International Journal of Management, IT & Engineering
(ISSN: 2249-0558)

CONTENTS

Editorial Board

Editorial Advisory Board

Title

THE PROBLEM OF HIGH DIMENSIONALITY WITH LOW DENSITY IN CLUSTERING

Author(s)

Prof. T. Sudha
Research Supervisor

Swapna Sree Reddy. Obili
Ph.D Research Scholar,
Sri Padmavathi Women's University,
Tirupati.

ABSTRACT:

In many real-world applications, there are a number of dimensions having large variations in a dataset. The dimensions of the large variations scatter the cluster and confuse the distance between two samples in a dataset. This degrades the performances of many existing algorithms. This problem can be happened even when the number of dimensions of a dataset is small. Moreover, no existing method can distinguish whether the dataset has the highly repeated problem or low-density's problem. The only way to distinguish the problem is by a prior knowledge, which is given by the user.

There are many methods to resolve this type of high dimensionality problem. The common way is to prune the non-significant features so that the features having large variations are removed and high-density cluster centers are obtained. Much research work has been carried out based on this criterion. *The subspace clustering method* is one of the well-known tools. The feature space is first partitioned into a number of equal length grids. Then, the density of each interval is measured. The features having low density are discarded and the clustering is conducted on the high density regions. Although these methods work very well on synthetic datasets, the pruned dimensions can carry useful information and hence, pruning them may increase the classification error rates.

1 Introduction:

In this paper, a new clustering algorithm is developed to handle this problem. Here, we introduce a new concept called sub-dimension. The key concept is to measure the similarity between two objects in several sub-dimensions. Here, we introduce a new concept called sub-dimension. A dataset is separated into p parts, which are not disjoint. Each part has the same number of input samples as the original data, but a smaller number of dimensions. In our formulation, each part has the same number of dimensions and we call each of these dimensions a sub-dimension. If more than half the features of two objects belong to the same group, these two objects are said to belong to the same group. For example, assume that x_1, x_2 and x_3 are 10 dimensional data vectors. The data point x_3 is said to be closer to x_1 than x_2 if more than half of the dimensions of x_1 and x_3 are closer to x_1 than x_2. Thus, if two patterns are very similar except

for a small number of features, this measure will preserve the similarity. Experiment results show that the clustering algorithm using this method gives better results than other methods.

The organization of this paper is as follows. In Section 2, we introduce the low-density problem and the basic concept of our new similarity measure. After that, in Section 3, we introduce the new clustering method. Experiment results are given in Section 4. Conclusions are given in Section 5.

2 Problem Statement:

In this section, we indicate the problem of existing similarity to the large variation dimension. Then, we introduce our method. We now consider three 10- dimension data points x_1 = [0.5878, 0.9511, 0.9511, 0.5878, 0.0000, -0.5878, -0.9511, -0.9511, -0.5878, - 0.0000]^T, x_2 = [4.9550, 3.1490, 3.3364, 2.0429, 0.0620, 0.9109, -0.6682, 3.7762, 3.3466, 4.2073]^T and x_3 = [0.5878, 0.9511, 0.9511, 0.5878, 0.0000, -0.5878, -0.9511, -0.9511, -0.5878, 10.0000]^T. These three patterns are shown in Figure 2.1. The one at the bottom is x_1 while the one above x_1 is x_2. The third vector x_3 (marked by •) is almost the same as x_1 except for the 10th dimension, which is far away from the x_1. The variation in the 10th dimension of x_3 is due to the presence of a noisy non-significant condition. The shapes of x_1 and x_3 are almost the same and they almost certainly belong to the same group. However, if we measure their similarity using the l_2 norm, we will find that x_1 and x_2 are more likely to be in the same group. The l_2 norm distance between x_1 and x_2 is $||x_1-x_2||$ = 9.4641 while the l_2 norm distance between x_1 and x_3 is $||x_1-x_3||$=10. Because of this, a clustering algorithm may produce unreliable results.

