ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(3), DOI: 10.17485/ijst/2016/v9i3/71221, January 2016*

An Empirical Analysis on Effect of Data Expansion for Clustering Low Dimensional Data

Smita Prava Mishra[1], Debahuti Mishra[2]* and Srikanta Patnaik[2]

[1]Computer Science and Information Technology, Institute of Technical Education and Research, Siksha O Anusandhan University, Bhubaneswar – 751003, Odisha, India
[2]Computer Science and Engineering, Institute of Technical Education and Research, Siksha O Anusandhan University, Bhubaneswar – 751003, Odisha, India; debahutimishra@soauniversity.ac.in

Abstract

The researchers of the data mining domain presume that the study of traditional clustering techniques is saturating day by day. But, a deep insight into those techniques unfolds many silhouettes which could lead to many more applications in diverged domains. In clustering, the attributes of the data provide the information needed for data segregation. There may exist some real world data with less number of attributes but more information contained in them and may be of interest for some applications. Because of less number of attributes, the data may not be well separated by any of the clustering techniques. Data expansion techniques are methods for constructing more number of attributes from less number of attributes. With the application of these techniques, an expanded data set may be reconstructed from a given data set during data preprocessing. The current work pronounces the fact that, the expanded data at times yield better clustering results than the real data. This paper is an attempt to empirically evaluate and analyze the effects of data expansion on clustering results where validity of the results are established through internal indexing techniques and probabilistic validation measures.

Keywords: Cluster Analysis, Cluster Validity, Data Expansion, Internal Indexing, Probabilistic Measures

1. Introduction

Data mining research literature gives us enough evidences on methodologies to deal with data of high dimensions[1-5] whereas, how to extract more information from data with lower dimensions, is paid less attention. The real world data is unpredictable in nature. Some applications may demand clustering of data with lower dimensions. Clustering is the grouping of data based on their intrinsic characteristics[6-8]. The features taken for clustering may be poor in number but rich in information content. There had been research works justifying the need of data expansion techniques[9-13], supporting better supervised learning before data classification. The current work is an empirical analysis, showing the effect of pre-clustering data expansion techniques applied on various data sets collected from diverged domains. Validity of the clusters formed after data expansions are established through probabilistic measures[14-17] as well as internal indexing techniques[18-21]. The clustering techniques applied here are k-means, k-medoids, c-means and expectation maximization applied for data clustering[7-8,15,22]. The data expansion techniques used prior to clustering are trigonometric expansion up to the fifth power, exponential expansion up to the fifth power and differential expansion taking the difference of each attribute from every other attribute. The probabilistic measures taken for clustering result validation are Normalized Mutual Information (NMI), Normalized Variation of Information (NVI) and Adjusted Random Index (ARI)[23-27]. The clustering outcomes are also validated through internal indexing techniques like Modified Goodman-Kruskal

*Author for correspondence

(GKmodified) index[28], Dunn's index[18-21,29,30] and Davies-Bouldin index[18-21,31].

The rest of the paper is organized as follows. Section 2 describes the background knowledge which motivated us to take up the work done in this paper. Section 3 describes the theoretical foundation that the readers may need to understand this paper. Section 4 narrates the work taken up in this paper with a schematic representation of the same. Section 5 gives a description of the experimental setup of the under taken work. Finally the conclusion and future scope of the empirical analysis is presented in section 6. In addition to the given layout the Appendix-I given in the form of table presents the numeric values obtained from the experimentation and also summarizes the results.

