Optimization of Data Mining in Dynamic Environments based on a Component Search Neural Network Algorithm

Ching-Lien Huang, Yung Hui Chen, Tian-Long John Wan

Department of Industrial Management, Lunghwa University of Science and Technology, lynne.line@msa.hinet.net
Department of Computer Information and Network Engineering, Lunghwa University of Science and Technology, cyh@mail.lhu.edu.tw
Department of Information Management, Lunghwa University of Science and Technology, Johnwan@mail.lhu.edu.tw
*Corresponding author e-mail: cyh@mail.lhu.edu.tw

Abstract

The searching for patterns and model construction in Data-mining is the important contribution. Previous works report that the component search (CS) algorithm is successful and effective for data-mining. Conventional research in pattern search and modeling in data-mining only consider data in a static state. Studies data-mining in dynamic environments are scarce. The artificial neural network (ANN) algorithm can solve dynamic condition problems. This study integrates the CS and ANN algorithm to create the novel (CS-ANN) algorithm that solves pattern-recognition problems and can simply construct manufacturing inspection models for dynamic environments. Simply experimental results indicate that the CS algorithm is computationally efficient for solving pattern-recognition problems and that the ANN algorithm enables simple and efficient modeling of dynamic systems. The CS-ANN algorithm is effective for pattern-recognition and modeling of dynamic systems.

Keywords: Component search (CS), Data-Mining (DM), Artificial Neural Network (ANN)

1. Introduction

Data-mining is simply a process of analyzing data for consistent patterns and/or systematic relationships between attributes and then validating findings by applying the detected patterns to new subsets of a system [1]. Notable research in pattern search includes a genetic algorithm (GA) and fuzzy methods used to construct association rules [2]. A Linear Correlation Discovering (LCD) method of pattern recognition [3], and Component value selection for active filters using parallel tabu search algorithm to generate new inspection patterns [4]. Combining principal component analysis, genetic algorithm and tabu search in 3D molecular similarity to get new patterns has also been reported [5]. Exarchos, et al. [6] used a three-stage online association rule to mine context information and information patterns. [7] adopted a neural network to generate automated trend analysis of proteomics data. Simultaneously, Daskalaki et al. [4] reviewed works in system lifecycle evaluation and estimation attributes. Classification and clustering domains, such as visualization, web data search, position clustering, and graphs classification, have also been discussed [3, 7, 8, 9, 10, 11]. Model construction include GA algorithm used to construct a bankruptcy prediction model [12], an artificial neural network to predict subsidence [13], and biblio-mining frameworks used to generate a usage-based forecasting rule [7]. And, in dynamic system model construction include GA algorithm, an artificial neural network, and Fuzzy set, etc, which used to construct the dynamic system model [1, 11, 14, 15, 16, 17, 18, 19, 20].

Simply the CS algorithm is clearly effective for pattern search [21]. The CS developed by Motorola is a novel method of finding key components that represent a product. The CS is a diagnostic and forecasting method simply designed to accurately predict numerous attributes by choosing the representative component. [21] used a CS based to resolve pattern recognition problems. Simply the literature confirms that the CS algorithm is effective for data-mining. Conventional research in pattern search 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 can
solve dynamic and multi-response condition problems. This study integrates the CS and ANN algorithm to obtain a novel method for dynamic environments restated, this work adopts the Back-propagation neural network (BPN). Simply because it can map the complex relationship between input data and corresponding outputs [22].

In summary, the CS-ANN algorithm is utilized for pattern construction and modeling in dynamic systems. A case study of an electronics company demonstrates the effectiveness of the proposed method.

2. Model construction

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

First, identify the key functions of the assemblage. Second, differentiate the good and bad assemblages. Third, measure the significant differences. Fourth, evaluate assemblage performance by exchanging components to identify effective factors. Finally, perform a confirmation experiment to test the model.