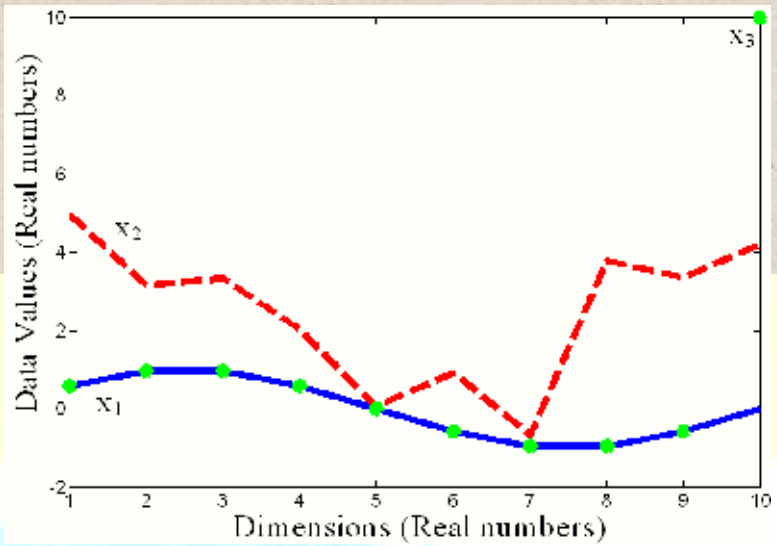


Figure 2.1. Illustration of the effect of a non-significant features on the similarity measure.

Now, we introduce our method by reformulating the distance measure as follows. Let $X = A_1 \times A_2 \times \ldots A_{d-1} \times A_d$, where A_j represents the j-th dimension of the data with $1 \leq j \leq d$. We re-define the dimension of X as follows: $X = B1 \times B2 \times \ldots B_{p-1} \times B_p$, where $p \leq d$, and $B_j = A_{j1} \times A_{j2} \times \ldots A_{j(s-1)} \times A_{js}$, where $1 \leq j \leq p$, $1 \leq j_1, j_2, \ldots, j_s \leq d$ and s is the number of features in each sub-dimension and $s \leq d$. Here the original dataset X is represented by d non-overlapping subsets A_1, A_2, … and A_d, where A_j simply represents the set of data values of all samples along dimension j. To work with the sub-dimensions, we decompose X into many overlapping subsets B_1, B_2, … and B_p, where B_j is the union of s subsets A_{j1}, A_{j2}, … and A_{js}. An input data vector x_i is now decomposed into p vectors, $x_i(B_j)$. In the sub-dimension based similarity measure, x_a is closer to x_b than to x_c if

$$\text{Card}\quad j : \left\| x_{a\,Bj} - x_{b\,Bj} \right\| < \left\| x_{a\,Bj} - x_{c\,Bj} \right\| \quad > \text{Card}\quad j : \left\| x_{a\,Bj} - x_{b\,Bj} \right\| \geq \left\| x_{a\,Bj}\, x_{c\,Bj} \right\|$$

where card(S) refers to the cardinality (or the number of elements) of the set S. The above equation defines a new similarity measure between two objects, under which x_a is closer to x_b than to x_c if $x_{a(Bj)}$ is closer to $x_{b(Bj)}$ than to $x_{c(Bj)}$ in more sub-dimensions. The above classification criterion can also be written as

$$\frac{\text{Card}\quad j:\left\|x_{a\ Bj}-x_{b\ Bj}\right\|<\left\|x_{a\ Bj}-x_{c\ Bj}\right\|}{p}>\frac{1}{2} \qquad (2.1)$$

since

$$\text{Card}\quad j:\left\|x_{a\ Bj}-x_{b\ Bj}\right\|<\left\|x_{a\ Bj}-x_{c\ Bj}\right\| \quad+\text{Card}\quad j:\left\|x_{a\ Bj}-x_{b\ Bj}\right\|\geq\left\|x_{a\ Bj}-x_{c\ Bj}\right\|$$

= the number of sub-dimension vector sets = *p*

This means that x_a is classified into the class of x_b if for more than 50% of sub-dimensions x_a is closer to x_b than to x_c, otherwise it is classified into the class of x_c. To have a more reliable classification, we can require this ratio to be greater than 50%, for example, we can set it to 60%. However, in doing so, we may have to reject x_a, that is, we cannot make a decision with enough confidence, if the ratio is between 50% and 60%. In practical applications, we can adjust this ratio to trade off between false positive and rejection rates in a pattern classification system.

Now, we apply this concept to the three patterns x_1, x_2 and x_3. We first decompose these data vectors into sub-dimensional ones. For example, we can decompose x_1 into 8 sub-dimensional vectors, each of which has three dimensions,

$\left[x_1^{\,1},x_1^{\,2},x_1^{\,3}\right]^T,\left[x_1^{\,2},x_1^{\,3},x_1^{\,4}\right]^T,\ldots,\left[x_1^{\,8},x_1^{\,9},x_1^{\,10}\right]^T$. Then we measure the similarity between all corresponding sub-dimensional vectors using the l_2 norm. After this calculation, we say that objects x_1 and x_3 are closer than x_1 and x_2 if more sub-dimensional vectors between x_1 and x_3 suggest they are closer. Obviously, for the three patterns x_1, x_2 and x_3, all the sub-

dimensional vectors between x_1 and x_3 give a smaller value than x_1 and x_2 except the last one, which contains the 8th, 9th and 10th dimensions. Thus, we say that x_1 and x_3 are closer than x_1 and x_2.