2. Background Knowledge

Common data mining functionality like clustering relies on the discriminability of the features of a data set. The intrinsic dimensionality of the feature space discriminates the individual data items and results in segregation of the data set into clusters. Researchers[11] have tried to establish the inherent relationship of data dimensions with its discriminability. It in turn, establishes the effect on density and distance distributions of the data set. Some research outcomes[32,33] reveal the fact that, increase in dimensionality of the data loses their discriminative ability. Conversely some other research works[9,10,12] on the supervised learning domain, unfolds the fact that expanded data sets when fed to the classifier models for training purpose, yields better classification accuracies than unexpanded data sets. Wang et al.[12] have used image data expansion technique when fewer images are available for training purpose of a classifier. The classifier gives 27% improved accuracy as compared to unexpanded image data used for classifier training. Similar work had been evidenced by Imani et al.[10] where image data had been expanded to create pseudo-training samples prior to image classification. In their work also the expanded data gives better classification results than unexpanded training data. In another work Mili et al.[9] have done a comparative study on various expansion techniques applied on a specific classifier model and their experimental study declares the trigonometric expansion technique to be the preferable expansion mechanism for training data expansion prior to classification purpose. Zhang et al.[13] have taken a very different approach on application of expansion technique. They have used Taylor expansion mechanism for integrating old and new test data to reduce the amount of data needed to be saved for classifier adaptation. Some works elaborate on similarity measures[34] and correlation based similarity measures[35] for evaluation of performances of various clustering techniques on the respective application domains. DasGupta et al.[36] have led a mathematical elaboration on small-set expansion problem which could lead to an application specific clustering solution which has been designated by the authors as unique game interpretation via communities in social networks.

Such contradictory findings motivate us to study the effect of data expansion prior to unsupervised learning methodologies. The study of clustering domain of data mining does not give any evidence of applications citing use of expansion techniques. The present work is an experimental evaluation of the effect of data expansion techniques on pre clustered data sets. It will lead other researchers to find suitable expansion technique as well as their validation techniques for their application specific research domains as applicable to clustering. As clustering techniques do not have known, fixed, benchmarked outcomes, they rely on either probabilistic measures or internal indexing techniques for validating clustering results. The normalized forms of the probabilistic measures are taken for this experimental evaluation for keeping the quantified values within a range of 0 to 1, so that it would be easy for comparison and evaluation. These two mechanisms do not require any prior knowledge about the class or concept of interest for validating the clustering results. They rely only on intrinsic feature characteristics for measurement of compactness of the clusters.

3. Theoretical Foundation

3.1 Expansion Functions used

Any set of attribute values that we take to represent an object may be viewed as an approximation of some function deriving the attributes. At times when the set of attributes are inadequate to describe or separate objects then the expansion functions may be used to expand the approximate values to higher approximations which result in a bigger attribute set to describe the object. Some of the expansion functions may be described as follows[9-13].

3.1.1 Trigonometric Expansion Functions

Trigonometric functions[9] are mathematical functions of an angle. In modern applications we may define them as infinite series, extending them to arbitrary positive and negative values. In our case, let $A = (a_1, a_2, \ldots, a_n)$ be the set of attributes and expansion size is k.

The expansion of each attribute may be done as per Equation (1)

$$EXPANSION(a_i) = \{a_i, \sin(\pi a_i), a_i\cos(\pi a_i), a_i\sin(2\pi a_i)\cos(2\pi a_i), ..., ai\sin(k\pi a_i), a_i\cos(k\pi a_i) \quad (1)$$

If $X = [0.3\ 0.2\ 0.1]$ is to be expanded to 2nd degree so each dimension a_i will expanded to 5 dimensions, such as:

$[a_i, a_i\sin(\pi a_i), a_i\cos(\pi a_i), a_i\sin(2\pi a_i), a_i\cos(2\pi a_i)]$

For a_1 i.e 0.3

$Ex1(a_1) = 0.3$

$Ex2(a_1) = 0.3\times\cos(\pi\times 0.3) = 0.1763$

$Ex3(a_1) = 0.3\times\sin(\pi\times 0.3) = 0.2427$

$Ex4(a_1) = 0.3\times\cos(2\times\pi\times 0.3) = -0.0927$

$Ex5(a_1) = 0.3\times\sin(2\times\pi\times 0.3) = 0.2853$

Similarly doing for all dimensions we will be having 15 total dimensions as having values:

[[0.3000 0.2427 0.1763 0.2853 –0.0927]… for 0.3
0.2000 0.1176 0.1618 0.1902 0.0618]… for 0.2
0.1000 0.0309 0.0951 0.0588 0.0809]] … for 0.1

3.1.2 Exponential Expansion Functions

Any point in space may be viewed as a function of infinite sum of terms those are derivatives of that function at that point. Common practice is to take a finite number of terms of that function. So, the set of attributes of an object may be expanded further by adding few more terms of the function. For the exponential expansion taken here, let $A = (a_1, a_2, \ldots, a_n)$ be the set of attributes.

