International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-1 Issue-4, April 2012

Density Based Clustering Scheme Using Dynamic Dissimilarity Measures

V. Kavitha, R.Manavalan

Abstract — Clustering methods are used support estimates of a data distribution have newly attracted much attention because of their ability to generate cluster boundaries of arbitrary shape and to contract with outliers efficiently. This paper proposes, Density based clustering using dynamic dissimilarity measure based on a dynamical system associated with Density estimating functions. Hypothetical basics of the proposed measure are developed and applied to construct a clustering method that can efficiently partition the whole data space. Clustering based on the proposed dissimilarity measure is robust to handle large amount of data and able to estimate the number of clusters automatically by avoid overlap. The dissimilarity values are evaluated and clustering process is carried out with the density values.

Index Terms — Clustering, kernel methods, dynamical systems, equilibrium vector, support, density.

I. INTRODUCTION

Recently, many researchers have successfully applied clustering methods based on the estimated support of a data distribution to solve some difficult and diverse unsupervised learning problems [2]. These methods, inspired by kernel machines such as kernel-based clustering [7] and support vector clustering, consist, in general, of two main stages: estimating a support function and clustering data points based on geometric structures of the estimated support function. The latter clustering stage is highly computer-intensive even in middle-scale problems and often shows poor clustering performance. Several researchers have therefore developed various techniques to reduce its computational complexity for the real applications, which include approximated graph techniques, spectral graph partitioning strategy [11], ensembles combined strategy [12], chunking strategies [3], pseudo hierarchical technique, or equilibrium-based techniques [6]. Despite their advantages over other clustering methods, the existing support-based clustering algorithms have some drawbacks. First, out-of-the sample points outside of the generated cluster boundaries cannot directly be assigned a cluster label. Second, the clustering results are very sensitive to the choice of kernel parameters used for a support estimate since the boundaries can show highly fluctuating behavior caused by small changes of the kernel parameters [13]. Finally, it is difficult to control the number of clusters when they are applied to
clustering problems with a priori information of the cluster numbers. To obtain number of cluster, perform computationally intensive parameter tuning process that involves repeated calls of support estimating step and cluster labeling step. To overcome these intrinsic handicaps, this paper proposes novel dissimilarity measure that can be applied to support-based clustering. Starting from a support function that estimates the support of a data distribution, and builds its associated dynamic process to partition the whole data space into so-called basin cells of equilibrium vectors and then construct a weighted graph consisting of equilibrium vectors. The constructed graph then defines a novel dissimilarity measure among equilibrium vectors with which we can perform inductive clustering that is, assigning cluster labels to out-of-the sample points as well as in-sample points.

Unlike the traditional SVC that focuses on the support vectors located on the cluster boundaries, the proposed dissimilarity measure focuses on the equilibrium vectors located inside the generated clusters and can be applied to any kernel-based support or density estimating functions if they can reveal clusters of a data distribution well. Finally, we perform experiments to show that clustering based on the proposed dissimilarity measure is robust to the choice of kernel parameters and is able to generate the user-specified number of clusters without the parameter tuning process.

This paper Organized as follows, Section1 describes Dynamic Dissimilarity Measures to avoid overlap, Section2 demonstrate Clustering based on Dynamic Dissimilarity Measure and Section 3 estimate the performance analysis of both Support based clustering and Density based clustering.

II. DYNAMIC DISSIMILARITY MEASURE

Dynamic dissimilarity measure is used to avoid overlapping of cluster. This technique is used to determine both Supports based clustering and Density based clustering.

A. Support of a data distribution

A support function (or quantile function) is roughly defined as a positive scalar function $f : R^n \rightarrow R^+$, where a level set of f estimates support of a data distribution. It can normally be decomposed into several separate connected components Ci; i = 1, . . .;m, i.e.,

$$L_f(r) = \{ x \in R^n : f(x) \le r \} = C_1 \cup ... \cup C_m \qquad -(1)$$

Manuscript published on 30 April 2012.
* Correspondence Author (s)
Kavitha V*, Department of Computer Science, Periyar University, KSR College of Arts and Science ,Tiruchengode.(e-mail:kavitha.vaidya@gmail.com).
Manavalan, Department of Computer Science, Periyar University, KSR College of Arts and Science, Tiruchengode.(e-mail: manavalan_r@rediffmail.in).

