# Finding the Number of Clusters using Visual Validation VAT Algorithm

G.Komarasamy [#1],Amitabh Wahi[*2]
[1]Assistant Professor–Senior Grade, Department of Computer Science and Engineering,
Bannari Amman Institute of Technology, Sathyamangalam, India
[2] Professor, Department of Information Technology,
Bannari Amman Institute of Technology, Sathyamangalam, India
[1] gkomarasamy@gmail.com
[2] awahi@rediffmail.com

***Abstract*—Clustering is the process of combining a set of data in such a way that data in the same group are more similar to each other than the groups (clusters). K-Means is an algorithm for widely used in clustering techniques. But in this algorithm some of the issues are determined i.e. K-value selected by user is the main disadvantage. To overcome the drawback visual methods such as the VAT algorithm generally used for cluster analysis, also it is used to obtain the k-value prior to clustering. But the estimated result does not match with the true (but unknown) value in many cases. Then Spectral VAT algorithm was implemented. This spec-VAT algorithm is more efficient than VAT algorithm for complex data sets. The Spec-VAT based algorithms such as A Spec-VAT, P Spec-VAT and E Spec-VAT is also used to find out the cluster value efficiently. But the range of k value is either directly or indirectly given to spectral based VAT algorithms. In this paper we propose direct visual validation method and divergence matrix. In this proposed work the value of k or the range of k is neither directly nor indirectly specified by the users. Instead of k value, we propose a new method of comparing objects and from that result. We choose an object which is closer than other object, From the V$^2$VAT (Visual Validation VAT) algorithm the experimental result shows that the proposed algorithm is much better than the other algorithms.**

**Keyword-VAT algorithm, visual validation, divergence matrix, V$^2$VAT algorithm**

## I. INTRODUCTION

Data Clustering is a technique in which, the data set is divided into sub data sets and then similar data is physically stored together. Sometimes, Clustering techniques are also referred as unsupervised technique because there is no particular dependent. Most often clustering and classification are confused, but there are certain differences among them. The data are assigned as pre defined classes in classification. But in clustering method the classes are also defined.

Clustering algorithms are mainly divided into two categories: Hierarchical algorithm and Partition algorithm. A hierarchical clustering algorithm divides the given data set into smaller subsets in hierarchical fashion. A partition clustering algorithm partitions the data set into desired number of sets in a single step. Many methods have been proposed to solve clustering problem.

## II. RELATED WORK

One of the most popular and simple clustering algorithms is K-means, it has been initially published in 1955. However, this requires a prior knowledge of the number of clusters and the subsequent selection of their centroid. This selection of k-value itself is an issue and sometimes it is hard to predict before the number of clusters that would be there in data. Various clustering methods have been proposed to estimate k-value (number of cluster), e.g., [7], [9], [12], [15], [16], [18] and [19] by choosing the best partition among a set of various partitions.

In contrast, cluster tendency assessment attempts to estimate k-value before clustering occurs. Estimating the number of clusters in data sets automatically, it is based on Visual Assessment of cluster Tendency (VAT) of a data set. This existing model focuses on one method for generating RDIs, namely VAT of Bezdek and Hathaway [1]. The basic input for most of the clustering algorithms is number of clusters. It is very difficult to determine previously. A new method called Dark Block Extraction (DBE) for automatically estimating the number of clusters in unlabeled data set, which is based on existing algorithm VAT using several common images and signal processing techniques of Wang, Leckie&Bezdek [21]. But in this system many issues like perplexing, and its inability in histogram overlapping. Therefore it is implemented by a new technique called Extended Dark Block Extraction of Asadi, Saikrishna&Subbarao [22].