The expansion of each attribute may be done as per Equation (2)

$$EXPANSION(a_i) = \{a_i, \exp(-a_i), \exp(-2a_i), \ldots\} \quad (2)$$

If $X = [0.3\ 0.2\ 0.1]$ is to be expanded using the above expansion up to k dimensions

Let's say $k = 3$ ai is to be expanded to 3rd term as follows:

$[a_1\ \ \exp(-a_i)\ \ \exp(-2a_i)\ \ \exp(-3a_i)]$

For a_1 i.e 0.3

$Ex1(a_1) = 0.3$

$Ex2(a_1) = Exp□(-0.3) = 0.7408$

$Ex3(a_1) = Exp□(-2\times 0.3) = 0.5488$

$Ex4(a_1) = Exp□(-3\times 0.3) = 0.4066$

So the final expansion will be

[[0.3000 0.7408 0.5488 0.4066]… for 0.3
0.2000 0.8187 0.6703 0.5488]… for 0.2
0.1000 0.9048 0.8187 0.7408]]… for 0.3

3.1.3 Differential Expansion Functions

A point in space is characterized by its distance from the axis, same in number as that of its attributes. The separation of the attributes may also be viewed as attributes of that object in space. For the differential expansion considered here, let $A = (a_1, a_2, \ldots, a_n)$ be the set of attributes.

The expansion of each attribute may be done as per Equation (3)

$$EXPANSION(a_i) = (a_1, a_2, \ldots, a_n, a_1 - a_2, a_1 - a_2, \ldots, a_2 - a_n, a_2 - a_n, \ldots, a_2 - a_n, \ldots \ldots) \quad (3)$$

If $X = [0.3\ 0.2\ 0.1]$ is to be expanded using differential expansion, then the terms would be:

[0.3 0.2 0.1 (0.3-0.2) (0.3-0.1) (0.2-0.1)]

So final expansion values are [0.3000 0.2000 0.1000 0.1000 0.2000 0.1000]

If there are n terms then expanded set will have $n \times (n + 1) / 2$ terms.

3.2 Clustering Techniques used

Clustering or cluster analysis is the method of data segregation, based on maximization of intra-cluster similarity and minimization of inter-cluster similarity. Out of several clustering methods, partitioning based[7,8,15,22] data clustering are one of the primitive types. They segregate the data set into a predefined number of partitions, such that each data object belongs to exactly one partition. The k-means clustering algorithm forms the clusters based on minimization of Euclidian distance of data objects from the randomly chosen cluster centers. The cluster centers are iteratively recomputed from the mean of data objects assigned to the respective clusters. Though performance of k-means is challenging, it undergoes critics like, being sensitive to noise and choosing non-member cluster centers. The k-medoids clustering algorithm is appointed to neutralize these biases. Medoids of a cluster are the data objects whose average dissimilarity to all the objects assigned to that cluster are minimal. The c-means clustering algorithm is soft or fuzzy based clustering algorithm which assigns membership levels of each data object with respect to each cluster. Each data element may belong to more than one cluster, as per the fuzziness. While using this method for hard or crisp clustering, the data objects are assigned to the clusters depending on the maximum degree of belongingness towards the clusters. The expectation-maximization[14,15] algorithm is an optimization method which determines the maximal-likelihood estimate of an unknown parameter. This algorithm when used for data clustering uses the Gaussian component estimate for the random variable, which is the expected cluster. It relies on the fact that, the data objects those belong to one cluster are generated from identical Gaussian component.