Further, the ANN algorithm is applied during the second stage. The purpose of ANN is to construct a dynamic system and form a system model by performing two steps. The first step is model mapping. The second step tests this model. That is, the formula \( Y_i = F(Y_j) \) \( i = 1 \) to \( k \), \( j = 1 \) to \( n \) is generated and applied to a dynamic complex system. In summary, the algorithmic procedure consists of the following three steps, as shown by Figure 1.

Step1. Determine the measurement index

This stage identifies the key components for which parameters \( d \) and \( D \) were generated and evaluated. The following five sub-step procedures are performed.

(1) Determine the experimental goal
   Simply identify the key functions of the system/assemblage.

(2) Prepare the good and bad difference system/assemblage
   A good assemblage can reach its measure value, but a bad assemblage can only obtain the value.

(3) Identify significant differences between good and bad system
   A good system/assemblage carries on the measure value with the bad system/assemblage to obtain \( G_1 \) and \( B_1 \), which again obtains system/assemblage rewiring and gets \( G_2 \) and \( B_2 \), and compute \( d \) and \( D \). If \( d < 5d \), the significant difference is tenable, and the formula of \( d \) and \( D \) is as below.

   \[
   d = AVG[G_1 - G_2] + |B_1 - B_2| \tag{1}
   
   D = AVG[G_1 + G_2] - |B_1 - B_2| \tag{2}
   
   The assemblage reveals the performance difference of all components.

(4) Evaluate the system/assemblage performance by exchanging one of the components.
   Simply this stage measures system/assemblage performance by exchanging a good component for and bad one.

(5) Identify effective factors
   The exits obvious differences among these components of the system/assemblage, and find the important components. The process of this stage requires the following five sub-steps.

   (1) Determine the research goal.
   (2) Prepare the good / bad system/assemblage.
   (3) Measure the significance difference.
(4) Evaluate the system/assemblage performance by exchanging one of the components.
(5) Identify the effective factors.

**Step2. Simply assess system/assemblage performance**

Simply this stage confirms the key components selected in the above step. Simply and identifies differences between components.

Here, this stage is sum in following one sub-step:

(1) Do the confirm experiment.

**Step3. Construct the dynamic model**

The third step includes two details for the ANN algorithm. The third step includes two processing steps of the ANN algorithm. The first step is model mapping. The second step tests this model and its model function formula as follows.

\[ Y_j = F(v_i) \]  

Where \( i = 1 \text{ to } k \), \( j = 1 \text{ to } n \), \( k \) is attribute items of a data set, \( n \) is experiment times. Thus, this stage is sum in following sub-steps:

(1) Map & Test the Model.

![Figure 1. The flowchart of CS-ANN Algorithm](image-url)
3. Case application

Assume an electric company has five components that require inspection. These components are coded from A, B, C, D, and E. Eleven datasets are selected from the good data set, and nine datasets are extracted from the bad data set. The CS-ANN algorithm is applied as the following:

**Step1. Determine the measurement index**

First, \( d \) and \( D \) are calculated for the system/assemblage. Table 1 shows the \( d/D \) value of the system/assemblage.

<table>
<thead>
<tr>
<th>No</th>
<th>Good Result</th>
<th>Bad Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( A_g ) and ( R_g )</td>
<td>( A_g ) and ( R_b )</td>
</tr>
<tr>
<td>2</td>
<td>( B_b ) and ( R_g )</td>
<td>( B_g ) and ( R_b )</td>
</tr>
<tr>
<td>3</td>
<td>( C_b ) and ( R_g )</td>
<td>( C_g ) and ( R_b )</td>
</tr>
<tr>
<td>4</td>
<td>( D_b ) and ( R_g )</td>
<td>( D_g ) and ( R_b )</td>
</tr>
<tr>
<td>5</td>
<td>( E_b ) and ( R_g )</td>
<td>( E_g ) and ( R_b )</td>
</tr>
</tbody>
</table>

**Step2. Simply assess system/assemblage performance**

For this stage, Table 1 shows the obvious differences among these components. To find important components and to perform above analysis, and the key components which will be found out, and they are A and D. And, to examine the difference among these components, which shows in Table 2.