3 The New Clustering Algorithm:

In this section, we introduce the new clustering method. There are a total of five steps in our method with the number of group ctotal given by the user. These steps are given as follows.

Step 1: Let $X=\left[x_i^{1}, x_i^{2}, ..., x_i^{d-1}, x_i^{d}\right]^T$ (for $1\leq i\leq n$) be the original dataset sorted in ascending order with respect to the standard derivation of each dimension. This dataset is divided into lower and upper half: $G^1 = [x_i^1, x_i^2, \dots, x_i^{d/2}]^T$ and $G^2 = [x_i^d, x_i^{d-1}, \dots, x_i^{d/2+1}]^T$. Then, G^1 and G^2 are mixed in an alternative manner: $G_s = [x_i^d, x_i^{d/2}, x_i^{d-1}, x_i^{d/2-1} \dots, x_i^{d/2+1}, x_i^1]^T$. If the difference in the standard derivation between two consecutive dimensions of G_s is larger than a threshold θ (which is taken as 2 in all the experiments), we will take $G_s = X$. We take G_s as a sorted dataset in the second step of the method.

Step 2: We divide the datasets $X \in R^d$ into several sub-dimensional sets which have smaller dimensions: $X_{(j)}=\{x_{i(j)}\}$ where $x_{i(j)}\in R^s$, j is the jth sub-dimension of the data $1\leq j\leq p$ and $s\leq d$. In this thesis, s=2 and s=3 are adopted. For example, x_i has 10 dimensions, its R^2 and R^3 sub-dimensional sets will be $[x_i^1,x_i^2]^T$, $[x_i^2,x_i^3]^T$,… $[x_i^9,x_i^{10}]^T$, and $[x_i^1, x_i^2,x_i^3]^T$, $[x_i^2,x_i^3,x_i^4]^T$, …, $[x_i^8,x_i^9,x_i^{10}]^T$ respectively.

Step 3: We apply the FCM algorithm to each of the sub-dimensional sets with the input parameter from $c = 2$ to $c = c_{total}$ and evaluate the clustering results. This is equivalent to conducting the cluster validity on each sub-dimensional set. For each sub-dimensional set, only the clustering result with the largest $I_{mod}(c)$ is considered. The original I-index can be found in the paper [Maulik and Bandyopadhyay 2002].

$I_{mod}(c)$ is a modified version of the I-index. The equation for $I_{mod}(c)$ is given as follows:

$$I_{mod}\ c = \left(\frac{1}{c} \times \frac{E_1}{E_c} \times D_c \right)^Q \qquad (3.1)$$

where the power Q is used to control the contrast between different cluster configurations. In this thesis, we take Q = 1. E_c and D_c are defined as

$$E_c = \sum_{i=1}^{n} \sum_{k=1}^{c} \delta_{ik} \left\| x_i - v_k \right\|^2 \qquad (3.2)$$

$$D_c = \max_{i,j=1}^{c} \left\| v_i - v_j \right\| \qquad (3.3)$$

where v_k is the prototype of class *k* generated by the clustering algorithm. δ_{ik} is a binary variable. If x_i is a data point closest to v_k, δ_{ik} =1. Otherwise, δ_{ik} =0. The difference between the proposed modified I-index and the original I-index is that the function E_c has a square power in the modified I-index while having no square power in the original I-index.

Step 4: Based on the results in Step 3, we are able to get a partition matrix P^j_s. This partition matrix is a binary matrix. If the sub-dimensional points $x_{p(j)}$ and $x_{q(j)}$ belong to the same group, P^j_s (p,q)=1. Otherwise, P^j_s (p,q)=0. Now, we define the variable P_s as the mean of these partition matrices.

$$P_s = 1 - \frac{1}{d-s} \sum_{j=1}^{d-s+1} P_s^{\,j}. \qquad (3.4)$$

As we adopt s=2 and s=3 for sub-dimensional sets, there are in total two variables P_2 and P_3. The average partition matrix P is defined as $P = (P_2+P_3)/2$. Thus, if there are two data points x_p and x_q and their conditions are very similar to each other, the value P(p,q) will be small.