There are many methods for assessing VAT of squaredissimilarity matrix based on crisp, fuzzy or probabilistic models [3], [2]. These methods are only applicable for square matrix. So implemented the Method

that assesses cluster tendency in rectangular (non square i.e. m × n) dissimilarity matrices of Bezdek, Fellow & Hathaway [10]. In VAT analysis methods automatically estimating the number of clusters in larger data sets becomes difficult. To overcome these problem, existing system performs visual cluster analysis on a large data set to apply the Spectral VAT algorithm of Ramamohanarao, Wang &Leckie[23]. It automatically derives the number of clusters in the SpecVAT images using these techniques A-SpecVAT, P-SpecVAT and E-SpecVAT algorithm.

It should be mentioned that a previous version of this work has appeared in [23]. In this paper, main issue is existing index based validation is not suitable for our visual algorithm. In contrast to the preliminary version, major changes of this paper are summarized as follows:

- We propose a method of divergence matrix instead of dissimilarity matrix in the preliminary version paper.
- We provide a solution for problem in the previous system i.e. direct visual validation method is implemented for an index based validation method.
- We propose a new method in this algorithm. So there is no need to give a k-value initially by the user.
- We provide experiments on several synthetic and real data sets to evaluate this new algorithm, and the results demonstrate with effectiveness.

## III. VISUAL ASSESSMENT OF CLUSTER TENDENCY(VAT)

The VAT algorithm is used as a visual method for clustering analysis. It displays an image of dissimilarities data. In gray scale VAT image each pixel is represent the scaled dissimilarities value. High dissimilarities viewed as white pixels and low dissimilarities viewed as black pixels. Definitely Each object is similar with itself, so the diagonal element of the dissimilarity matrix is always zero. The elements in the off diagonal is scaled to the range of (0,1). Several algorithms have extended VAT for assessment problems [5],[8],[11],[17] and [21].

We can generate a RDI by using any of the existing scheme in [1],[4],[13],[14] and [24]. Here Reordered Dissimilarity Image (RDI) is generated using VAT, RDI highlights potential clusters as a set of "dark blocks" along the diagonal of the image.A dark block along the diagonal of the $I(D_*)$ is a submatrix of "similarly small" dissimilarity values. From this, the dark block is refers a cluster of datas which are similar to each other, VAT algorithm is shown bellow.

### *A. VAT Algorithm Steps*

**Input**: An n × n scaled matrix of pairwise dissimilarities $D=[d_{ij}]$, with $1 \geq d_{ij} \geq 0$; $d_{ij}= d_{ji}$; $d_{ii}=0$, for $1 \leq i,j \leq n$

(1) Set I=ø, J={1,2,. . . . , n} and π=(0,0,….,0)
Select (i,j) Є $\arg_{p\in J,q\in J}\max\{d_{pq}\}$.
Set π(1)=i, I←{i} and J←J-{i}.

(2) Repeat for t=2,3,…,n
Select (i,j) Є $\arg_{p\in I,q\in J}\min\{d_{pq}\}$.
Set π(t)=j, update I←I ∪ {j} and J←J-{j}.

(3) From the reordered matrix $\widetilde{D}=[d_{ij}]=[d_{\pi(i)\pi(j)}]$,for $1 \leq i,j \leq n$

**Output:** A scaled gray-scale image $I(\widetilde{D})$,in which $\max\{\tilde{d}_{ij}\}$ corresponds to white and $\min\{\tilde{d}_{ij}\}$ to black.

Many possible ways to obtain a RDI, here we use VAT to generate RDIs of unlabeled data. Let $O=\{o_1,o_2,\ldots,o_n\}$ denote n objects in the data (e.g., fish, flowers, beers, etc.). Vectorial data have the form $F=\{f_1,f_2,\ldots,f_n\}$, $f_i \subset R^h$, where each coordinate of the vector $f_i$ provides a feature value of each of h attributes (i.e., $a_j$, j=1,2,….,h) corresponding to an object $o_i$. Sometimes relational data are directly recorded, such as pairwise dissimilarities (or similarities) between objects, represented by an n × n symmetric matrix D.

