A Novel Approach for Mixed Data Clustering using Dynamic Growing Hierarchical Self-Organizing Map and Extended Attribute-Oriented Induction

Hari Prasad¹, M. Punithavalli²

¹Assistant Professor (Selection Grade), Department of Computer Applications, Sri Ramakrishna Institute of Technology, Coimbatore, India.
²Director, Department of Computer Applications, Sri Ramakrishna Engineering College, Coimbatore, India.
E-mail: hari.research@yahoo.com

Abstract: Data clustering is one of the most important data mining techniques which groups data supported on their similarity. A number of approaches are existing for clustering numerical data and the problem of clustering mixed data is still unresolved. The standard clustering techniques are in general used for numeric data and are not probable to handle mixed data for the reason that of their computational incompetence. The requisite for an enhanced mixed data clustering approach is becoming vital and it is turning out to be a hot research area. By the sort of resolving this issue, Growing Hierarchical Self-Organizing Map (GHSOM) and Extended Attribute-Oriented Induction (EAOI) for clustering mixed data type is previously projected except it does not have any capability to control the growth of the map and in addition the structure of GHSOM is static. To overcoming this issue, a Dynamic Growing Hierarchical Self-Organizing Map (DGHSOM) with EAOI is projected in this paper for handling the mixed data. The main importance of DGHSOM is that it has the ability to grow or modify the structure to represent the application enhanced. The experimentation for the proposed technique is approved with the help of UCI Adult Data Set and Cleve Dataset and it is fond that it is superior to previous approaches based on the number of resultant clusters and outliers with substantial reduction in the processing time. The Clustering error also reduced.

Keywords: Mixed Data Clustering, Extended Attribute-Oriented Induction (EAOI), Self-Organizing Map, Dynamic Growing Hierarchical Self-Organizing Map (DGHSOM), Controlled Growth.

1. Introduction

Due to the major development in together computer hardware and software, a enormous quantity of data is produced and gathered regularly. It is to be acknowledged that data are meaningful only at what time one can extract the hidden information inside them. On the other hand, “the most substantial complexity for obtaining high quality knowledge from data is due to the insufficiency of the data itself” (Wiederhold et al.,1996). These major complexities of gathered data move toward from their increasing size and adaptable domains.

Clustering is one of the important tools in data mining. The main objective of data clustering is focused on segmenting the data set into quite a few different groups in order that objects contains a high degree of resemblance to each other in the related group and have a high degree of difference to the ones in different groups (Han et al.,2001). Every generated group is acknowledged as a cluster. Useful patterns may be acquired by analyzing each cluster. In case, clustering customers with similar characteristics depending on their purchasing behaviors in transaction data may find out their previously unidentified patterns. The acquired information is useful to take decisions in the field of marketing.

Majority of the existing clustering techniques is proficient of processing moreover categorical data or numeric data. On the other hand, in recent times numerous mixed datasets together with categorical and numeric values came into existence. One of the well-known practices is to group the mixed dataset to convert categorical values into numeric values and after that they carry out any numeric clustering algorithm. A further common approach is to compare the categorical values directly, wherever two discrete values result in distance 1 on the same time like identical values result in distance 0. Though, these two techniques do not consider the similarity information embedded among categorical values. As a result, the similarity structure in grouping outcomes is not showing clearly in the dataset (Hsu C., 2005, 2006).

One of the familiar neural networks for mixed data clustering is Kohonen’s Self-Organizing Map (KSOM) developed primarily for visualization of nonlinear associations of multi-dimensional data (Kohonen,1995). The KSOM is famous for its helpfulness in numerous real world applications (Haritopoulos et al., 2002, Kohonen et al., 2000, Papadimitriou et al., 2001).

The Self-Organizing Map (SOM) has been exploited as a tool for mapping high-dimensional data into a two or three dimensional feature map.
(Kohonen, 1995). It is then sufficient to visually recognize the clusters from the map. The main benefit is that it would be probable to accomplish some idea of the structure of the data through examining the map, owing to the topology preserving nature of the SOM. It has been uncertainly exposed that the SOM in its original structure does not capable to handle mixed data and numerous attempts have been made by several researchers to overcome this limitation (Villmann et al., 1997, Ritter et al., 1992).