<table>
<thead>
<tr>
<th>Test</th>
<th>Good Result</th>
<th>Bad Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>( A_g D_g R_g )</td>
<td>( A_g D_b R_b )</td>
</tr>
<tr>
<td>1</td>
<td>( A_g D_g R_g )</td>
<td>( A_g D_b R_b )</td>
</tr>
<tr>
<td>2</td>
<td>( B_b A_g D_g R_g )</td>
<td>( B_g A_b D_g R_b )</td>
</tr>
<tr>
<td>3</td>
<td>( C_b A_g D_g R_g )</td>
<td>( C_g A_b D_g R_b )</td>
</tr>
<tr>
<td>4</td>
<td>( A_g D_g R_g )</td>
<td>( A_g D_b R_b )</td>
</tr>
<tr>
<td>5</td>
<td>( E_g A_g D_g R_g )</td>
<td>( E_g A_b D_g R_b )</td>
</tr>
</tbody>
</table>

**Step3. Construct the dynamic model**

(1) **Map the model with BPN.**

In this stage, the ANN algorithm is used to develop a dynamic model. Thus, as Table 3 shows, the second system/assemblage was selected and its \( d/D \) was computed, and was shown in Table 3.

The formula for the main effect of A as follows.
The formula for the main effect of D as follows.
\[ D = AVG(A_1, A_3) - AVG(A_2, A_4) \]  
(5)

The formula for the interaction of AD is follows.
\[ AD = AVG(A_2, A_3) - AVG(A_1, A_4) \]  
(6)

From above analysis, the A and D are the most important component of the system.

<table>
<thead>
<tr>
<th>Table 3. Cross Effects of the Important Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.55</td>
</tr>
<tr>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. The Input Parameters and Their Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>

(2) Construct the dynamic system

In this stage, the ANN algorithm is going to run the modeling of a dynamic model if a dynamic system is in a good condition or not. The process sub-steps are bellow.

(i) Map the Model

And then the ANN algorithm is used to develop a dynamic complex model. Thus, the second group of its input and output data will be selected, a BP model is also constructed to map the model. The initial input/output parameters, and the results of mapping, which lists in Eq. 3, Table 4, and Table 5.
(ii) Test the Model.

The relationship and responses model is developed by using BP neural network, in which inspection data of 27 for training as well as 5 lots are used for testing. The model structure is selected with 6-4-6 (input-hidden-output) owing to the RMSE is the lowest value among others’ which has analysis by the experiment but never be involved into this work, in which the architecture 6-4-6 is chosen to get a convergence performance.

The RMSE of training error is 0.01, and the testing error is 0.01. Both the two RMSE values of training and testing are all going into convergence, which shown in Figure 2.

![Figure 2. The RMSE Values of Testing Structure](image)

Therefore, according to the data for the second group, the formula of desirable functions is used to applied for a dynamic system can map in 6-4-6 ANN structure successfully.

4. Discussion

The discussing of the CS algorithm is also confirmed by changing the number of experiment. Thus, if the number of experiments is test by sequential and random. And then in testing, this study will be determined that the reliability of the CS algorithm is the same. Therefore, this study determined that the above two changing states would shows that the correct rate not be changed. Restated, the new pattern chosen has the following variables: $v_1$, $v_2$, $v_3$, $v_4$, and $v_5$. And the results are compare with the MTS method of [20], its accuracy is shown in Figure 3.
5. Conclusion

In above analysis, forming the new pattern from second group, which show the number value of parameters are changed from 30 to 6. From above mentions, we know that the proposed algorithm is successful applied in pattern forming.

In modeling a dynamic system aspect, the ANN algorithm shows us the 6-4-6 structure is the optimal selection from other’s architecture, which shows RMSE is convergence in 0.01.

From the results, the methodology of the CS 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 CS-ANN algorithm is good at pattern-recognition and model construction of dynamic systems.

We conclude that the CS-ANN algorithm can successfully be applied to dynamic environments for data-mining problems.

6. Acknowledgment

This work was supported by the National Science Council, R.O.C., under Grant NSC 100-2511-S-262-002-.

7. References