3.3 Internal Indexing Techniques used for Cluster Validation

3.3.1 Dunn's Index

Maximized inter-cluster distance and minimized intra-cluster distance maximizes the contribution to the Dunn's index value. The cluster that obtains maximized index value is taken as optimal[28]. For each cluster partition, the Dunn index can be calculated by the formula given in Equation (4):

$$D = \min_{1\le i\le n}\left\{\min_{1\le j\le n, i\ne j}\left\{\frac{d(i,j)}{\max d'(k)}\right\}\right\} \tag{4}$$

where

$d(i, j)$ represents the distance between clusters i and j,

$d'(k)$ measures the intra-cluster distance of cluster k and

n represents the number of clusters.

3.3.2 Davies-Bouldin Index

Davies-Bouldin index is determined by the ratio of the sum of the within-cluster scatter to between-cluster separation. Smaller values of the calculated index indicate good clustering[28]. It can be calculated by the formula given in Equation (5):

$$DB = \frac{1}{n}\sum_{i=1}^{n}\max_{i\ne j}\left\{\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right\} \tag{5}$$

where

n is the number of clusters,

c_x is the centroid of cluster x,

σ_x is the average distance of all objects in cluster x to centroid c_x and

$d(c_i, c_j)$ is the distance between centroids c_i and c_j.

3.3.3 Modified Goodman-Kruskal (GKmodified) Index

The original Goodman-Kruskal indexing technique[28] considers all possible quadruples for a given data set. The quadruples may be either concordant or disconcordant, however the quadruples are disjoint in nature. A good cluster is one with many concordant and few disconcordant quadruples. In contrast Modified Goodman-Kruskal ($GK_{modified}$) index takes triplets instead of quadruples to validate the clusters by avoiding disjointedness. A good cluster is one with many concordant and few disconcordant quadruples. Let Nc and Nd denote the number of concordant and disconcordant triplets, respectively. The cluster with larger value of ($GK_{modified}$) indicates a good clustering. Then the modified GK index, ($GK_{modified}$), is defined as per the formula in Equation (6):

$$GK_{modified} = \frac{N_c - N_d}{N_c + N_d} \tag{6}$$

3.4 Probabilistic Measures used for Cluster Validation

Probability measures are a way of agreement between two clustering results[23-27,34]. They measure the degree of agreement among the results. The probabilistic measures taken for this empirical study are as follows:

3.4.1 Normalized Mutual Information (NMI)

NMI is a probabilistic method to measure the similarity between two candidates clustering.

Let the information contained, which is also called entropy of cluster

Co be $E(Co) = -\sum_{o}\frac{n_o}{n}\log\left(\frac{n_o}{n}\right)$ and

Cc be $E(Cc) = -\sum_{c}\frac{n_c}{n}\log\left(\frac{n_c}{n}\right)$

Similarly the joint entropy of Co and Cc be $E(Co, Cc) = -\sum_{oc}\frac{n_{oc}}{n}\log\left(\frac{n_{oc}}{n}\right)$. The Normalized Mutual Information value may be computed as per the formula given in Equation (7).

$$NMI(Co, Cc) = \frac{E(Co) + E(Cc) - E(Co, Cc)}{\sqrt{E(Co) + E(Cc)}} \quad (7)$$

3.4.2 Normalized Variation of Information (NVI)

On the basis of variation of information, another information theoretic measure is determined for cluster validation. It behaves consistently if data sets of different sizes and clustering with different number of clusters are considered. NVI value may be computed as per the formula given in Equation (8).

$$NVI(Co, Cc) = 1 - \frac{2 \times [E(Co) + E(Cc) - E(Co, Cc)]}{E(Co) + E(Cc)} \quad (8)$$

3.4.3 Adjusted Random Index (ARI)

ARI is a measure of the similarity between two data clusters. It has a value between 0 and 1, with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same. The Adjusted Random Index value may be computed as per the formula given in Equation (9).