Popular support functions that are trained from data and are shown to characterize the support (or quantile) of a class conditional distribution include:

A support function generated by the support vector domain description (SVDD) method in [8],

$$f(x)=1-2\sum_{j\in J}\alpha j e^{-q|x-x_i|^2}+\sum_{i,j\in J}\alpha_i\alpha_j e^{-q|x-x_i|^2} \quad -(2)$$

where x_i and α^j for $j\in J$ are support vectors and their corresponding coefficients constructed by optimizing the SVDD model. q denotes the parameter of the Gaussian kernel.

A Gaussian process support function generated by the Gaussian process clustering in [5],

$$f(x)=k(x)TkC-1(x) \quad -(3)$$

Where C is a positive-definite covariance matrix with elements $C_{ij}=C(x_i\ x_j;\Theta)$ and k(x) = (C(x; x1), . . . ; C(x,xN))T , One commonly used covariance function is

$$C(x_i\ x_j;\Theta=v_0\exp\left\{-\frac{1}{2}\sum_{m=1}^{n}l_m\left(x_j{}^m-x_j{}^m\right)\right\}+v_{i+\delta_{ij}}v$$ Where a set of hyper parameters, $\Theta=\{v_0,v_1,v_2,l_1,\ldots,l_n\}$, can be determined by maximizing the marginal likelihood. Traditional clustering methods based on a support function assign each connected component Ci to a separate cluster. Hence, they are not inductive since the set of clusters Ci; i = 1. . . m does not cover the whole input space.

B. *Graph Associated with a Support*

With the aid of the dynamical system approach, we can construct a weighted graph that preserves and simplifies the cluster structures of Lf (r) [14]. To be specific, we build the following dynamical system associated with f:

$$\frac{dx}{dt}=F(x):=-\nabla f(x) \quad -(4)$$

The existence of a unique solution (or trajectory) x(0) : $R\to R^n$ for each initial condition x(0) can be guaranteed under some mild conditions imposed on F. A state vector x satisfying F(x) = 0 is called an equilibrium vector of system (4). We say that an equilibrium vector x of (4) is hyperbolic if the Jacobean matrix of F at x, denoted by JF (x), has no zero eigenvalues. A hyperbolic equilibrium vector is called (1) an (asymptotically) stable equilibrium vector if all the eigenvalues of its corresponding Hessian are positive, or 2) an unstable equilibrium vector (or an UEV) otherwise.

A hyperbolic equilibrium vector x is called an index-k equilibrium vector if its Hessian has exactly k negative eigenvalues [10]. The basin cell of a SEV, s, is defined as the closure of the set of all the points converging to s when process (4) is applied, get the following equation

$$\overline{A(S)}:cl\left\{x(0)\in R^n:\lim_{t\to\alpha}x(t)=s\right\}. \quad -(5)$$

The boundary of the basin cell defines the basin cell boundary, denoted by $\partial\overline{A(S)}$. One distinguished feature of system (4) is that we can partition the whole data space into several separate basin cells of the SEVs under some mild conditions, [9]

$$R^n=\bigcup_{i}^{M}\overline{A(S_i)} \quad -(6)$$

Where {si: i = 1. . . M} is the set of the SEVs of system (4). Since each data point converges almost surely to one of the SEVs when system (4) is applied, can easily identify a basin cell to which a data point belongs by its corresponding SEV. An SEV, s_a, is said to be adjacent to another SEV, s_b, if there exists an index-one saddle equilibrium vector $d\in\overline{A(S_a)}\bigcap\overline{A(S_a)}$. Such an index-one saddle equilibrium vector, d, is called a transition equilibrium vector (TEV) between s_a and s_b. From a practical viewpoint, the notions of adjacent SEVs and TEVs enable us to build a weighted graph Gr = (Vr, Er) describing the connections between the SEVs with the following elements:

The vertices Vr of Gr consist of SEVs, $s_{i,}$ in V with f ($s_{i,}$) < r. 2. The edge Er of Gr is defined as follows: ($s_{i,}$, s_j)$\in$ Er with the edge weight distance, dE ($s_{i,}$, s_j) = f (d), if there is a TEV, d, between $s_{i,}$ and s_j with dE ($s_{i,}$, s_j) < r.