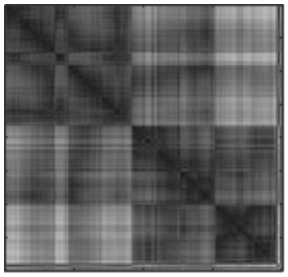

Fig.1. VAT image

We can always convert F into dissimilarities D as $d_{ij} = \|f_i - f_j\|$, $1 \leq i,j \leq n]$ in any vector norm in $R^h$. Generally, the dissimilarity matrix satisfies $1 \geq d_{ij} \geq 0$; $d_{ij}= d_{ji}$; $d_{ii}=0$, for $1 \leq i,j \leq n$.The VAT algorithm displays

a dissimilarity matrix D as a gray-scale image and element of the dissimilarity matrix are in the rande between (0,1) as shown in Fig 1.

Even RDI are widely used, it is only effective at highlighting cluster tendency in data sets that contain compact separated clusters.But, in many practical applications involve highly complex structure in data sets. So, in this paper the new approach was implemented to obtain RDI which combines VAT and spectral analysis.

## IV. $V^2$VAT ALGORITHM (VISUAL VALIDATION VAT)

Three important points about $V^2$ VAT are mentioned in the following are:

1. Only a pairwise divergence matrix D is required as the input. When the vectorial forms are available, it is easy to convert them into D using some form of divergence measures.

2. Although the VAT image suggests both the number of object and reordering matrix produces neither a partition nor a hierarchy of clusters. Hidden structure can be viewed as an illustrative data visualization for estimating the number of clusters before clustering. However, from the reordered matrix hierarchical structure is detected if the diagonal subblocks exist within the larger diagonal blocks.

3. We have a prior knowledge of k value in the previous paper. But, in this algorithm we propose a new model which makes this algorithm much efficient.

### *A. $V^2$VAT Algorithm steps*

**Input**: D = $[d_{ij}]$ : An n × n scaled matrix of pairwise divergence matrix.

(1) Compute a local scale $\sigma_i$ for each $o_i$ as $\sigma_i = d(o_i,o_k) = d_{ik}$,where $o_k$ is nearest neighbour of $o_i$ without user defined k value i.e.,$n(o_i)$ = if(( $\sum_{i=1}^{n}\sum_{j=1}^{n} d(oi, oj)$< ( $\sum_{j=1}^{n}\sum_{k=1}^{n} d(oj, ok)$).

(2) Choose the m indices from {1,2,….,n} randomly to form the sample index set $I_s$ and the set of the remainig object indices $I_r$, which are respectively used to get sub-matrices $D_S$ and $D_B$ from D.

(3) Construct the matrices S$\in R^{m\times m}$ from $D_S$ and B $\in R^{m\times(n-m)}$ from $D_B$ using the weighted Gaussian function $\exp(-d_{ij}d_{ji}/(\sigma_i\sigma_j))$.

(4) Perform eigendecomposition of S and compute the approximate eigenvectors $\widetilde{U}_F$.

(5) Choose the columns of $\widetilde{U}_F$ that will increase the divergence of matrix and using that columns form $V_k \in R^{n\times k}$ and normalize the rows of $V_k$ to unit Euclidean norm to generate $V'_k$.Treat each row of $V'_k$ as a new instance to compute a new pairwise divergence matrix $D'_s \in R^{m\times m}$ between the sample instances to obtain sample SpecVAT images $I(D'_s)$.

(6) Construct the weighting matrix W $\in R^{n\times n}$ by defining $w_{ij} = \exp(-d_{ij}d_{ji}/(\sigma_i\sigma_j))$ for i≠ j, and $w_{ii} = 0$.

(7) Construct the normalized Laplacian matrix $L' = M^{-1/2} W M^{-1/2}$.

(8) Choose the k largest eigenvectors of L' to form the matrix $V = [v_1,\ldots,v_k] \in R^{n\times k}$ by stacking the eigenvectors in columns.

(9) Normalize the rows of V with unit Euclidean norm to generate V'.

(10) For i = 1,2,…,n, let $u_i \in R^k$ be the vector corresponding to the i-th row of V' and treat it as a new instance (corresponding to $o_i$). Then construct a new pairwise divergence matrix D' between instances.

