Unsupervised Neural Network Adaptive Resonance Theory 2 for Clustering

Aina Musdholifah\textsuperscript{1} and Difla Yustisia

Department of Computer Science and Electronics
Gadjah Mada University (UGM)
Yogyakarta, Indonesia
non.nana@gmail.com\textsuperscript{1}

\textbf{Abstract}—Clustering is one of pattern recognition techniques which are often used for extract information from large amount dataset to get more benefit for the data owner. Clustering which is an unsupervised technique assigns the input data into clusters based on their similarity degrees. In this paper, the Unsupervised Neural Network Adaptive Resonance Theory 2 is used for clustering data. To evaluate the results, a technique called Cohesion and Separation is utilized. Furthermore, in order to validate the algorithm, this study uses patent data. The proposed ART2 algorithms and validation techniques scale well and gain considerable performance due to the resulted cluster.

\textbf{Keywords:} Clustering, Unsupervised Neural Network, Adaptive Resonance Theory 2.

I. INTRODUCTION

Pattern recognition is one of techniques that place objects to classes or groups. The pattern recognition is called supervised pattern recognition when there are training data with labels in each formed classes. But, there are many cases with no training data. For that condition, the main objectives of the pattern recognition is how to calculate the similarities among objects and then group the similar objects to same cluster. This process is called unsupervised pattern recognition or clustering.

In clustering, there are no predefined classes and no examples. Objects are grouped based on self-similarity. From this practice perspectives, clustering play a important part in data mining applications, such as scientific data explorations, extracting information, mining, spatial databases applications, Web analysis, CRM, marketing, medical diagnosis, and biology computations. Many clustering algorithms have been developed, such as \textit{K-means} [1], \textit{density-based partitioning}, \textit{grid based}, and \textit{unsupervised neural network} [2].

Unsupervised neural networks (UNN) have ability to cluster multivariate data. The number of clusters is decided by UNN and directly depends on data inputs (Clare dan Cohen, 2001). SOM (Self-Organizing Maps) is an UNN technique that commonly used. Another technique, Adaptive Resonance Theory (ART) has some types, such as Adaptive Resonance Theory 1 (ART 1) that is designed to binary inputs, and Adaptive Resonance Theory 2 (ART 2) that is designed to continuous (numeric) values [3].

In this paper, ART 2 is chosen, because this algorithm can accept input in numeric form that is usually used many data. Changing of data symbols to numeric form is also simpler than to another form, such as binary form. Besides that, Fausett [3] states that ART was designed to stable networks.

II. PREPROCESSING DATA

Preprocessing data is predecessor step before clustering. This step is needed to form or arrange input in order that can be used to next steps [4]. In preprocessing, it is needed to convert the data after get the data, because the data may be still has various formats or types, such as texts or symbolic data. Otherwise, in many neural networks, the input should be in certain type, such as binary type or continuous type.

The steps of converting symbolic or text data to numeric data are (1) an integer code is assigned for each symbol; (2) and then, change them to binary or continuous data. To normalizing the continuous data, we need to simple analyzing the data, i.e., calculate the number of variation for each attributes, and then define the maximal values for each feature. After that, by using the maximum values and simple equation 1, normalize the values of each features.

\[
N_f = \begin{cases} 
1, & \text{if } f > \text{Max}f \\
\frac{f}{\text{Max}f}, & \text{if } f \leq \text{Max}f
\end{cases}
\]

(1)

Where \(F \), \(f \), \(\text{Max}f\) and \(N_f\) is feature / field, features values, Maximum value for \(F\) that can be received and normalized values of \(F\), respectively.

III. UNSUPERVISED NEURAL NETWORK (UNN)

Artificial Neural Networks is a system that process information like human neuron characteristics. Artificial neural network was developed as a mathematic model generalization of neural biology based on assumption:

1. Information processing is done to many simple elements, that is called neuron.
2. Signal is send between the neurons by connection link.
3. Each the connection link has suitable weight that will strengthen or weaken the signal.
4. Each neuron use an activation function (usually it is non linear) that is operated to sum of the inputs that received to deciding their output signals. And then the output signal will be compared to a limit threshold.