Step 5: We consider the average partition matrix P as a similarity matrix in a hierarchical clustering algorithm. We adopt the complete link method in the hierarchical clustering algorithm to get the clustering result. Table 3.1 summarizes these five steps.

Algorithm:

1. The dataset is sorted according to its dimensions.
   For s = 2 to 3, perform steps 2 to 4:
2. The dataset X is divided into several sub-dimensional sets, each with dimension s.
3. The FCM algorithm is applied to each sub-dimensional set with the input parameter *c varied from c=2 to* $c=c_{total}$*. Then, in each sub-dimensional set, the clustering results* with largest $I_m(c)$ value will be considered.
4. By making use of these clustering results, the matrix P_s can be obtained.
5. The average partition matrix $P=(P_2+P_3)/2$ is computed. By taking P as a distance matrix, the hierarchical clustering algorithm is applied to produce the final result.

Table 3.1. Summary of our clustering algorithm.

4 Experiment Results:

In this section, we conduct eight experiments to test the robustness of our method. Four different clustering algorithms are chosen to compare with the performance of our method. They are the GMM, the FCM algorithm, the hierarchical clustering method with complete link (HC) and the HARP algorithm, which is a subspace clustering method.

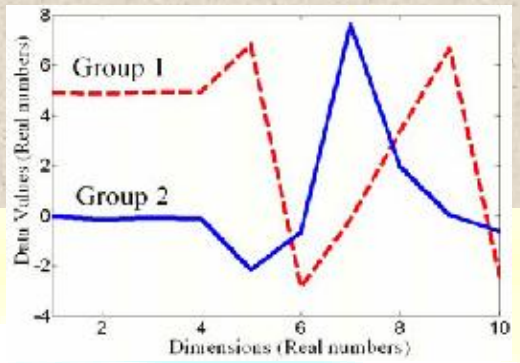
Each algorithm except HC will be performed 10 times to each real world dataset. In this paper, the data samples we adopt from the real world datasets have class labels. We make use of these labels for evaluating the algorithms. For example, after applying the FCM algorithm, we obtain c partitions $C_1,\ldots C_c$. In each original group, we find the number of objects correctly recognized in $C_1,\ldots C_c$ so that the sum of these numbers reaches maximum. Based on the number

of correctly classified objects, we will compare the algorithms in three ways. They are the maximum, mean and standard derivation of the number of correctly recognized objects in the 10 runs. Also, we will show the number of correctly recognized objects in each group for the best clustering result among 10 runs.

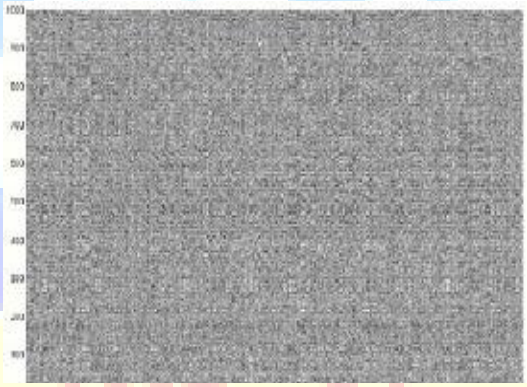
4.1 Synthetic Dataset

We perform three experiments based on three synthetic datasets. In each of these datasets, some of the dimensions have very large variance compared with other dimensions. This situation is similar to the one we introduced in Section 2. This can make the conventional distance measure error prone.