(11) Set I=ø, J={1,2,. . . . , n} and π=(0,0,….,0)
Select (i,j) $\in \arg_{p\in J,q\in J}\max\{d'_{pq}\}$
Set π(1)=i, I←{i} and J←J-{i}

(12) Repeat for t=2,3,…,n
Select (i,j) $\in \arg_{p\in I,q\in J}\min\{d'_{pq}\}$
Set π(t)=j, update I←I ∪ {j} and J←J-{j}

(13) From the reordered matrix $\widetilde{D}'=[d'_{ij}]=[d'_{\pi(i)\pi(j)}]$,for 1 ≤ i,j ≤ n. A scaled gray-scale image $I(\widetilde{D}')$ is obtained.

(14) Compute an optimal threshold $T^*_k$ that can maximize $\sigma_B^2$ for the image $I(\widetilde{D}'_k)$, i.e., $T^*_k$ = arg $\max_{1\le T<L}\sigma_B^2(T)$.

(15) Obtain the corresponding "goodness" measure for each SpecVAT image $GM(k) = \sigma_B^2(T^*_k)$.

(16) Determine the number of clusters as c = arg $\max_k$ GM(k).

(17) For the sample SpecVAT image $I(D'_{k*})$, set the genome of each individual $x_i$ ( i = 1~b ) as a binary string of length n-1, corresponding to the indices of the first n-1 samples.

(18) Randomly set c-1 element in each $x_i$ to '1' and other to '0' to create the initial population.

(19) Set a fitness function as taking the input $x_i$ and calculating the candidate partition U from $x_i$, and returning the result of objective function.

(20) Apply GA algorithm until there is no improvement within g = 10 generation to find the optimum genome $x^*$.

(21) Transform $x^*$ into cluster partition $U^*$(which is equivalent to obtaining the sizes of each cluster $\{n_1,\ldots,n_c\}$). The position $p_1$ of the first '1' in $x^*$ means the firstcluster partition is from sample 1 to $p_1$.

The position $p_j$( j = 2,…,c-1 ) of the j-th '1' means the j-th cluster partition is first sample ($p_{j-1}$ + 1) to $p_j$. The c-th cluster partition is from the sample ($p_{c-1}$ + 1) to n.

(22) Retrieve real objects indices in each cluster $C_i$ with the permutation index π(), i.e., $C_1 = \{o_{\pi(1)},\ldots, o_{\pi(n1)}\}$, and $C_i = \{o_{\pi(ni-1+1)},\ldots, o_{\pi(ni-1+ni)}\}$ for i = 2,…,c. We obtain the data partitioning $\{C_1,\ldots,C_c\}$.

(23) Perform out-of-sample extension using $V'_{k*}$ to obtain the cluster labels of the remaining objects indexed by $I_r$.

**Output:** The number of clusters c and the data partitioning results.

Assume that an n × n symmetric positive semidefinite matrix, say F, can be decomposed as $F = U_F \Sigma_F U^T_F$, where $\Sigma_F$ are the eigenvalues of F and $U_F$ are the associated eigenvectors. Suppose m columns of F are randomly sampled without replacement. Let A be the n × m matrix of these sampled columns, and S be the m × m matrix consisting of the intersection of these m columns with the corresponding m rows. We can rearrange (Eq.1) the columns and rows of F without disturbing the generality such that,

$$F = \begin{pmatrix} S & B \\ B^T & C \end{pmatrix} \quad \text{with} \quad A = \begin{pmatrix} S \\ B^T \end{pmatrix} \tag{1}$$

where $B \in R^{m\times(n-m)}$contains the elements from the sampled objects to the remaining ones, and $C \in R^{(n-m)\times(n-m)}$ contains the elements between all of the remaining objects. In the case of interest, $m \ll n$, S is very small but C is usually large.