An artificial neural network is decided by three terms [3]:
1. Connection patterns between the neurons, it is called network architecture,
2. Methods to decide the connection’s weight, it is called learning algorithms,
3. Activation function.

The method to decide the connections’ weight is one of special characteristic from various neural networks. Generally, such learning divide two types, i.e., supervised and unsupervised [3]. Typical neural network with supervised learning is often used. This network learns used many vectors or patterns, and each vectors is connected to suitable output vectors. This process is often known as supervised learning/ training [3].

Whereas the neural network type with unsupervised learning is a self-organizing network. This network can group input vectors that have similarities without using training data to decide what kinds of the members in one group, or which output vectors that an input should joins.

The neural network then modified weights so that the most similar inputs will be in same output unit (cluster). This neural network also produce example vector for each clusters that produced. The networks that can be categorized as unsupervised neural network are Kohonen Network (or it is often called Self-Organizing Map (SOM), and Adaptive Resonance Theory [3].

IV. ADAPTIVE RESONANCE THEORY (ART-2)

One of type of ART is ART 1, which is special designed for clustering of binary vectors. Whereas another types, is ART 2, received vector input by continuous values. The network is clustering input based on unsupervised learning. When each a pattern (in vector format) is represented, a suitable cluster will be chosen and its weight will be adapted until the cluster unit can earn pattern that was inputted before. The weight of a cluster unit is represented in vector code form or exemplar form for patterns in that cluster.

F_1 layer consist of 6 unit type (i.e.: W, X, U, V, P, and Q). There are n units for each the unit types, where n is dimension of an input pattern. There is only one extra unit for each unit types, which is shown in figure 1. An extra unit between W unit and X unit receives signals from all of W units and sends the signals to all of X units. Another extra unit between P unit and Q unit has same task. As is an extra unit before, an extra unit between unit V and U unit has same task. Both X unit and Q unit connect to V unit.

The symbols of connection path between the units in F_1 layers in figure 1, indicate signal transformation from a type unit to next type unit. But, connection between P unit in F_2 layer and Y in F_3 layer represent weight that multiply transmitted signals by the path. F_3 unit activation that win, is d, where 0<d<1. The symbol ---> indicate normalization (Fausett, 1994).

The learning algorithms use some facts, i.e.:
1. Reset cannot happen during resonance (step 8)
2. A new winner unit cannot be chosen while resonance.

But, this algorithm did not use yet facts that:
1. Usually, in slow learning use N_IT = 1, and step 10th can be passed away.
2. In fast learning, for first pattern that be learned by a cluster, u will parallel with t by learning cycles and equilibrium weight will be:

\[ t_{j} = \frac{1}{1-d} u_i \]
\[ b_{j} = \frac{1}{1-d} u_i \]

(4)

Many another impossible stop conditions are:
1. Repeat step 8 until weight change under certain tolerance. For slow learning, repeat step 1 until limit of certain tolerance. Whereas for fast learning, repeat step 1 until placing the patterns of cluster unit do not change from one epoch to next epoch.
2. Step 3 until 11 form a learning trial that is a representation of a pattern). It is better to reference performance of a learning trial for each input pattern as a pattern. The pattern is not needed to be present with same order for each epoch. Update activation F1 during resonance phase in phase learning mode cause value u, become different during learning is going on (Fausett, 1994).

![Figure 1. Typical architecture of ART2](image)
V. CLUSTER VALIDATION

One of type of ART is ART 1, which is special designed for clustering of binary vectors. Whereas another type, is ART 2, received vector input by continuous values. The network is clustering input based on unsupervised learning. When each a pattern (in vector format) is presented, a suitable cluster will be chosen and its weight will be adapted until the cluster unit can earn pattern that was inputted before. The weight of a cluster unit is represented in vector code form or exemplar form for patterns in that cluster.

The main clustering problem is how to decide the number of optimal cluster that suitable to the data set. Sometimes the clusters results are not represented to the real data structure. Hence, the quantitative measures are needed to evaluate the result of clustering algorithms need, next we called cluster validation [5].