Let :

Co = The original set of clusters

Cc = The set of clusters obtained by applying clustering

no = The number of objects in Co

nc = The number of objects in Cc

noc = The number of objects in |Co ∩Cc|

$$ARI(Co, Cc) = 1 - \frac{\sum_{oc}\binom{oc}{2} - \frac{\left[\sum_{o}\binom{n_o}{2}\sum_{c}\binom{n_c}{2}\right]}{\frac{n}{2}}}{\frac{1}{2}\left[\sum_{o}\binom{n_o}{2} + \sum_{c}\binom{n_c}{2}\right] - \frac{\sum_{o}\binom{n_o}{2} + \sum_{c}\binom{n_c}{2}}{\frac{n}{2}}} \quad (9)$$

4. Schematic Representation

The schematic representation, pictorially represented in Figure 1 describes the flow of the work carried out sequentially for the purpose of the experimental study done in this paper. At first the data sets are collected and their class labels were removed as the study aims at cluster analysis. Then the expansion techniques were applied in the respective data sets to reconstruct bigger data sets from the existing data sets. In the present work, bigger refers to attribute construction from the existing attributes. On those expanded data sets, various clustering algorithms were applied. In addition to that, clustering techniques were also applied to the unexpanded data sets. Then for the study of clustering qualities, two categories of validation mechanisms were used. The first category refers to probabilistic measures of cluster validation which has a mathematical foundation in measuring the similarity among objects belonging to the same cluster. The other category refers to internal indexing techniques for cluster validation. It is to reassure the cluster qualities with reference to the basic clustering criteria i.e. maximizing intra cluster similarity and minimizing inter cluster similarity. The results obtained after implementation of the experimental setup, shown in Figure 1, were considered for a comparative analysis.

5. Experimental Setup

5.1 Data Set Description

For establishing a comparative setup needed for analysis purpose, data sets were chosen from UCI Machine Learning Repository[37]. To avoid the influence of any specific domain, six different data sets were collected from diverged domains. For ease of computation, some data tuples with noisy and missing data were dropped. After preprocessing, the final size of the data sets was as described in Table 1. The Table describes the number of final data instances and the number of features considered for clustering purpose, after removing the class labels in addition to the number of natural classes of the individual data sets taken for this analytical study.

Table 1. Data set Description

Data sets	No. of Instances	No. of Features	No. of Classes
Iris	150	4	3
Wine	178	13	3
Breast Cancer Data	98	25	3
WDBC	569	31	2
Connectionist Bench (Sonar, Mines vs. Rocks)	208	60	2
Parkinsons Disease Data Set	195	22	2