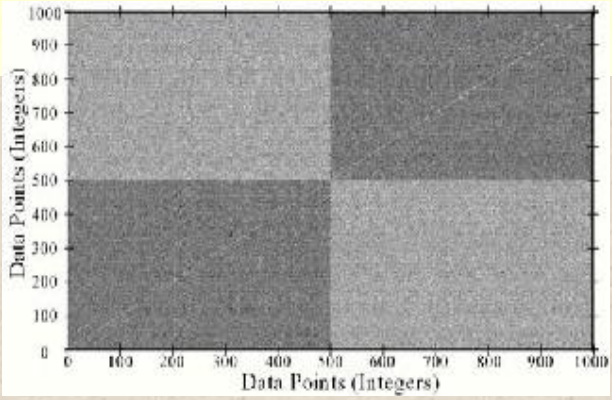
Example A Synthetic Dataset 1: We consider a ten dimensional dataset with two groups. The first four dimensions are the same and they are generated by two normally distributed functions $N(0,1)$ and $N(5,1)$. Each of them consists of 500 points. The last six dimensions are generated by a normally distributed function which is $N(0,100)$ with 1000 points. Thus, the data matrix consists of a size of 1000×10. In this example, there are six components, which are non-significant conditions having large variations for the two groups, while there are four components, which contain information of the two groups. Figure 4.1(a) shows mean values of the two groups. The two groups can be clearly separated in terms of the first four dimensions but they cannot in terms of the last six dimensions. If we apply the FCM algorithm to the first four dimensions of the dataset, 100% accuracy will be obtained. However, the insertion of non-significant information from the extra six dimensions degrades the clustering result significantly. The clustering results for this dataset are given in Table 4.1. We can see that the GMM, FCM and HC clustering algorithms obtain only half the accuracy rate. The HARP algorithm has a much higher accuracy than the non- subspace clustering algorithm. As the higher dimensions are pruned in HARP, the subspace clustering algorithm has a better performance than non-subspace clustering algorithm. Our method obtains a 100% accuracy rate. One may think that the clustering result of our method may be unreliable since the total number of dimensions that do not contain the information of the two groups is more than the total number of dimensions that contain the information of the two groups.



(a) The mean values of the two groups for each of the 10 dimensions.



(b) P2 for the last six dimensions of the dataset, which is a random matrix.



(c) The partition matrix for Synthetic Dataset1.

Figure 4.1. The mean values of the synthetic dataset 1 and its partition matrices.

However, in our method, we divide the dataset into several sub-dimensional sets and conduct data clustering for each of them. In the third and fourth steps of our method, we found that only the first four dimensions shared the same clustering results. However, the last six dimensions produce very different clustering results in each sub- dimensional set. Thus, the variable P_2 for the last six dimensions is just a random matrix and does not contribute much to the average partition matrix P. The matrix P_2 is given in Figure 4.1(b). The darker pixels represent larger values in the partition matrix and vice versa. Figure 4.1(c) shows the average partition matrix P for synthetic data '1' after permutating the matrix P so that the first group consists of the first 500 elements. We can clearly see that our method can detect 500 points in one group and another 500 points in another group.

Groups	GMM	FCM	HC	HARP	Our method
1	297	255	421	478	500
2	206	249	99	492	500
Total(Max)	503	504	520	970	1000

Table 4.1. Clustering results for Synthetic Dataset 1.

Example B Synthetic Dataset 2: Now, we consider a twenty-four dimensional dataset with three groups. The first four dimensions are generated by three normally distributed functions $N(0,1)$, $N(5,1)$ and $N(10,1)$. Each of these consists of 500 points. The last 21 dimensions are generated by the normally distributed function with a difference variance from 5 in the fifth dimension to 100 in the twenty-fourth dimension and the variances between consecutive two dimensions have a difference of 5. If we put the variances from the fifth to

twenty fourth dimensions in a vector, it will become $[5, 10, 15, 20, \_, 100]^T$. Thus, the data matrix consists of a size of 1500×24. This synthetic dataset is different from the previous one. The elements of the Synthetic Dataset 1, which are non-significant conditions, have exactly the same variances. However, in this dataset, the variances are not the same but are monotonic increasing. The clustering results for this dataset are given in Table 4.2. Again, we can see that the GMM, FCM and HC algorithms obtain only half the accuracy rate. For the HARP algorithm, the result is not as good as the one given in Synthetic Dataset 1 (Table 4.1). Its accuracy is reduced to around 75%. This shows that the HARP algorithm could not prune the noise dimensions well if they are very different. Our method obtains a 99% accuracy rate. In this experiment, we can see that the new technique is able to yield more accurate results than conventional methods.

Groups	GMM	FCM	HC	HARP	Our method
1	200	236	109	402	494
2	209	192	153	498	497
3	134	113	155	205	494
Total(Max)	534	541	517	1105	1485

Table 4.2. Clustering results for Synthetic Dataset 2.

Example C Synthetic Dataset 3: Now, we consider a ten dimensional dataset with three groups. The first four dimensions are generated by three normally distributed functions $N(0,1)$, $N(5,1)$ and $N(10,1)$. Each of them consists of 500 points. The last six dimensions are generated by the normally distributed function with two different variances. The fifth to eighth dimensions are generated by a normal distribution function with variance 10 while the ninth to tenth dimensions are generated by a normal distribution function with variance 10000. Thus, the data matrix consists of a size of 1500×10. The clustering results for this dataset are given in Table 4.3. Similar to Synthetic Dataset 2, the GMM, FCM and HC clustering algorithms have only half the accuracy rate. The HARP algorithm has an 83% accuracy rate. Our method obtains 100%

accuracy. In these experiments, we can see that our method is able to yield more accurate results although both subspace and traditional clustering algorithms cannot.