Growing Hierarchical Self-Organizing Map (GHSOM) and Extended Attribute-Oriented Induction (EAOI) are used for managing the mixed numeric and categorical data. However it does not have any potential to control the growth of the map and additionally the structure of GHSOM is static. In order to solve these issues, Dynamic Growing Hierarchical Self-Organizing Map (DGHSOM) is planned in this paper.

2. Related Works

The different existing clustering techniques are discussed in this section which is proposed by different authors.

Dharmendra et al., (Roy et al., 2010) proposed a Genetic K-Means Clustering Algorithm for Mixed Numeric and Categorical Data Sets. A novel method (Juha Vesanto et al., 2000) was set onwards by Juha et al., for clustering of Self-Organizing Map. According to the way proposed in this paper the clustering is carried out by means of a two-level approach, where the data set is initially clustered with the SOM, and then, the SOM is clustered.

Mark Girolami presents a Mercer Kernel-Based Clustering (Mark Girolami, 2002) algorithm in Feature Space. This paper presents a method for both the unsubstantiated partitioning of a sample of data and the assessment of the feasible number of inbuilt of clusters which generate the data.

A new supervised clustering algorithm was predictable by Li et al., in (Shijin Li et al., 2007). They recommended their algorithm for data set with mixed attributes. As of the complexity of data set with mixed attributes, the conventional clustering algorithms suitable for this kind of dataset are not many and the result of clustering is poor. K-prototype clustering is one of the most frequently used methods in data mining for this kind of data. They borrowed the thoughts from the multiple classifiers combing technology which uses k-prototype as the basis clustering algorithm in order to design a multi-level clustering assemble algorithm, which adaptively selects attribute for re-clustering. Jian et al., in (Jian Yin et al., 2005) proposed an proficient algorithm for clustering mixed type attributes in huge dataset.

Incremental Grid Growing (IGG) was developed by Blackmore (Blackmore J., 1995) which construct the network incrementally by vigorously modify its structure and connectivity supported on the input data.

A robust and scalable clustering algorithm was set onwards by Tom et al., in (Tom Chiu et al., 2001). The author working this clustering algorithm for mixed type attributes in large database environment. In their paper, they projected a distance measure that make possible grouping data with both continuous and categorical attributes. This distance measure is resultant from a probabilistic model that the distance between two clusters is equivalent to the decrease in log-likelihood function as a result of integration. A sum of this measure is memory efficient as it depends only on the integrating function as a result of integration. A sum of this measure is memory efficient as it depends only on the integrating function as a result of integration.

3. Methodology

3.1 Growing Hierarchical Self-Organizing Map (GHSOM)

The Growing Hierarchical Self-Organizing Map consists of a hierarchical structure of multiple layers in which each layer contains numerous independent growing Self-Organizing Maps (Rauber et al., 2002). Starting from a top-level map, every map which is related to the Growing Grid model, grows in size to symbolize a assembling of data at a particular level of detail. After a distinct development concerning the granularity of data representation is accomplished, the units are studied to observe whether they symbolize the data at a specific minimum level of granularity. Those units that symbolize too diverse input data are extensive to create a new small growing SOM at a succeeding layer, where the related data shall be characterized in more detail. These new maps however again grow in size until a specific development of the quality of data representation is reached. Units representing a previously quite homogeneous set of data, in opposition, will not need any additional expansion into subsequent layers. The acquired GHSOM consequently is completely adaptive to reflect, by its very architecture, the hierarchical structure inbuilt in the data, assigning additional space for the demonstrating of inhomogeneous areas in the input space.
Figure 1. Trained GHSOM

The GHSOM evolves to a structure of SOMs reflecting the hierarchical structure of the input data. A graphical demonstration of a GHSOM is presented in figure 1. The map in first layer encloses 3 X 2 units and suggests a rather rough grouping of the chief clusters in the input data. The six independent maps in the second layer offer a more detailed view of the data. The input data for single map is the subset so as to have been mapped onto the related unit in the upper layer. Two units positioned in one of the second-layer maps contain additional been extensive into third-layer maps to recommend sufficiently granular input data representation. It has to be seen that the maps includes various sizes supported on the structure of the data that moderate the problem of previously defining the structure of the architecture. The layer 0 serves up as an illustration of the complete data set and is essential for the managing of the growth process.