The cluster validation technique can be divided three types, i.e., unsupervised, supervised, and relative. Unsupervised cluster validation technique measures as well as clustering structures without see external information aspects. One of the techniques is Sum Squared Error (SSE). Unsupervised measurements usually are divided two classes, i.e., cluster cohesion that decide how close the objects in same cluster, and cluster separation that decide how separate among the clusters [6].

A. Sum Square Error (SSE)

SSE calculates error of each point of data, then compute total of error square. Errors in this context can be a Euclidean distance to closest centroid. Formally, SSE can be defined:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} \text{dist}(c_i, x)^2$$

(5)

which k is the number of clusters, x is an object in cluster, c_i is cluster-i, c_i is c_i's centroid, and dist is Euclidean distance among two objects [6].

B. Cohesion and Separation

Cluster validation of a set with K cluster, can be defined as a sum of weighted individual cluster validation,

$$\text{overall validity} = \sum_{i=1}^{K} w_i \text{ validity}(C_i)$$

(6)

where w_i is weight of cluster-i, that depend of the cluster validation measurement technique. Validation functions can be cohesion, separation, or combination of cohesion and separation [6].

Cohesion of the cluster can be defined as the sum of closeness of the cluster to the centroid. Separation between two clusters can be measured by the closeness of the cluster prototype and the centroid. The illustration can be shown in figure 3 which the cluster centroid is assigned by "*".
Learning algorithm of the artificial neural network used learning algorithm ART 2. This experiment used slow learning and fast learning.

<table>
<thead>
<tr>
<th>Types</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster measurements</td>
<td>Cluster weights</td>
</tr>
<tr>
<td>( \sum_{C_i} \text{proximity}(x,c_i) )</td>
<td>( m \times \text{objek dalam cluster ke-i} )</td>
</tr>
<tr>
<td>( \text{proximity}(c_i,c) )</td>
<td>( \text{Prototype-based Cohesion} )</td>
</tr>
<tr>
<td>( \text{Prototype-based Separation} )</td>
<td></td>
</tr>
</tbody>
</table>

B. Clustering

In this process, there are two steps of learning used UNN ART 2 and cluster validation. Clustering process is done many times with Max_m and Min_m parameters. It is done many times because to get the best clustering result, that has optimal SSE total and SSB total.

VII. ANALYSIS OF EXPERIMENTAL RESULT

This study utilizes the daily medical record of Local Government Clinic. The data has some attributes, but in this experiment only use five attributes including Dates, Address, Age, Sex, and medical diagnosis.

From the experiment results, has shown that if the clusters that have smallest total SSE then they have biggest total SSB too. Besides that, all of the experiments have shown that the best clustering begins in certain m values until upper limit of the range, and also in the same clustering results.

VIII. CONCLUSION

Unsupervised neural network ART 2 (ART-2) algorithm has been implemented. From the experimental result, the algorithm achieved clusters with smallest total SSE and biggest total SSE. The proposed ART2
algorithms and validation techniques scale well and gain considerable performance due to the resulted cluster.

![Flowchart of clustering process.](image)

**Table II. The Experiments and Attributes Combinations**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Dimension</th>
<th>The Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Date Address Age Sex Diagnosis</td>
</tr>
</tbody>
</table>

**Table III. The Result of The Experiments**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Attributes</th>
<th>Ranges</th>
<th>Best Result (Based on Total SSE and Total SSB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>m with smallest SSE</td>
</tr>
<tr>
<td>1</td>
<td>Address, Diagnose</td>
<td>10-20</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>Date, Address, Diagnose</td>
<td>10-20</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Date, Address, Age, Diagnose</td>
<td>10-20</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Date, Address, Age, Sex, Diagnose</td>
<td>10-20</td>
<td>15</td>
</tr>
</tbody>
</table>

**Acknowledgment (Heading 5)**

This work is supported by a research grant from Ministry of National Education (Mendikbud). The authors gratefully acknowledge many helpful comments by reviewers in improving the publication.

**References**