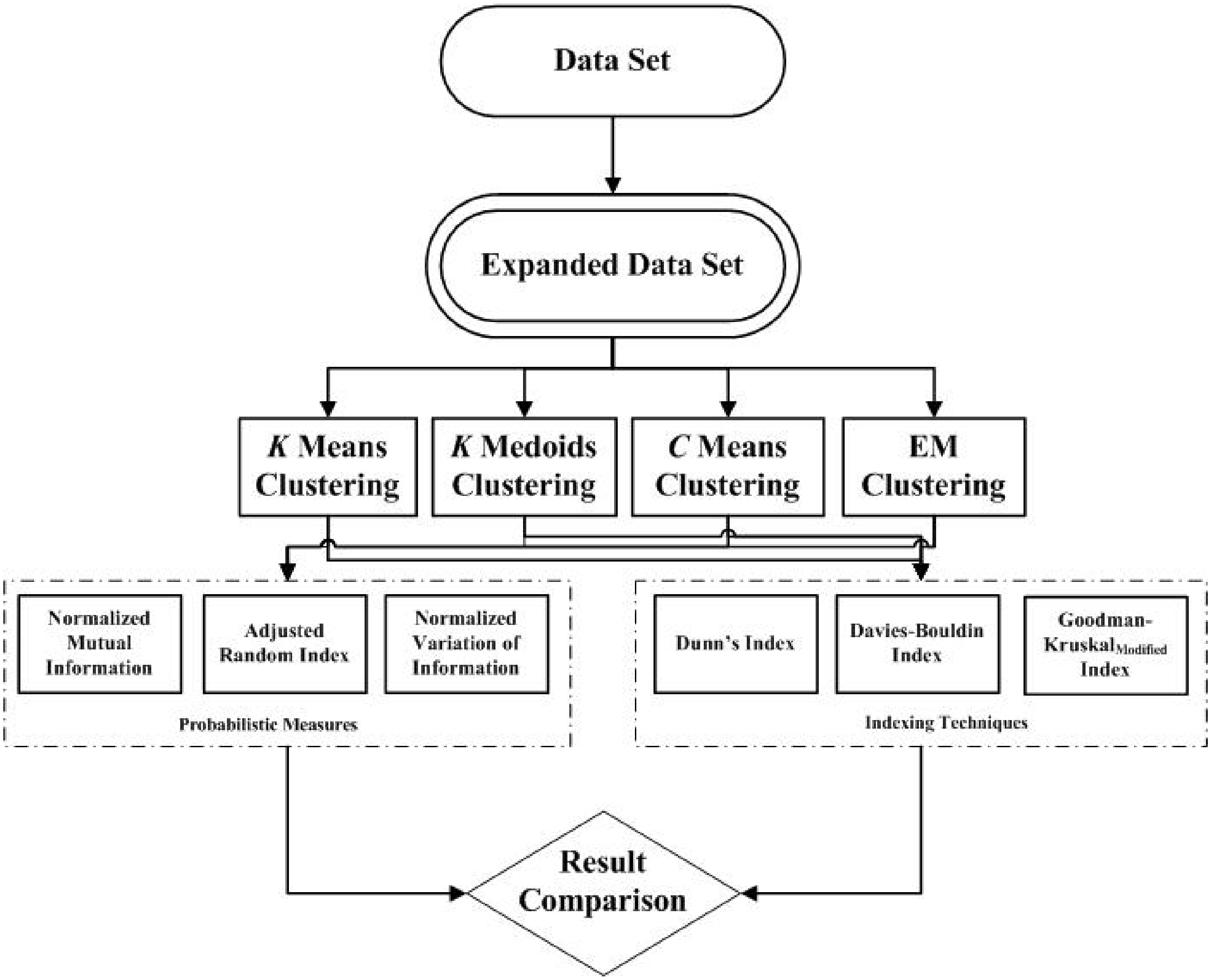

Figure 1. Schematic representation of the analytical study.

5.2 Experimentation

In the experimental setup for this work, initially as described in the data set description section, different data sets from diverged domains are gathered. For computational simplicity and interpretable results, data tuples with missing or noisy values were removed from the data sets. The data sets obtained thereafter are as presented in Table 1. The data sets taken for this experimental analysis are at first treated with the pre-discussed clustering techniques and their standard values for the various cluster validation mechanism were taken for benchmarking purpose. There after three different data expansion techniques such as Trigonometric Expansion, Exponential Expansion and Differential Expansion were applied on the same data sets to regenerate bigger expanded data sets with higher dimensions. Four different clustering techniques with different characteristics were chosen for the experimentation to neutralize the biasing effect of any specific clustering technique. The clustering techniques hence chosen are k- means, k- medoids, c-means and expectation maximization technique applied for clustering purpose. Once the clusters were obtained, now the clustering qualities are to be verified. For this purpose, instead of relying on any single method of cluster validation we had chosen six different methods, out of which three are internal indexing techniques of cluster validation and other three are probabilistic measures of cluster validation. The internal indexing techniques of cluster validation were preferred for the study because they are unsupervised methods and they determine the cluster qualities based on the attribute characteristics of the data sets. The internal indexing techniques are Dunn's index, Davies-Bouldin index and Modified Goodman-Kruskal (GKmodified) index. Dunn's index and Davies-Bouldin index are well known and well accepted as internal indexing based cluster validity methods whereas Modified Goodman-Kruskal ($GK_{modified}$) indexing is a method which overcomes the drawbacks of traditional Goodman-Kruskal indexing technique for cluster validation. The other category of cluster validation techniques taken for experimentation are probabilistic measures, which may be considered as supervised methods of cluster validation and are also comparable to external indexing techniques of cluster validation. Their probabilistic nature leads to a mathematical basis of validation mechanism and also considers the class label information for determining the cluster qualities. The probabilistic measures taken for experimentation are NMI, NVI and ARI. All the discussed techniques generate a value in the range of 0 to 1 such that it would be preferable for comparison purpose. Same clustering techniques and cluster validation techniques were then applied on the expanded data sets to obtain the new validity measures and were compared with the benchmarked values obtained earlier from unexpanded data sets.