3.2 Initial Setup and Global Network Control

The advantage of the GHSOM architecture is its modification to the training data. The worth of this adjustment is deliberated by means of divergence among a unit's model vector and the input vectors indicates by this specific unit. Mainly, two dissimilar approaches can be utilized for the management of the growth process with the help of either the mean quantization error of a unit (normally utilized as a quality measure for data depiction with SOMs), or the absolute value, i.e. the quantization error of a unit. Officially, the mean quantization error of a unit i is computed based on the equation 1 as the mean Euclidean distance among its model vector \( m_i \) and the \( n_c \) input vectors \( x \) that are elements of the set of input vectors \( C_i \) mapped onto this unit i:

\[
mqe_i = \frac{1}{n_c} \sum_{x_j \in C_i} ||m_i - x_j||, n_c = |C_i|, C_i \neq \emptyset
\]  

(1)

The preliminary point for the GHSOM training procedure is the calculation of a mean quantization error \( mqe_0 \) of the unit forming the layer 0 map as characterized in equation 2 in which \( n_i \) specifies the number of every input vectors \( x \) of the input data set \( I \) and \( m_0 \) represents the mean of the input data (Rauber et al., 2002).

\[
mqe_0 = \frac{1}{n_i} \sum_{x \in I} ||m_0 - x||, n_i = |I|
\]  

(2)

The \( mqe \) calculates the variation of each input data mapped onto a specific unit and will be make use of to manage the growth procedure of the neural network. Mainly, the minimum characteristic of data representation of each unit will be indicates as a fraction, specified by a parameter \( r_2 \), of \( mqe_0 \).

Also, each unit must represent their individual subsets of data at a mean quantization error lower than a fraction \( r_2 \) of \( mqe_0 \), i.e. convincing the global termination criterion indicated in equation 3:

\[
mqe_i < r_2 \cdot mqe_0
\]  

(3)

On the other hand, the quantization error of the unit \( qe \) can be make use of as an alternative of the mean quantization error, resulting in a global termination condition.

\[
qe_1 = \sum_{x_i \in C_i} ||m_i - x_i||
\]  

(4)

\[
qe_1 < r_2 \cdot qe_0
\]  

(5)

3.3 Training and Growth Procedure of a Growing SOM

A newly generated map is qualified depending on the standard SOM training process. After definite predetermined number of training iterations the \( qes \) of all units as presented in Expression 5 are analyzed. A high \( qe \) indicates that an inhomogeneous part of the input space including different data, or at least a quite
large set of input data from a highly homogenous part of
the input space is characterized by this unit. Consequent-
ly, new units are necessary to offer more space for suitable
data representation. The unit with the highest \( qe \) is con-
sequently selected and is characterized as the error unit. The error unit is explained as \( e \). Ac-

Accordingly, the majority of different adjacent unit \( d \) by
the way of input space distance is chosen. This is carried
out by the way of contrasting the model vectors of every
neighboring unit with the model vector of the error unit
\( e \). A new row or column of units is commenced between
the error unit \( e \) and its most different neighbor \( d \). The
model vectors of the new units are kept as the average of
their equivalent neighbors.

In general, the growth procedure of a growing
SOM can be explain as follows. Let \( C_i \) specifies the
subset of vectors \( x_i \) of the input data that is mapped onto
unit \( i \), i.e., \( C_i \subseteq \mathbb{1}; \) and \( m_i \) specifies the model vector of
unit \( i \). After that, the error unit \( e \) is calculated as the unit
with the maximum quantization error as represented in
equation 6:

\[
 e = \arg \max_i \left( \sum_{x_i \in C_i} ||m_i - x_i|| \right), n_c = |C_i|, C_i \neq \emptyset \tag{6}
\]

After the error unit is selected, its enormously
dissimilar neighbor \( d \) is known as listed in equation 7,
where \( N_e \) represents the set of neighboring units of the
error unit \( e \):

\[
d = \arg \max_{\text{N}_e} \left( ||m_i - x_i|| \right), m_i \in \text{N}_e \tag{7}
\]

A row or column of units is integrated between
\( d \) and \( e \). To obtain a smooth positioning of the recently
included units in the input space, their model vectors are
primarily set as the means of their individual neighbors.
After including the learning rate and neighborhood
range are reorganizing to their initial values, and training
goes on in a SOM-like fashion for the next \( \lambda \) iterations.
This training method of single growing SOM is
eNumerously related to the Growing Grid model. The

distinction up to now is that a decreasing learning rate is
used and a decreasing neighborhood range moderately
fixed values.

