment for data-mining are scarce. The artificial neural network (ANN) algorithm can solve dynamic condition problems. This study integrates the MTS and ANN algorithm to create the novel (MTS–ANN) algorithm that solves the pattern-recognition problems and can be applied to construct a model for manufacturing inspection in dynamic environments. From the results of the experiment, we find that the methodology of the MTS algorithm can easily solves pattern-recognition problems, and is computationally efficient as well as the ANN algorithm is a simple and efficient procedure for constructing a model of a dynamic system. The MTS–ANN algorithm is good at pattern-recognition and model construction of dynamic systems.

1. Introduction

Large collections of data are potential loads of valuable information. So, in data-mining, search and extraction can be difficult and exhaustive processes (Keim & Kriegel, 1994). In other words, data-mining analysis has two important processes: searching for patterns and model construction. Data-mining is a search for relationships and global patterns that exist in large databases, but are ‘hidden’ in vast amounts of data. Therefore, data-mining is an analytical process that explores data in search of consistent patterns and/or systematic relationships between attributes, and then validates findings by applying the detected patterns to new subsets of a system.

In the search for suitable patterns, in a previous work, A genetic algorithm (GA) and Fuzzy methods based to construct association rules (Kaya & Alhajj, 2005), A linear correlation discovering (LCD) method based for pattern recognition (Chiang, Huang, & Lim, 2005), and Authors used a three-stage online association rule to mine context information and information patterns (Wang, Tseng, & Hong, 2006). Authors adopted a neural network to generate automated trend analysis of proteomics data (Nicholson, 2006). Simultaneously, estimation of works has been discussed in terms of system lifecycle evaluation and estimation attributes (Daskalaki, Kopanas, Goudara, & Avouris, 2003). Additionally, the classification and clustering domains, such as visualization, web data search, position clustering, and graphs classification, have been extensively discussed (Chang & Ding, 2005; Coenen & Leng, 2007; Das & Datta, 2007; Nasraoui, Rojas, & Cardona, 2006).

In the field of model construction, a GA algorithm based to model a bankruptcy prediction model (Kim & Han, 2003), an artificial neural network based to predict subsidence (Ambrozic & Turk, 2003), and the use of biblio-mining frameworks to generate a usage-based forecasting rule (Nicholson, 2006).

Obviously, when searching for patterns, the MTS is a good and an effective algorithm (Das & Datta, 2007). The MTS developed by Taguchi is a novel method that combines the Mahalanobis distance (*MD*), orthogonal arrays (OA) and the signal-to-noise (SN) ratio. The MTS is a diagnostic and forecasting method. The main aim of the MTS is to make accurate predictions in multidimensional attributes by constructing a global measure meter. The application of MTS in the pattern-recognition area is as follows: a MTS based to resolve classification problems (Das & Datta, 2007).

Literature review findings demonstrate that the MTS algorithm is successful and effective for data-mining. Conventional research

* Corresponding author. Address: Department of Industrial Management, Lunghwa University of Science and Technology, No. 300, Section 1, Wanshou Road, Guishan, Taoyuan County 33306, Taiwan, ROC. Tel.: +886 939800339; fax: +886 2 82093211x6110.
E-mail address: lynne.line@msa.hinet.net (C.-L. Huang).

0957-4174/$ - see front matter Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.06.120

in searching for patterns and modeling in data-mining is typically in a static state. From above mentions, studies using a dynamic environment for data-mining are scarce. The artificial neural network (ANN) algorithm to solve dynamic and multi-response condition problems. This study integrates the MTS and ANN algorithm as a novel method for dynamic environments, that is, this work adopts the Backpropagation neural network (BPN) owing to its ability to map the complex relationship between input data and corresponding outputs (Keim & Kriegel, 1994).

In summary, the MTS–ANN algorithm is utilized to construct a pattern and model a dynamic system. A case study of an electronics company is used to verify and validate the proposed methodology.

2. Model construction

Model construction using the MTS–ANN algorithm begins with the MTS process. The selected data are processed using the following steps.

First, the process of the proposed algorithm in terms of establishing the MTS, which computes the *mean*, *standard deviation*, and *SN* ratios from a normal dataset. Next, compute the same values as an abnormal dataset. From the two groups of dataset which classifies is applied verify the system. Next, calculate the Mahalanobis distance (*MD*), transfers the *MD* values into observations y_{ij}, the *threshold* value is used to determined whether attributes are important and, the *SN* ratios are obtained to selects these attributes that the model reduced to a pattern.