As a technique for finding numerical approximations to eigenfunction problems, the Nystrom method has been used for fast Gaussian process classification and regression, low-rank approximation to kernel matrices, large-scale manifold learning and image segmentation. Simply, it uses S and A ( Eq.2) to approximate F as

$$F \approx \tilde{F} = A S^+ A^T \tag{2}$$

where "+" is the pseudoinverse. The Nystrom method implicitly approximates C by $BS^+ B^T$ , and the resulting approximate eigenvalues and eigenvectors of F are (Eq.3)

$$\tilde{\Sigma}_F = \left(\frac{n}{m}\right)\Sigma_S \quad \text{and} \quad \tilde{U}_F = \sqrt{\frac{m}{n}} A U_S \Sigma_S^+ \tag{3}$$

where $S = U_S \Sigma_S U_S^+$. In a more explicit form, the approximated eigenvectors are written as (Eq.4),

$$\tilde{U}_F = \begin{pmatrix} U_S \\ B^T U_S \Sigma_S^+ \end{pmatrix} \tag{4}$$

It can be seen that only A (or S and B) is needed to compute the approximated eigenvectors of F. In this procedure, eigendecomposition on the small sample matrix $S \in R^{m\times m}$ is practical, and multiplication with the matrix B (i.e., $B^T U_S \Sigma_S^+$ ) is also feasible.

The normalized Laplacian matrix [6],

$$\hat{L} = M^{-1/2} (M - W) M^{-1/2} = I - L' \tag{5}$$

where, I is the identity matrix. Replacing I - L' with L' in our algorithm only changes the Eigen values from $1-\lambda_i$ to $\lambda_i$and not the eigenvectors (Eq.5).

Apply the A-SpecVAT and P-SpecVAT algorithms to a sample set for perform visual cluster analysis on a large data set and then extend the sample clustering result to obtain (approximate) clusters for the remaining objects. After obtaining the approximate eigenvectors of F using the Nystrom method [20], we treat each row of $\tilde{U}_F$as a new "instance" in the spectral space (implicitly corresponding to an original object in O). We choose the new instances corresponding to the choosen samples to obtain the sample SpecVAT images. These SpecVAT images are used to determine the number of clusters c by using A-SpecVAT, and then apply P-SpecVAT on the best sample SpecVAT image to obtain the data partitioning, i.e., obtaining the class labels of the sampled objects $\{o_1^s, o_{2,\ldots,}^s o_m^s\}$, say $\{l_1,l_2,\ldots,l_m\}$, where $l_i = \{1,\ldots, c\}$. The set of all non degenerate c-partition matrices for O is (Eq.6)

$$H_{hcn} = \{U \in R^{c\times n} \mid 0 \le u_{ik} \le 1, \forall i, k\} \tag{6}$$

$$\text{With} \quad \sum_{i=1}^{c} u_{i,k} = 1, \forall k \text{ and } \sum_{k=1}^{n} u_{i,k} > 0, \forall i$$

The objective function is defined as (Eq.7)

$$E(U,\tilde{D}') = E_b - E_w \tag{7}$$

A good U should maximize this objective function (Eq.8),

$$U^* = \arg\max E(U,\tilde{D}') \text{ where } U \in H^*_{hcn} \tag{8}$$

Accordingly, we select the best SpecVAT image as the one with the maximum goodness value (Eq.9) and calculate the number of clusters as

$$c = \arg\max \sigma_B^2(T^*, k) \tag{9}$$

In addition, this visual partitioning procedure can also operate on the VAT image or other reordered divergence images. Finally we got number of clusters(c) and data partitioning.

## V. EXPERIMENTAL RESULTS

To evaluate the proposed algorithm, we have to perform certain experiments by using unlabeled data sets. In this section we compare the existing VAT algorithm with proposed $V^2$VAT algorithm to evaluate the accuracy, precision and recall. We have carried out a data sets for a experiments. The data sets are wine, iris and vehicle from the UCI Machine Learning Repository [25]. The number of cluster for this data sets are calculated as per steps defined in VAT algorithm and $V^2$VAT algorithm.