Groups	GMM	FCM	HC	HARP	Our method
1	200	198	117	471	500
2	209	173	323	470	500
3	134	155	66	295	500
Total(Max)	534	526	506	1236	1500

Table 4.3. Clustering results for Synthetic Dataset 3.

4.2 Real World Data

In this section, the robustness of our method is shown by real world datasets. Five real world datasets are used. They are; iris, wdbc, wine, yeast cell cycle and sporulation datasets. The information for the first three datasets can be found in the following table 4.4.

Name	Full Name	No. of features	Total no. of samples	No. of groups	Normalization
Iris	Iris Plant Database	4	150	3	Yes
Wbcd	Wisconsin Breast Cancer Databases	9	683	2	No
Wine	Wine Recognition	3	178	3	Yes

Table 4.4. Information of datasets.

The last two datasets are microarray datasets and their information is given as follows:

Yeast cell cycle data: This dataset was published by Cho *et al* [Cho *et al*. 1998]. It consisted of 6220 genes with 17 time points taken at 10 minute intervals. In the study of Yeung *et al*., a subset of 384 genes is adopted [Yeung *et al*. 2001]. This subset of datasets can be found at the website (http://faculty.washington.edu/kayee/model/). We normalize each gene expression profile with zero mean and unit variance. This dataset has five cycle phases. They are; early G1 phase, late G1 phase, S phase, S2 phase and M phase.

Sporulation data: This dataset consists of 6118 genes and can found at the website (http://cmgm.stanford.edu/pbrown/sporulation). We only take the genes with the value of root mean square of the log2 transformed data greater than 1.13. After the pre-processing, we get a subset of the data, which contains 1136 genes of the following seven phases: rapid transient induction (‘metabolic’), early I induction, early II induction, early-middle induction, middle induction, mid-late induction and late induction.

The clustering results for these five datasets are given in Tables 4.5, 4.6, 4.7, 4.8 and 4.9. Except for the wine data, the HARP algorithm does not yield better results than the three clustering algorithms (GMM, FCM and HC). Our method yields the largest numbers of correctly classified objects in all cases. Also, the standard derivation of our method is very small. This implies that our method is very stable and able to yield robust solutions.

Groups	GMM	FCM	HC	HARP	Our method
1	50	50	50	49	50
2	40	47	49	38	48
3	49	37	27	12	43
Total(Max)	139	134	126	99	141
Mean	124.9	134	/	99	141
Std	19.1512	0	/	0	0

Table 4.5. Clustering results for the iris data.

Groups	GMM	FCM	HC	HARP	Our method
1	57	50	56	55	58
2	48	45	473	40	62
3	24	27	21	44	45
Total(Max)	129	122	120	139	165
Mean	124.7	122	/	139	165
Std	3.0569	0	/	0	0

Table 4.6. Clustering results for the wine data.

Groups	GMM	FCM	HC	HARP	Our method
1	351	356	357	354	344
2	155	130	20	67	192
Total(Max)	506	486	377	421	536
Mean	484.1	486	/	421	516.9
Std	23.8768	0	/	0	13.4367

Table 4.7. Clustering results for the wdbc data.

Phases	GMM	FCM	HC	HARP	Our method
Early G1	60	50	48	41	48
Late G1	115	67	112	116	120
S phase	31	10	1	23	28
G2	24	38	46	31	35
M Phase	22	51	49	32	52

Total (Max)	252	216	256	243	283
Mean	199.9	216	/	243	282.3
Std	29.6403	0	/	0	0.9487

Table 4.8. Clustering results for the yeast cell cycle data.

	GMM	FCM	HC	HARP	Our method
Metabolic	4	1	1	0	3
Early G1	172	172	244	241	173
Early G2	24	7	4	0	38
Early middle	95	66	8	5	43
Middle	29	32	21	9	60
Mid-late	0	2	2	1	3
Late	0	0	0	0	0
Total (Max)	324	280	280	256	353
Mean	306.9	280	/	256	340.4
Std	13.4449	0	/	0	21.8744

Table 4.9. Clustering results for the sporulation data.