### 3.4 Termination of Growth Process

The growth process goes on only until the
map’s mean quantization error, indicated as \( MQE \) in
capital letters, accomplishes a certain fraction \( r_1 \) of the
\( qe_u \) of the respective unit \( u \) in the upper layer
(specifically the unit encompassing the layer 0 map for
the first-layer map). The MQE of a map is calculated as
the mean of all units’ quantization errors \( qe_i \) of the
subset \( U \) of the maps’ units onto which data is mapped:

\[
 MQE_m = \frac{1}{n_u} \cdot \sum_{i \in u} qe_i, \quad n_u = |u| \tag{8}
\]

In general, the end condition for the growth of
a single map \( m \) is defined as:

\[
 MQE_m < r_1 \cdot qe_u \tag{9}
\]

where \( qe_u \) is the quantization error of the
respective unit \( u \) in the upper layer. It is obvious that the
smaller the parameter \( r_1 \) is preferred the larger the
resulting map will be, explaining its data at a higher
granularity. In case of a larger \( r_1 \), more detailed data
representation will be handed over to additional maps
promote down the hierarchy. The parameter \( r_1 \) as a
result acts as the control parameter for the depth/shallo-

ness of the resulting hierarchical GH-

SOM architecture.

### 3.5 Dynamic GHSOM (DGHSOM) with Controlled

Growth

It is essential for all the knowledge discovery
applications to have definite control on the growth of
the map. This can be accomplished by controlling the
control parameter \( r_1 \) (Alahakoon et al., 2000). The
requirement for a measure for controlling the growth of
the GHSOM is very important.

In case of using feature maps to recognize the
clusters, it is helpful if there is a way for initially observe
the most significant clusters and this will assist the data
analyst to obtain some idea of the whole data set, to get
finer clusters. In addition, this will also support the data
analyst in building decisions on regions of the data that
are not of attention and tune the finer clustering only to
regions of interest.

In order to accomplish this control, a process is
developed to point out the amount of spread required by
identifying a control parameter (Alahakoon et al., 2000).

The DGHSOM make use of a threshold value
called the Growth Threshold \( GT \) to build a decision
when to initiate new node growth. \( GT \) will make a
decision of the amount of spread of the feature map to
be created. As a consequence, when only an abstract
picture of the data is necessary, a large \( GT \) will result in
a map with a less number of nodes. In the same way, a
smaller \( GT \) will cause the map’s spreading out more.

The node growth in the DGHSOM is in
progress when the mean quantization error value of a
node goes ahead of the \( GT \). The mean quantization error
value for node \( i \) is calculated as

\[
 MQE_i = \sum_{j=1}^{D} \sum_{H_i} (x_{ij} - w_j)^2 \tag{10}
\]

where \( H_i \) represents the number of hits to the
node \( i \) and \( D \) is the dimension of the data. \( x_{ij} \) and \( w_j \)
represents the input and weight vectors of the node \( i \), in
The major reason of the GT is to permit the map to grow new nodes by presenting a threshold for the error value and the minimum error value is 0, it can be argued that for growth of new nodes
\[
0 \leq GT \leq \sum_{h_{max}}^{D} (x_{ij} - w_{j})^2 \quad (18)
\]

H_{max} can uncertainly be infinite, (18) becomes
\[
0 \leq GT \leq \infty . \text{ It is necessary to identify a function } f(r) \text{ such that }
0 \leq r \leq 1 \quad (19)
\]

and
\[
0 \leq D \times f(r) \leq \infty \quad (20)
\]

In other words, a function f(x) that takes the values 0 to ∞, when x takes the values 0 to 1, is to be identified.

A Napier logarithmic function of the type
\[
y = -a \times \ln(1 - x)
\]

is one such equation that satisfies these requirements. If \( \eta = 1 - r \) and
\[
GT \leq -D \times \ln(1 - \eta) \quad (21)
\]

Then
\[
GT = -D \times \ln(r) \quad (22)
\]

As a consequence, rather than providing a GT, which would take different values for various data sets, the data analyst can now present a value r, which will be used by the system to compute the GT value supported on the dimensions of the data. This will allow the DGHSOM’s to be acknowledged with their control parameters and can form a basis for comparison of different maps.