TABLE I
C value for different data set using VAT and $V^2$VAT

| Data set | Wine | Iris | Vehicle |
|---|---|---|---|
| Number of instances | 178 | 150 | 946 |
| Number of attributes | 13 | 4 | 18 |
| Actual Number of clusters | 3 | 3 | 4 |
| C value in VAT | 2 | 4 | 5 |
| C value in $V^2$VAT | 3 | 3 | 4 |

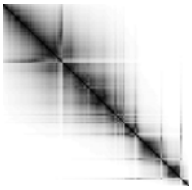

Fig.2. VAT image for wine

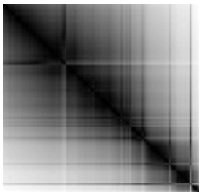

Fig.3. $V^2$VAT image for wine

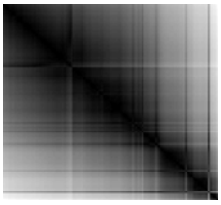

Fig.4. VAT image for iris

Fig.5. $V^2$VAT image for iris

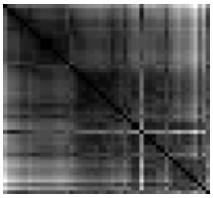

Fig.6. VAT image for vehicle

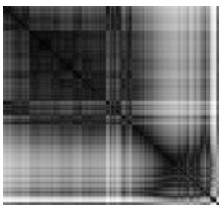

Fig.7. $V^2$VAT image for vehicle

From that result we obtain VAT and $V^2$VAT image for each data set. This images are compared in this section. The output image of the VAT algorithm and $V^2$VAT algorithm are given below (Fig.2 to Fig.7). By these VAT image for data set are not clear. So it is difficult to determine the number of clusters exactly. Viewers may detect different estimate of C value. But in the image of $V^2$VAT we clearly detect the number of cluster(c). In the table.1 give the difference of C value for $V^2$VAT from VAT algorithm. From the comparison we can identify the proposed paper is more efficient than the existing paper.

### *A. Histogram of divergence matrix in $V^2$VAT algorithm*

Histogram is a graphical representation of data distribution means that it gives the probability distribution of data. It is also used to estimate the probability density. In structure, Histogram is similar to the bar chart. In this place we can view the histogram of divergence matrix of data sets (Fig.8 to Fig.10). For each data set showed determine the divergence matrix. Divergence matrix is the basic input for proposed $V^2$VAT algorithm.

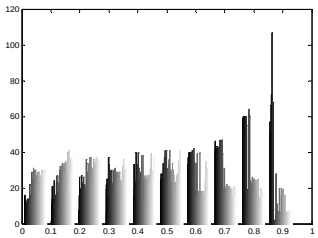


Fig.8. Wine dataset

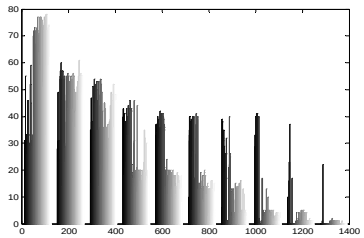


Fig.9. Iris dataset

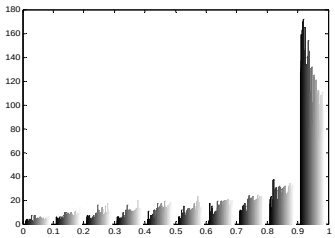


Fig.10. Vehicle dataset

Let the pixels of an image be represented in L gray levels . The number of pixels at level l is denoted by $m_l$ and the total number of pixels by $N = \sum_{l=1}^{L} m_l$. Such a gray-level histogram may be normalized and regarded as a probability distribution, i.e $p_l = m_l/N$ , $p_l > 0$ , $\sum_{l=1}^{L} p_l = 1$. In the above histogram are shown in (Fig.8 to Fig10) the distribution of data for wine, iris and vehicle datasets.