5 Conclusion:

In this paper, we have introduced a new similarity measure to resolve the problem of high dimensionality with low-density. The new algorithm does not prune any dimension in the dataset. The key concept of this algorithm is to measure the similarity between two samples in a number of sub-dimensions. Such a similarity measure reduces the effects of noise in the data. We

have performed eight experiments to test the robustness of the method including three synthetic datasets, three real world datasets and two microarray datasets. We have also compared our method with four different clustering algorithms. Experiment results show that our method yields better results than existing clustering algorithms.

REFERENCES:

- [Agrawal 1981] - H. Agrawal, "Extreme Self-Organization in Networks Constructed from Gene Expression Data," *Physical Review Letters*, Vol. 89, pp. 268702 [4 pages], 2002.
- [Agrawal *et al*. 1998] - R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications," *In Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 94-105. ACM Press, 1998.
- [Ball and Hall 1965] - G. H. Ball and D. J. Hall, "A Clustering Technique for Summarizing Multivariate Data_, *Behav. Sci.*, Vol. 12, pp. 153-155, 1967.
- [Banfield and Raftery 1993] - J. D. Banfield and A. E. Raftery, "Model-based Gaussian and Non-Gaussian Clustering," *Biometrics*, Vol. 49, pp. 803-821, 1993.
- [Barbara *et al*. 2002] - D. Barbara, Y. Li, and J. Couto. Coolcat, "An Entropy Based Algorithm for Categorical Clustering," *In Proceedings of the Eleventh International Conference on In Formation and Knowledge Management*, ACM Press, pp. 582-589, 2002.
- [Barnett and Lewis 1994] - V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd edition, John Wiley, 1994.
- [Brown 1983] - B. M. Brown, "Statistical Use of Spatial Median," *Journal of the Royal Statistical Society, Series B*, Vol. 45, pp.25-30, 1983.
- [Burden and Faires 1997] - R. Burden and J. Faires, *Numerical Analysis*, 6th edition, Pacific Grove, Calif.: Brooks/Cole Pub, 1997.
- [Castleman 1996] - K. Castleman, *Digital Image Processing*, Englewood Cliffs, N. J. : Prentice Hall, 1996.

- [Chakraborty *et al*. 1998] - B. Chakraborty, P. Chaudhuri and H. Oja, "Operating Transformation Retransformation on Spatial Median and Angle Test," *Statistica Sinica*, Vol. 8, pp.767-784, 1998.
- [Chintalapudi and Kam 1998b] - K. K. Chintalapudi and M. Kam, _ The Credibilistic Fuzzy C-means Clustering Algorithm, " *In Proc. IEEE Int. Conf. Systems Man Cybernetics*, pp. 2034–2040, 1998.
- [Chu *et al*. 1998] - S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. Brown, I. Herskowitz, "The Transcriptional Program of Sporulation in Budding Yeast," *Science*, Vol. 282, pp. 699-705, 1998.
- [Dash *et al*. 2002] - M. Dash, K. Choi, P. Scheuermann and L. Huan, "Feature Selection for Clustering - A Filter Solution," *IEEE International Conference on Data Mining*, pp. 115-122, 2002.
- [Dave 1991] - R. N. Dave, "Characterization and Detection of Noise in Clustering," *Pattern Recognition Letters*, pp. 657-664, 1991.
- [Evans 1998] - L. C. Evans, *Partial Differential Equations*, Providence, R.I.: American Mathematical Society, 1998.
- [Friedman and Meulman 2004] - J. H. Friedman and J. J. Meulman, "Clustering Objects on Subsets of Attributes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 66, pp. 815-849, 2004.
- [Gonzalez 1991] - T. F. Gonzalez, "Covering a Set of Points in Multidimensional Space," *Information Processing Letters*, Vol. 40, pp. 181-188, 1991.
- [Huber 1981] - P. Huber, *Robust Statistics*, Wiley, New York, 1981.
- [Jennifer and Brodley 2000] - G. D. Jennifer and C. E. Brodley, "Feature Subset Selection and Order Identification for Unsupervised Learning," *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 247-254, 2000.
- [Kaplan 1991] - W. Kaplan, *Advanced Calculus*, 4th Edition, Reading, Mass.: Addison-Wesley, 1991.
- [Kaufman and Rousseeuw 1990] - L. Kaufman and P. J. Rousseeuw, *Finding Groups In Data: An Introduction To Cluster analysis*, New York: Wiley, 1990.
- [Krishnapuram and Keller 1993] - R. Krishnapuram and J. M. Keller, "A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, Vol. 1, pp. 98-110, 1993.