### 3.6 Extended Attribute-Oriented Induction

To trounce the disadvantage of major values and numeric attributes, an extension to the conventional AOI (Han et al., 1993) is used in this paper (Chung-Chian Hsu et al., 2006). This supply the ability of discovering the major values and an option for processing numeric attributes. For the exploration of major values, a parameter majority threshold \( \beta \) is initiated. If some values (i.e., major values) take up a major portion (exceeding \( \beta \)) of an attribute, the Extended AOI (EAOI) preserves those major values and generalizes other non major values. If no major values exist in an attribute, the EAOI proceeds like the AOI, producing the same results as that of the conventional approach. In addition, if \( \beta \) is set to 1, the EAOI degenerates to the AOI.

For solving the problems of making subjectively numeric concept hierarchies and generalizing boundary values, an alternative for processing numeric attributes is projected: Users can desire to compute the average and deviation of the aggregated numeric values in its place of generalizing those values to discrete concepts. Under this alternative, only definite attributes are generalized. The average and
deviation of numeric attributes of the combined tuples are calculated and then replace the original numeric values. The computed deviation discloses the dispersion of numeric values; the less the deviation is, the more concentrated the values are; or else, the more diversified the values are.

The EAOI algorithm is outlined as follows (Hsu, 2004):

**Algorithm:** An extended attribute-oriented induction algorithm for major values and alternative processing of numeric attributes

**Input:** A relation W with an attribute set A; a set of concept hierarchies; generalization threshold \( \theta \), and majority threshold \( \beta \).

**Output:** A generalized relation P.

**Method:**
1. Determine whether to generalize numeric attributes.
2. For each attribute \( A_i \) to be generalized in W,
   2.1 Determine whether \( A_i \) should be removed, and if not, determine its minimum desired generalization level \( L_i \) in its concept hierarchy.
   2.2 Construct its major-value set \( M_i \) according to \( \theta \) and \( \beta \).
   2.3 For \( v \in \text{Dom}(A_i) \), if \( v \in M_i \) construct the mapping pair as \( (v, v_i - M_{Li}) \), otherwise, as \( (v, v) \).
3. Derive the generalized relation P by replacing each value \( v \) by its mapping value and computing other aggregate values.

In Step 1, if numeric attributes are not to be generalized, their averages and deviations will be computed in Step 3. Step 2 aims at arranging the mapping pairs of attribute values for generalization. First, in Step 2.1, an attribute is removed either since there is no concept hierarchy defined for the attribute, or their higher-level conceptions are expressed in terms of other attributes. In Step 2.2, the attribute’s major-value set \( M_i \) is constructed, which consists of the first \( \alpha(<\theta) \) count leading values if they take up a major portion \( (\geq \beta) \) of the attribute, where \( \theta \) is the generalization threshold that sets the maximum number of distinct values permitted in the generalized attribute.

In Step 2.3, if \( v \) is one of the major values, its mapping value remains the same, i.e., major values will not be generalized to higher-level concepts. Otherwise, \( v \) will be generalized by the concept at level \( L_i \) by excluding the values enclosed in both the major-value set and the leaf set of the \( v_i \) subtree (i.e., \( v_i - M_{Li} \)), where \( M_{Li} = \text{Leaf}(v_i) \cap M_i \). Note that, if there are no major values in \( A_i \), \( M_i \) and \( M_{Li} \) will be empty. For that reason, the EAOI will act like the AOI. In Step 3, aggregate values are calculated, together with the accumulated count of merged tuples, which have identical values after the generalization, and the averages and deviations of numeric attributes of combined tuples if numeric attributes are determined not to be generalized.

4. Experimental Results

The future DGHSOM with EAOI mixed data clustering technique is experimented by means of UCI Adult Data Set and cleve dataset.

4.1 UCI Adult Data Set

This data set contains 15 attributes that contains eight categorical, six numerical and one class attributes. 10,000 tuples from the 48,842 tuples are preferred randomly for the evaluation.