*B. Accuracy comparison*

In this section performance is evaluated in terms of accuracy. In this graph Fig.11. We have taken two parameters called accuracy and datasets. It helps to analyze the existing VAT and proposed $V^2$VAT algorithm. Accuracy can be calculated from formula given below.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

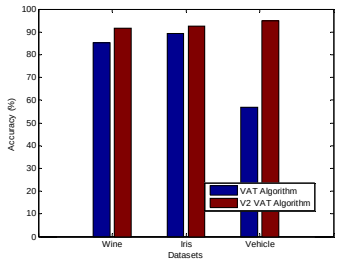


Fig.11. Accuracy comparison

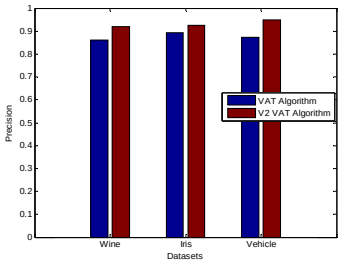


Fig.12. Precision comparison

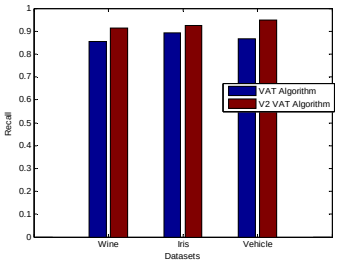


Fig.13. Recall comparison

The accuracy is shown in Fig.11 as the parameter is noted as Y axis and the number of dataset is noted as X axis. The blue line represents the existing system and the red line represents the proposed system. From this chart we identify that the accuracy of the proposed system is higher than the existing system. From this we can easily understand the proposed system is more efficient than the exiting one.

*C. Precision comparison*

In this section the performance is evaluated in terms of precision. The following chart gives the precision comparison between the existing and proposed. It can be defined as,

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

The Fig.12 the chart shows that X-axis will be three data sets and Y-axis is the precision parameter. Even in the three data sets also our proposed system has more precision comparing to existing VAT algorithm. From this graph, proposed concept is better in precision parameter.

*D. Recall comparison*

In this section, we compare the recall parameter between VAT and proposed $V^2$VAT algorithm. Recall means information retrieval. It is mathematically calculated by using the formula as,

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

In the Fig.13 chart shows that X-axis is the datasets and Y-axis is the recall rate. In view of this recall comparison graph we can conclude as the proposed algorithm is effective in recall performance comparing to theexisting VAT algorithm.

## VI. CONCLUSION AND FUTURE WORK

In this paper the VAT algorithm are developed to $V^2$VAT algorithm. $V^2$VAT algorithm is more effective in many cases. This new approach is an efficient solution to overcome the problems in existing VAT algorithm. The direct visual validation method is proposed to be advantageous for this paper. The new method obtains the k value in simple representation and easy to understand. The divergence matrix used instead of dissimilarity matrix where increases the efficiency of the algorithm. A series of experiments on different data set have demonstrated that our algorithm performs well in terms of both number of clusters and data partitioning. In future we will use higher order statistical method like linear moment (summarize the shape of probability distribution), pareto index (specifying the pareto distribution) and nonparametric skew (measures a skewness of random variable's distribution) which is suitable for the complex data types.

## REFERENCES

[1] J.C. Bezdek and R.J. Hathaway, "VAT: A Tool for Visual Assessment of Cluster Tendency," Proc. International Joint Conference Neural Networks, pp. 2225-2230, 2002.