It can then be shown [14] that two SEVs, $s_{i,}$ and s_j, are in the same connected component of the graph Gr if, and only if, $s_{i,}$ and s_j are in the same cluster of the level set Lf (r); that is, each connected component of Gr corresponds to a cluster of Lf (r). This result enables us to build a simplified graph Gr that preserves the topological structure of the level set Lf (r).

C. *Dynamic Dissimilarity Measure*

The basic elements for the constructed graph Gr are the SEVs, $s_{i,}$, whose basin cell $\overline{A(S_i)}$ can be identified as the set consisting of similar objects with respect to system (4). Now to define a dissimilarity measure defined on all pairs of the SEVs, we first present the next theorem, which serves as a theoretical basis to extend the distance between two adjacent SEVs to the distance between any pair of SEVs by guaranteeing the existence of a connected graph Gr for some large r.

Theorem 1. For a given support function f defined by (2),assume that for any $x(0)\in R^n$, each connected component of the level set Lf (r) is compact, where r = f(x0). Then, there exists a γ > 0 such that the graph Gr is connected for all r > γ .

Proof: Choose $M<\max_{j\in J}\|x_i\|$.

Let $v(x)=\frac{1}{2}\|x\|^2$ an DM ={x: $\|x\|$ < M}.

Then, for any x0$\in${x : $\|x\|$ > M}, the trajectory x(t) of (4) starting from x(0) always satisfies

$$\frac{\delta}{\delta t}v(x(t))_{t=0}\ x_0{}^T(-\nabla f(x_0))=-x_0{}^T(x_0)-\sum_{j\in J}\alpha_j(x_0)x_j)$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$= -\|x_0\|^2 + \sum_{j\in J} \alpha_j(x_0) x_0^T x_j$$

$$\le -\|x_0\|^2 + 1\|x_0\| \max_i \|x_j\| < 0$$

Where

$$\alpha_j(x_0) = \frac{\alpha_j \exp\left(-q\|x_0 - x_j\|^2\right)}{\sum_{j\in J} \alpha_j \exp\left(-q\|x_0 - x_j\|^2\right)}$$

This implies that $\|x_0\|$ is always strictly decreasing when process (4) is applied and, in particular, all the equilibrium vectors of (4) are inside the bounded and connected set DM. Since all the trajectories of (4) converge to one of its equilibrium vectors [4], the trajectory of (4) starting from any x0 with $\|x_0\| > M$ always enters into the region DM.

Next, we show that there exists a $\gamma > 0$ such that DM $Lf(\gamma)$, where $Lf(\gamma) = \{x \in R^n : f(x) < \gamma\}$ For any $x_0 \in DM$,

Let's have $\|x_0 - x_j\| \le \|x_0\| + \|x_j\| < 2M, \forall j \in J$. Hence, where C = $1 + \sum_{i,j} \alpha_i \alpha_j e^{-q|xi - x_j|^2}$ and q is a preset parameter of the Gaussian kernel in (2).
By choosing γ _ > C _ e_4qM2.

$$f(x_0) = C - 2\sum_{j\in J} \alpha_j e^{-q|x_0 - x_j|^2} \le C - e^{-4qM^2}.$$

Finally, let r > γ . By the invariant property of Lf (r) (i.e., if a point is on a connected component of Lf (r), then its entire positive trajectory lies on the same component), we should have that for any point x0 ∈ (Lf (r)/ DM), the trajectory starting at x(0) = x0 should first hit ∂ DM, and then enter inside the region DM since all the equilibrium vectors are inside the region DM and all the trajectories converge to one of its equilibrium vectors. This implies that the set DM is a strong deformation retract of the level set Lf (r). Also, from the uniqueness of the trajectories [18], the boundary ∂ DM is homeomorphic to ∂ Lf (r), which implies that Lf (r) is connected from the connectedness of DM. Since the cardinality (i.e., the number of connected components) is the same between the graph G and Lf (r) [14], therefore, the graph Gr is connected. This theorem motivates us to define a dissimilarity measure on a connected graph G = Gr for r > γ as follows:

***Definition 1*:** (Dissimilarity measure). Let a connected graph G = (V, E) be given. For a pair of SEVs, s_i and s_j, in V, we can define the distance dG (s_i, s_j)as

$$d_G(s_i, s_j) = \min\left\{d_E(s_i, s_j), \max_{k-1,-h} d_E(s_{ik-1}, s_{ik})\right\}$$

for a path sequence (with no cycle) s_i = si0 , si1 , . . . ; sih-1 , $s_{ih,} = s_{i,}$ such that (s_{ik-1} , s_{ik}) ∈ E for each k = 1. . . . . j, which endows a graph G with a dissimilarity measure. Assume dE (s_i, s_j) = α if $s_i, s_j \notin E$.

Geometrically, the distance dG (s_i, s_j) takes the smallest function value along a path connecting two SEVs to escape from one SEV and move on to the other SEV.

III. CLUSTERING BASED ON A DYNAMIC DISSIMILARITY MEASURE

Generally speaking, a support function (e.g., f in 2) is often very sensitive to the choice of kernel parameters, so is the clustering structure described by Lf (r). For example, if clusters overlap in some region, it is difficult or even impossible to find the kernel parameter to separate it. Moreover, to control the number of clusters, we have to change the kernel parameters (hence, the support function f) by trial and error, where each alteration of kernel parameters entails repeated calls of a quadratic programming solver and a cluster labeling algorithm, which is computationally intensive.

The derived dynamic dissimilarity measure on the graph G can help us to overcome these drawbacks when it is applied to clustering. Specifically, with an input K > 1 denoting the number of clusters, we begin with every SEV representing a singleton cluster. Denote these clusters C1 = {s1}. . . Cv = {sv}. At each step, the closest two clusters (i.e., two separate clusters containing two adjacent SEVs with the least edge weight distance) are merged into a single cluster, producing one less cluster at the next higher level. This process is terminated when we get K clusters starting from v clusters. Kruskal's algorithm is used to find the distance between clusters.

Kruskal's Algorithm:

1: Given a number of clusters K

2: Rearrange the index k = 1; . . . ; e in such a way that f(d1) < f(d2) < ….< f(de)

3: Start with initial clusters as C1 = {s1}, . . ., CM = {sM}. In this initial step, the distance between two clusters is defined as

$$d(C_i, C_j) = \begin{Bmatrix} d_e(s_i, s_j) = f(d_k) if s_i, s_j \notin E \\ \infty otherwise \end{Bmatrix}$$

This algorithm possesses a monotonicity property. That is, the dissimilarity between merged clusters is monotone increasing with the level of the merger. Thus, the binary tree, called a dendrogram, can be plotted so that the height of each node is proportional to the value of the intergroup dissimilarity between its two daughters.

This nice property makes the method less sensitive to the choice of kernel parameters unlike the traditional support-based clustering algorithms such as the SVC. Also, Algorithm 1 enables us to control the number of clusters by manipulating the constructed graph without changing the kernel parameters.

The most time-consuming step in Algorithm 1 involves locating SEVs and TEVs for constructing a weighted graph G = (V, E). If we let m be the average number of iterations for locating SEVs from data

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

points via steepest descent process, then the time complexity of getting all SEVs and TEVs of system (4) is O(Nm) and O(M_2d), respectively. Here, M is the number of SEVs and d is the average number of computing TEVs between SEVs [14].