Number of Resultant Clusters and Outliers

For the attribute choosing, the process of relevance analysis based on information gain is used. The relevance threshold was set to 0.1 and seven qualified attributes are attained: Marital-status, Relationship, Education, Capital_gain, Capital_loss, Age and Hours_per_week. The first three are categorical, and the others are numeric.

The map size is 400 units. The training parameters are set to the same with that of the previous experiment.

<table>
<thead>
<tr>
<th>Distance Criteria</th>
<th>SOM Cluster</th>
<th>SOM Outlier</th>
<th>GHSOM Cluster</th>
<th>GHSOM Outlier</th>
<th>DGHSOM Cluster</th>
<th>DGHSOM Outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d=0 )</td>
<td>88</td>
<td>-</td>
<td>75</td>
<td>-</td>
<td>61</td>
<td>4</td>
</tr>
<tr>
<td>( d \leq 1.44 )</td>
<td>19</td>
<td>-</td>
<td>9</td>
<td>-</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>( d \leq 2.82 )</td>
<td>9</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( d \leq 3^8\text{Adj} )</td>
<td>14</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

The number of resultant clusters by using SOM, GHSOM and DGHSOM with dissimilar distance criteria is presents in table 1. It can be seen that the projected DGHSOM with EAOI technique consequences in better categorization and it detects the clusters and outliers successfully.

Processing Time

For the reason that of the flexible shape of the network, the DGHSOM can point out a set of data with a smaller amount of nodes when comparing against the SOM and GHSOM and it is exposed in figure 2.
This becomes an advantage when training a network with extremely large data set, because the reduction in the number of nodes will result in a reduction in processing time and also less computer resources.

4.2 Cleve dataset

The second data set was a cleve dataset which is Dr. Detrano’s heart disease dataset that generated at the Cleveland Clinic modified to be a real mixed dataset. The dataset has 303 instances, each being described by 6 numeric and 8 categorical attributes. The instances were also classified into two classes, each class is either healthy (buff) or with heart disease (sick). The cleve dataset has 5 missing values in numeric attributes, all of them are replaced with the value of 0.

The clustering accuracy is measured suppose, the final number of clusters is k, clustering accuracy r is defined as

\[ r = \frac{\sum_{i=1}^{k} a_i}{n} \]

Where n is the number of instances in the data set, \( a_i \) is the number of instances occurring in the both cluster i and its corresponding class, which as maximum value. Consequently, the clustering error is defined as \( e = 1 - r \).

The dataset using here is the cleve data set into different number of clusters, varying from 2 to 9. For each fixed number of clusters, the clustering errors for various methods were compared.

Table 2. Performance of Different Clustering Techniques

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Clustering Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOM</td>
<td>0.358</td>
</tr>
<tr>
<td>GHSOM</td>
<td>0.291</td>
</tr>
<tr>
<td>DGHSOM</td>
<td>0.158</td>
</tr>
</tbody>
</table>

The resultant clustering error is determined by using SOM, GHSOM and DGHSOM with dissimilar distance criteria is presents in table 2. It can be seen that the projected DGHSOM with EAOI technique having minimum clustering error.

From the Figure 3, came to know that the DGHSOM technique is the most efficient technique as compared with the other two techniques.

5. Conclusion

Mixed data clustering is on of the difficult task and it is a major challenge in the field of research to offer a better clustering technique that will be able to handle mixed data successfully. GHSOM does not have any competence to control the growth of the map and furthermore the structure of GHSOM is static. In this paper, Dynamic Growing Hierarchical Self-Organizing Map (DGHSOM) with EAOI is projected for handling the mixed data. This paper concentrates on proficient clustering technique for mixed category data. DGHSOM can efficiently control the growth of the map and it has the potential to grow or change the structure. The experiment is carried out with the help of UCI Adult data set and Cleve dataset and it can be observed that proposed DGHSOM with EAOI presents better clustering result when compared against the SOM and GHSOM. Furthermore the processing time of the projected DGHSOM with EAOI is very low and the clustering error also very low.

Corresponding Author:
D. Hari Prasad
Assistant Professor (Selection Grade)
Department of Computer Applications
Sri Ramakrishna Institute of Technology
Coimbatore, India.
E-mail: hari.research@yahoo.com

References


3/3/2013