[2] Mehmet Can, "Fuzzy Clustering Models and Algorithms for Pattern Recognition", SadinaGagula-Palalic.,Sarajevo, 2008.
[3] Anil K.Jain, Richard C Dubes, "Algorithm for Clustering data" QA278.J35,1988.
[4] R. Ling, "A Computer Generated Aid for Cluster Analysis," Comm. ACM, vol. 16, pp. 355-361, 1973.
[5] R. Hathaway, J.C. Bezdek, and J. Huband, "Scalable Visual Assessment of Cluster Tendency," Pattern Recognition, vol. 39, pp. 1315-1324, 2006.
[6] F. Chung, "Spectral Graph Theory",Handbook of Linear Algebra, second edition, CCR Press, vol. 92. Am. Math. Soc., 1997.
[7] X. Hu and L. Xu, "A Comparative Study of Several Cluster Number Selection Criteria," Intelligent Data Engineering and Automated Learning, pp. 195-202, Springer, 2003.
[8] J. Huband, J.C. Bezdek, and R. Hathaway, "Bigvat: Visual Assessment of Cluster Tendency for Large Data Sets," Pattern Recognition, vol. 38, no.11, pp. 1875-1886, 2005.
[9] U. Maulik and S. Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1650- 1654, Dec.2002.
[10] James C. Bezdek, Fellow, Richard J. Hathaway, and Jacalyn M. Huband"Visual Assessment of Clustering Tendency for Rectangular Dissimilarity Matrices",IEEE transactions on fuzzy systems, vol. 15, no. 5, october 2007.
[11] L. Zelnik-Manor and P. Perona, "Self-Tuning Spectral Clustering," Advances in Neural Information Processing Systems, MIT Press, 2004.
[12] J.C. Dunn, "Indices of Partition Fuzziness and the Detection of Clusters in Large Sets," Fuzzy Automata and Decision Processes, Elsevier, 1976.
[13] P.J. Rousseeuw, "A Graphical Aid to the Interpretations and Validation of Cluster Analysis," J. Computational and Applied Math., vol. 20, pp. 53-65, 1987.
[14] T. Tran-Luu, "Mathematical Concepts and Novel Heuristic Methods for Data Clustering and Visualization," PhD thesis, Univ. of Maryland, 1996.
[15] R.B. Calinski and J. Harabasz, "A Dendrite Method for Cluster Analysis," Comm. Statistics, vol. 3, pp.1-27, 1974.
[16] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data-Set via the Gap Statistics," J. Royal Statistical Soc. B, vol. 63, pp. 411-423, 2001.
[17] I. Sledge, J. Huband, and J.C. Bezdek, " Automatic Cluster Count Extraction from Unlabeled Data-Sets," Proc. Joint Fourth International Conf. Natural Computation (ICNC) and Fifth International Conf. Fuzzy Systems and Knowledge Discovery (FSKD), 2008.
[18] J.C. Bezdek and N.R. Pal, "Some New Indices of Cluster Validity," IEEE Trans. System, Man and Cybernetics, vol. 28, no. 3, pp. 301-315, June 1998.
[19] O.Maimon and L. Rokach,eds. "Decomposition Methodology for Knowledge Discovery and Data Mining", pp. 90-94.World Scientific, 2005.
[20] C. Williams and M. Seeger, "Using the Nystrcom Method to Speed up Kernel Machines," Advances in Neural Information Processing Systems, pp. 682-688, MIT Press, 2000.
[21] L. Wang, C. Leckie, J. Bezdek, and R. Kotagiri, "Automatically Determining the Number of Clusters in Unlabeled Data Sets," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 3, pp. 335-350, Mar. 2009.
[22] SrinivasuluAsadi, Ch D V SubbaRao, V Saikrishna, "Finding the Number of Clusters in Unlabeled Datasets using Extended Dark Block Extraction",International Journal of Computer Applications (0975 – 8887)Volume 7– No.3, September, 2010.
[23] Liang Wang, Senior Member, Xin Geng, James Bezdek, Christopher Leckie, Fellow and Kotagiri Ramamohanarao, "Enhanced Visual Analysis for Cluster Tendency Assessment and Data Partitioning", IEEE transactions on knowledge and data engineering, vol. 22, no.10, october 2010.
[24] I. Dhillon, D. Modha, and W. Spangler, "Visualizing Class Structure of Multidimensional Data", Proc. 30th Symp. Interface: Computing Science and Statistics, 1998.
[25] UCI Repository of Machine Learning Databases,http://www.ics.uci.edu/~mlearn/MLRepository.html, 2008.