- [Krishnapuram *et al*. 1995a] - R. Krishnapuram, H. Frigui and O. Nasraoui, "Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation. I," *IEEE Trans. on Fuzzy Systems*, Vol. 3, No. 1, pp.29-43, 1995.
- [Krishnapuram *et al*. 1995b] - R. Krishnapuram, H. Frigui and O. Nasraoui, "Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation. II," *IEEE Trans. on Fuzzy Systems*, Vol. 3, No. 1, pp. 44-60, 1995.
- [Lam and Yan 2004] - B. Lam, and H. Yan, "Robust Clustering Algorithm for Suppression of Outliers," *International Symposium on Intelligent Multimedia, Video & Speech Processing*, accepted, October 20-22, 2004.
- [Maulik and Bandyopadhyay 2002] - U. Maulik and S. Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices," *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. 24, pp. 1650-1654, 2002.
- [Melek *et al*. 1999] - W. W. Melek, M. R. Emami, A. A. Goldenberg, "An Improved Robust Fuzzy Clustering Algorithm," *IEEE International Fuzzy Systems Conference Proceedings*, Vol. 3, pp. 1261 - 1265, 1999.
- [Milasevic and Ducharme 1987] - P. Milasevic and G. R. Ducharme, "Uniqueness of the Spatial Median," *Annals of Statistics*, Vol. 15, pp.332-333, 1987.
- [Miller and Browning 2003] - D. J. Miller and J. Browning, "A Mixture Model and EM-based Algorithm for Class Discovery, Robust Classification, and Outlier Rejection in Mixed Labeled/Unlabeled Data Sets, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, pp. 1468-1483, 2003.
- [Milligan and Cooper 1985] - G. Milligan and C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, Vol. 50, pp. 159-179, 1985.
- [Modha and Spangler 2003] - D. Modha and S. Spangler, "Feature Weighting in HCM Clustering," *Machine Learning*, Vol. 52, pp. 217-237, 2003.
- [Nasraoui and Rojas 1997] - O. Nasraoui and R. Krisnapuram, "Clustering using a genetic fuzzy least median of squares algorithm," *Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS '97*, pp. 217-221, 1997.
- [Nasraoui and Rojas 2006] -O. Nasraoui, Carlos Rojas, "Robust Clustering for Tracking Noisy Evolving Data Streams," *SDM*, pp. 618-622, 2006.

- [Pal *et al*. 1997] - N. R. Pal, K. Pal and J. C. Bezdek, "A Mixed C-means Clustering Model," *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, Vol. 1, pp. 11-21, 1997.
- [Rao 1973] - C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley, 1973.
- [Roth *et al*. 2004] - T. Lange, V. Roth, M. Braun and J. Buhmann, "Stability-Based Validation of Clustering Solutions," *Neural Computation*, 16, pp. 1299-1323, 2004.
- [Rousseeuw and Leroy 1987] - P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, 1987.
- [Rudin 1976] - W. Rudin, *Principles of Mathematical Analysis*, 3rd Edition, New York : McGraw-Hill, 1976.
- [Runkler and Bezdek 1999] - T. Runkler and J. Bezdek, "Alternating cluster estimation: a new tool for clustering andfunction approximation," *IEEE Transactions on Fuzzy Systems*, Vol. 7, pp. 377-393, 1999.
- [Schwartz 1978] - G. Schwartz, "Estimating the Dimension of a Model," *Annals of Statistics*, Vol. 6, pp. 461-464, 1978.
- [Small 1990] - C. G. Small, "A Survey of Multidimensional Medians, " *International Statistical Review*, Vol. 58, pp.263-277, 1990.
- [Trunk 1979] - G. Trunk, "A Problem of Dimensionality: A Simple Example," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 3, pp. 306-307, 1979.
- [Yang *et al*. 2002] - J. Yang, W. Wang, H. Wang, and P. Yu, "d-clusters: Capturing Subspace Correlation in a Large Data Set, " *Proceedings of 18th International Conference on Data Engineering*, pp. 517-528, 2002.
- [Yeung *et al*. 2001] - K. Yeung, D. Haynor and W. Ruzzo, "Validating Clustering for Gene Expression Data," *Bioinformatics*, Vol. 17, No. 4, pp. 309-318, 2001.
- [Yip *et al*. 2004] - K. Y. Yip, D. W. Cheung and M. K. Ng, "HARP: a practical projected clustering algorithm," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, pp. 1387-1397, 2004.