5.3 Result Analysis

The results obtained after applying four clustering techniques such as *k*- means, *k*- medoids, *c*- means and expectation maximization technique applied for clustering purpose were treated with six cluster validity measures such as Dunn's index, Davies-Bouldin index, Modified Good man-Kruskal ($GK_{modified}$) index, NMI, NVI and ARI. There results were stored for benchmarking purpose so that they can be compared with the same mechanisms when applied on expanded data sets. The table in Appendix I depict the numeric values obtained from the above discussed experimental setup. The table in Appendix II illustrates a summarized representation of the comparison results. The benchmarked standard values obtained from unexpanded data sets are demarcated with the symbol (Θ), whereas the values which are more preferred as compared to the benchmarked results are demarcated with the symbol ($\uparrow$) and the values which are less preferred as compared to the benchmarked results are demarcated with the symbol ($\downarrow$). Though summary of the result is not entirely in the favor of expanding the data sets before clustering, still the effect of expansion cannot be ignored. Many a times the effect is significant and should be able to draw the attention of researchers towards domain specific effects of data expansion and also the effect of other expansion techniques on clustering results.

6. Conclusion and Future Scope

In real world data, the dimensionality may not be large always. Sometimes small set of attributes may have large information contained in them. Yet the data items may not be well separable. Expansion techniques may come to rescue at times by generation of well separable attributes. The effect of data expansion techniques on smaller data sets may not be ignored for clustering purposes. In future, more empirical studies may be done for the study of the effect of data expansion techniques on domain specific data sets. However, more vivid studies can also be made on performances of individual expansion techniques on domain specific data sets. Though increase in dimensionality of the data may add to computational complexities, the increased accuracy of clustered data may prove to be worth payoff of computational complexity for sensitive domains like disease prediction, drug analysis etc. Further studies on computational complexities and performance issues may open new avenues on the effect of data expansion on cluster analysis.

7. References

1. Steinbach M, Ertoz L, Kumar V. The challenges of clustering high dimensional data. New Directions in Statistical Physics. 2004. p. 273–309.
2. Bouveyron C, Girard S, Schmid C. High-dimensional data clustering. Computational Statistics and Data Analysis. 2007; 52:502–19.
3. Bouveyron C, Brunet-Saumard C. Model-based clustering of high-dimensional data: A review. Computational Statistics and Data Analysis. 2014; 71:52–78.
4. Parsons L, Haque E, Liu H. Subspace clustering for high dimensional data: a review. ACM SIGKDD Explorations Newsletter - Special issue on learning from imbalanced datasets. 2004; 6(1):90–105.
5. Sembiring RW, Zain JM, Embong A. Clustering high dimensional data using subspace and projected clustering algorithms. International Journal of Computer Science and Information Technology. 2010; 2(4):162–70.