IV. DENSITY BASED CLUSTERING SCHEME

The support based clustering model is enhanced in two ways. They are support function enhancement with optimal cluster boundary selection process. The density estimation function is used to replace the support estimation function. The clustering process is carried out with the density values. The system enhances the support function to improve cluster boundary estimation. Transaction distributions are used in the density analysis. Dynamic distance measure is used to estimate distance between transactions. Cluster boundary is extracted from the graph. The problem is divided into four modules. They are support estimation, density analysis, graph construction and clustering process.

Support estimation module is designed to extract attribute support ratio values. Density analysis is carried out to fetch attribute distributions. Graph construction module is used to construct distance based graph. Clustering process is used to partition the transaction data values.

A. Support Estimation

The support estimation is performed to find the support ratio for the attributes. Distance measures are calculated using support values. The attribute information and its values are used in the support estimation process. The attribute values and their ratio are used to estimate the support values. The static distance measure uses the transactional data values for the similarity analysis. Dynamic distance measure is used to estimate the distance between the transactions. The support ratio and transactional data values are used in the distance estimation process. Equilibrium vector is constructed with distance values.

B. Density Analysis

Attribute distributions are estimated in the density analysis. Attribute values and their distribution levels are estimated in the system. Attribute values and their intervals are considered in the distance estimation. Distribution intervals are used to fetch attribute relationship. Dynamic distance measure uses the density information. Density information and transactional data values are used in the distance estimation process. Transaction distances are used to update vector information. Both support and density functions are used to reflect global relationship for the transactional data values.

C. Graph Construction

The system uses two types of graphs. They are support based graph and density based graphs. Support based graph uses the distance values based on support functions. Density based graph uses the density based distance values. The graphs reflect the data relationship and their similarity levels. The graph supports the user s to estimate cluster intervals. Cluster boundaries are estimated from the graphs.

D. Clustering Process

Data partitioning operation is carried out under the module. Clustering process is initiated with the cluster count values. The data values are partitioned with respect to the cluster count. The graph data values are divided into the partitions using the cluster count. The user can perform the clustering process with out the cluster count value. The ystem also automatically estimates the cluster count. Data partitions are done with distance measures. Cluster boundaries are estimated using the distance values. The cluster count is decided with the support of the distance and graph values. Performance analysis is carried out for support based and density based clusters. The F-measure function, precision and recall metrics are used for the performance analysis. Figure 4.1 shows the process flow of the proposed system.

***Algorithm**: Clustering based on a dynamic dissimilarity measure*

Step 1:Consider Density function f and its associated weighted graph G = (V, E) where si, i = 1. . . M is the set of SEVs and dk, k = 1,.. . . , e is the set of TEVs.

Step 2: Initialize the number of clusters.

Step 3: Arrange the index k = 1; . . . ; e in such a way that f(d1) < f(d2) < ….< f(de)

Step 4: Start with initial clusters as C1 = {s1}, . . ., CM = {sM}

Step 5: Distance between two clusters is defined as

$$d\left(C_i, C_j\right) = \begin{Bmatrix} d_e\left(s_i, s_j\right) = f\left(d_k\right) if s_{i,} s_j \notin E \\ \infty otherwise \end{Bmatrix}$$

Pseudo code:

1: Initialize I = 1 and k = 1
2: Check the condition using while loop
While l < M - K do
3: Find the SEVs si; sj with its edge weight
dE(si, sj) = f(dk)
4: if si, sj are not in the same clusters then combine them using CM+l = Ca ∪ Cb where si ∈ Ca and sj ∈ Cb.
5: Find the minimum distance for all remaining cluster using following function
d(CM+l,Cu)= min{d(Ca,Cu) d(Cb,Cu
6: Add cluster CM+l as a new cluster and remove clusters Ca and Cb.
7: Increase the value of the variable I and K
I = I + 1; k = k + 1
8: else
9: k = k + 1
10: end if
11: end while