Furthermore, the ANN algorithm is applied during the second stage. The purpose of ANN is to construct a dynamic system, and form a system model, which is completed using the next two steps. The first step determines to map a model. The second step tests this model. That is, the formula $Y_{ij} = F(V_i)$ $i = 1 - k$, $j = 1 - n$ is generated and applied to a dynamic complex system. In summary, the algorithmic procedure has the following three steps:

Step 1: Construct a measure meter with the MTS process of a system

First, parameters – *mean*, *standardized values*, *relationship matrix* and *MD* – are computed. The formula for standardized values is Eq. (1): The *mean* and *standard deviation* are $\overline{Y_i} = \sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}/n$, and $S_{td} = \sqrt{\frac{\sum y_{nk} - m_k}{n}}$, and the number of k is expressed the attributes of a dataset

$$y_{ij} = \frac{Y_{ij} - m_i}{\sigma_i}. \tag{1}$$

Next, Eq. (2) generates the relationship matrix:

$$R = \begin{bmatrix} 1 & r_{12} & \cdot & r_{1k} \\ r_{21} & 1 & \cdot & r_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ r_{k1} & r_{k2} & \cdot & r_{kk} \end{bmatrix}. \tag{2}$$

Matrix A is the inverse of covariance for the attributes of the normal dataset, and is derived by Eq. (3). Finally, the MD is derived by Eq. (4), where y_{ij} is the standardized values of Y_{ij}, the formula is $y_{ij} = \frac{Y_{ij} - \overline{y}_i}{\sigma_i}$, $i = 1 - k$, and $j = 1 - n$, and Y_{ij} represents the value of the observational value of the *ij*th attribute

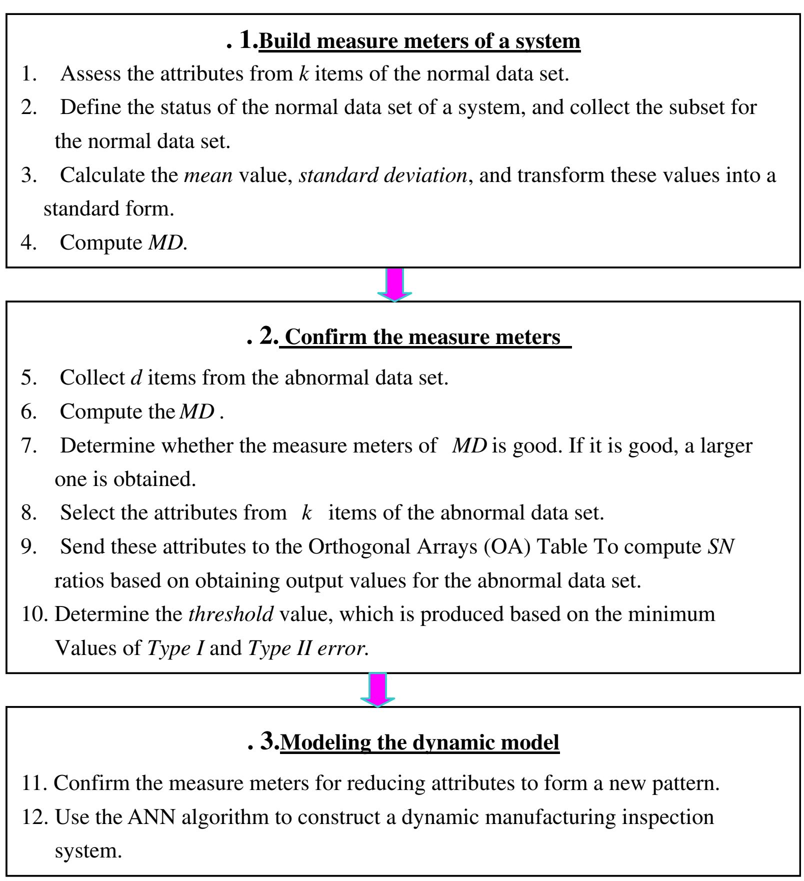

Fig. 1. Summary of the MTS–ANN algorithm.