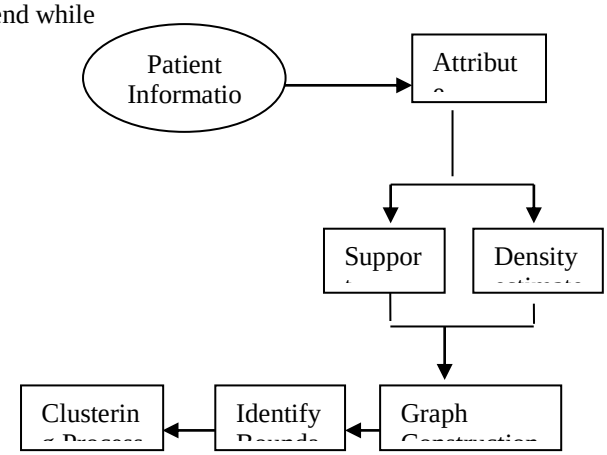

Figure 4.1 Structure of proposed System

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Patient Data set is analyzed to remove the noisy data. Support and density values are calculated using corresponding functions. Then the graph is constructed using those values.After the construction of graph the boundary is identified to perform clustering process.

V. EXPERIMENTAL ANALYSIS

The density ratio based data partitioning system is tested using the breast cancer diagnosis datasets. The dataset is collected from the UCI (University of California, Irwin) machine learning repository [20]. It provides information about the breast cancer patient diagnosis information. The class information and associated symptom details are provided in the dataset. The dataset may consist of noise records. Noise elimination process is also performed on the data sets.

A. Evaluation Method

The experimental evaluation is performed with the F-measure and their parameters to evaluate the accuracy of the clustering algorithms.

The F-measure measurement is used to represent the cluster accuracy information. The F-measure is a harmonic combination of the precision and recall values used in information retrieval. Each cluster obtained can be considered as the result of a query, whereas each pre-classified set of transactions can be considered as the desired set of transactions for that query. Thus, the precision P (i,j) and recall R(i,j) of each cluster j for each class i can be calculated.

If ni is the number of members of the class i, nj is the number of members of the cluster j and nij is the number of members of the class i in the cluster j, then P (i,j) and R(i,j) can be defined as follows in the table:

STATISTICAL PARAMETERS

PPARAMETERS	FORMULA
Precision P(i.j)	$P(i,j)=\frac{n_{ij}}{n_j},$
Recall R(i,j)	$R(i,j)=\frac{n_{ij}}{n_i}$
F-measure F(i,j)	$F(i,j)=\frac{2*P(i,j)*R(i,j)}{P(i,j)+R(i,j)}$
F-measure of the whole clustering result F	$F=\sum_i \frac{n_i}{n}\max_j (F(i,j)),$

Where n is the total number of transactions in the dataset. In general, F-measure will produce the better clustering result for large number of transactions.

B. Performance Analysis

The performance of density ratio based data partitioning scheme is carried out with a set of experiments.

The experiments are conducted in an Intel Dual Core processor with 2.5. GHz speed and 1GB of memory. The breast cancer dataset is used for the analysis. The datasets are collected from the UCI machine learning repository. The testing is performed with different transaction and cluster count. Static and dynamic clustering models are tested with different similarity measures. The cluster accuracy level is analyzed with F-measure and their parameters performance measure.

The F-measure is used to evaluate the cluster quality. The precision and recall measures are used in the F-measure calculation. Both the precision and recall measurements are calculated using the query based assessment. The actual class values are compared with the retrieved class values. The F-measure is estimated for all the clusters with different similarity measures.

The performance analysis is carried out with set of 1000 transactions. The comparative analysis is carried out to analyze the static and dynamic clusters accuracy levels. Static clustering technique partitions the transaction collection with a fixed cluster count collected from the user. In the dynamic clustering model the system automatically estimates the feasible cluster count.

The density ratio based similarity measure is used in the clustering process. The precision and recall values are calculated for both the techniques. The F-measure values are calculated and compared in figure 6.1. The comparative analysis shows that the dynamic clustering model produces more than 25% accuracy level than the static clustering model.

Table 5.1. F-Measure Analysis - Static Clusters Vs Dynamic Clusters

Records	Static Clusters	Dynamic Clusters
200	0.665	0.903
400	0.683	0.908
600	0.702	0.916
800	0.722	0.921
1000	0.741	0.939

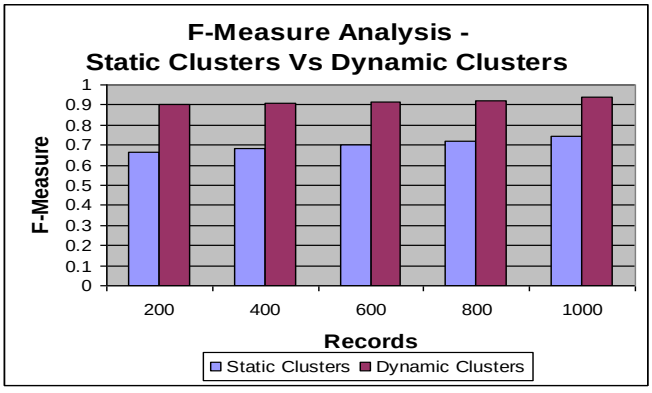

Figure 5.1. F-Measure Analysis - Static Clusters Vs Dynamic Clusters

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The static clustering process is analyzed for the support distance and density distance models. The support ratio is used in the support distance model. The distribution information is used in the density distance model. The performance of support distance and density distance models are compared in table 6.1. The results show that the density distance model achieves accuracy level more than 15% better than the support distance model.

Table 5.2. Static Clusters F-Measure Analysis - Support Distance Vs Density Distance

Records	Support Distance	Density Distance
200	0.573	0.665
400	0.590	0.683
600	0.603	0.702
800	0.619	0.722
1000	0.636	0.741

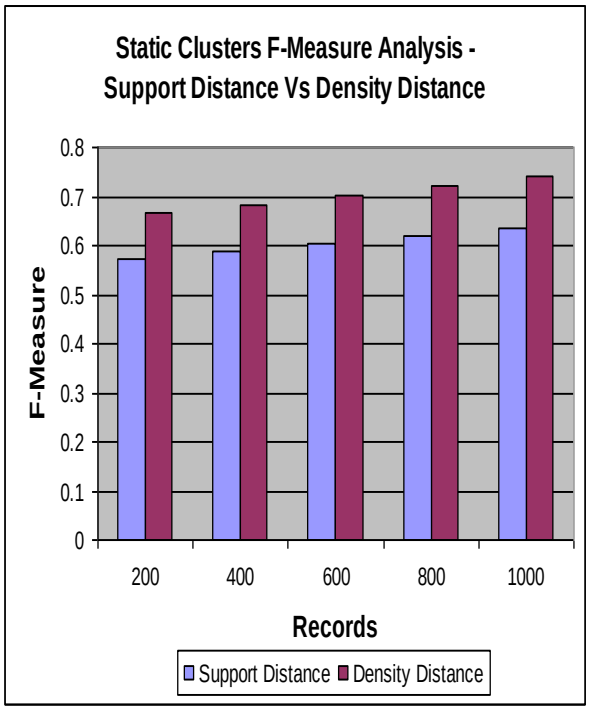

Figure 5.2. Static Clusters F-Measure Analysis - Support Distance Vs Density Distance

The dynamic clustering model automatically estimates the cluster count based on the distance intervals. The data partitioning is performed with the detected cluster count. The F-measure analysis is performed for both the support distance and density distance models. The comparative results are shown in figure 6.2. The results show that the density distance model produces accuracy level more than 10% than the support distance model. The dynamic clustering model with density distance is the feasible clustering mechanism for the data partitioning requirements.

Table 5.3. Dynamic Clusters F-Measure Analysis - Support Distance Vs Density Distance

Records	Support Distance	Density Distance
200	0.771	0.903
400	0.789	0.908
600	0.807	0.916
800	0.825	0.921
1000	0.841	0.939

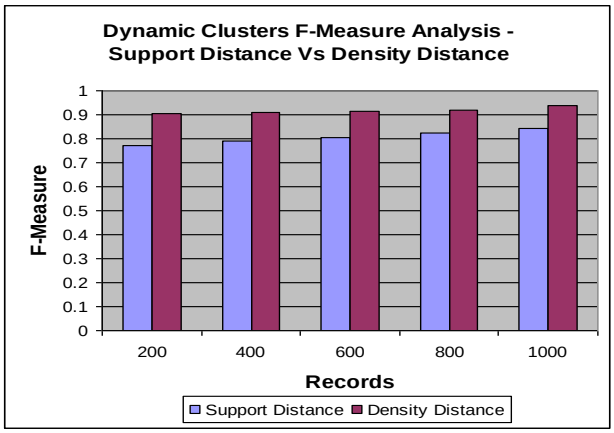

Figure 5.3. Dynamic Clusters F-Measure Analysis - Support Distance Vs Density Distance

VI. CONCLUSION

In this paper, dynamic dissimilarity measure for Density-based clustering has been proposed. The support ratio based model is not well suited for attributes with huge values. The density ratio based model is used to partition the transactions with huge amount of attribute values. The clustering scheme is tested with breast cancer patient data sets. The data values are analyzed with different intervals and cluster count. The clustering techniques are compared with their accuracy levels. The density ratio based data partition scheme produce more accurate results than the support ratio based data partition scheme. The automatic clustering model produces data partitions with efficient assessment of the distance intervals.

REFERENCES

[1] Daewon Lee and Jaewook Lee "Dynamic Dissimilarity Measure for Support-Based Clustering" IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 6, June 2010.
[2] F. Camastra and A. Verri, "A Novel Kernel Method for Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, May 2005.
[3] T. Ban and S. Abe, "Spatially Chunking Support Vector Clustering Algorithm," Proc. Int'l Joint Conf. Neural Networks, 2004.
[4] J. Lee and D. Lee, "An Improved Cluster Labeling Method for Support Vector Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, 2005.
[5] H.-C. Kim and J. Lee, "Clustering Based on Gaussian Processes," Neural Computation, vol. 19, no. 11, pp. 3088-3107, 2007.
[6] D. Lee "Equilibrium-Based Support Vector Machine for Semi-Supervised Classification," IEEE Trans. Neural Networks, 2007.
[7] D. Zhang, "A Novel Kernelised Fuzzy C-Means Algorithm with Application in Medical Image Segmentation," Artificial Intelligence in Medicine, 2004.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

[8] D. Lee, "Domain Described Support Vector Classifier for Multi-Classification Problems," Pattern Recognition, 2007.
[9] J. Lee, "A Novel Three-Phase Trajectory Informed Search Methodology for Global Optimization," J. Global Optimization, 2007.
[10] J. Lee, "An Optimization-Driven Framework for the Computation of the Controlling UEP in Transient Stablity Analysis," IEEE Trans. Automatic Control, 2004.
[11] J. Park, "Support Vector Clustering Combined with Spectral Graph Partitioning," Proc. 17th Int'l Conf. Pattern Recognition, 2004.
[12] W.J. Puma-Villanueva, "Improving Support Vector Clustering with Ensembles," Proc. Int'l Joint Conf. Neural Networks, 2005.
[13] M.S. Hansen and R. Larsen, "Robust Pseudohierarchical Support Vector Clustering," Proc. Scandinavian Conf. Image Analysis, 2007.
[14] J. Lee, "Dynamic Characterization of Cluster Structures for Robust and Inductive Support Vector Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, 2006.

AUTHOR PROFILE

Ms.KAVITHA.V Obtained MCA from Karunya University, Coimbatore, Tamilnadu, India, in the year 2009. She pursues M.Phil in Computer Science. Her areas of interest are Data mining and soft computing,

Mr. MANAVALAN R Obtained M.Sc., Computer Science from St.Joseph's College of Bharathidasan University,Trichy, Tamilnadu, India, in the year 1999, and M.Phil., in Computer Science from Manonmaniam Sundaranar University, Thirunelveli, Tamilnadu, India in the year 2002. He works as Asst.Prof & Head, Department of Computer Science and Applications, KSR College of Arts and Science, Thiruchengode, Namakwa, Tamilnadu, India. His areas of interest are Medical image processing and analysis, soft computing, pattern recognition and Theory of Computation.