6. Wikipedia, the free encyclopedia. Cluster Analysis. 2015. Available from: en.wikipedia.org/wiki/Cluster_analysis
7. Estivill-Castro V. Why so many clustering algorithms — A Position Paper. ACM SIGKDD Explorations Newsletter. 2002; 4(1):65–75.
8. Han J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann; 2000.
9. Mili F, Hamdi M. A comparative study of expansion functions for evolutionary hybrid functional link artificial neural networks for data mining and classification. International Conference on Computer Applications Technology. 2013 Jan. p. 1–8.
10. Imani M, Ghassemian F. Adaptive expansion of training samples for improving hyperspectral image classification performance. 21st Iranian Conference on Electrical Engineering; Mashhad. 2013 May. p. 1–6.
11. Houle ME. Dimensionality, discriminability, density and distance distributions. 13th International Conference on Data Mining Workshops; Dallas. 2013 Dec. p. 468–73.
12. Wang SY, Liao SW, Hsieh LC, Chen YY, Hsu WH. Learning by expansion: Exploiting social media for image classification with few training examples. Neurocomputing. 2012; 95(15):117–25.
13. Zhang C, Hamid R, Zhang Z. Taylor expansion based classifier adaptation: application to person detection. IEEE Conference on Computer Vision and Pattern Recognition; Anchorage. 2008 June. p. 1–8.
14. Dunham MH. Data Mining- Introductory and Advanced Concepts. Pearson Education; 2006.
15. Abass OA. Comparisons between data clustering algorithms. The International Arab Journal of Information Technology. 2008; 5(3):320–5.
16. Jain AK, Richard CD. Algorithms for Clustering Data. Advanced Reference Series: Computer Science. Englewood: NJ. Prentice-Hall, Inc.; 1988.
17. Theodoridis S, Koutroumbas K. Pattern Recognition. NY: USA. Academic Press; 1999.
18. Bezdek J, Pal N. Some new indexes of cluster validity. IEEE Transactions on System, Man and Cybernetics-Part B. 1998; 28(3):301–15.
19. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. Journal of Intelligent Information Systems. 2001; 17(2):107–45.
20. Rendon E, Abundez IM, Arizmendi A, Quroz EM. Internal versus external cluster validation indices. International Journal of Computers and Communications. 2011; 5(1):27–34.
21. Rendon E, Abundez IM, Zagal CG, Arizmendi A, Quroz EM, Arzate E. A comparison of internal and external cluster validation indexes. Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications; Montreux: Switzerland. 2011 Dec. p. 158–63.
22. Vinoth Kumar B, Karpagam GR, Vijaya Rekha N. Performance analysis of deterministic centroid initialization method for partitional algorithms in image block clustering. Indian Journal of Science and Technology. 2015 Apr; 8(S7):63–73.
23. Meila M. Comparing clusterings by the variation of information. Proceedings of Learning theory and kernel machines; Washington: USA. Springer Berlin Heidelberg; 2003 Aug. p. 173–87.
24. Wu J, Chen J, Xiong H, Xie M. External validation measures for k-means clustering: A data distribution perspective. Expert Systems with Applications. 2009 Apr; 36(3):6050–61.
25. Sripada SC, Sreenivasa Rao M. Comparison of purity and entropy of k-means clustering and fuzzy c means clustering. Indian Journal of Computer Science and Engineering. 2011; 2(3):343–6.
26. Topchy AP, Law MHC, Jain AK, Fred AL. Analysis of consensus partition in cluster ensemble. Proceedings of the 4th IEEE International Conference on Data Mining; Washington: DC. IEEE Computer Society; 2004. p. 225–32.
27. Ghosh J, Acharya A. Cluster Ensembles. WIREs Data Mining and Knowledge Discovery. 2011; 1(4):305–15.
28. Mishra SP, Mishra D, Patnaik S. A new improved cluster validity indexing technique: harnessed from goodman-kruskal validity index. International Journal of Information and Communication Technology. 2015; 7(1): 88–99.
29. Bezdek J, Pal N. Cluster validation with generalized dunn's indices. Proceedings of the 2nd New Zealand Two-Stream International Conference on Artificial Neural Networks and Expert Systems; Dunedin: New Zealand. 1995 Nov. p. 190–3.
30. Dunn JC. Well separated clusters and optimal fuzzy partitions. Journal of Cybernetics. 1974; 4(1):95–104.
31. Davies DL, Bouldin DW. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979; 1(4):224–7.
32. Weber R, Schek H J, Blott S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. Proceedings of the 24rd International Conference on Very Large Data Bases; New York: USA. 1998 Aug. p. 194–205.
33. Pestov V. On the geometry of similarity search: dimensionality curse and concentration of measure. Information Processing Letters. 2000; 73(1-2): 47–51.
34. Muruga Radha Devi D, Thambidurai P. Similarity measurement in recent biased time series databases using

different clustering methods. Indian Journal of Science and Technology. 2014 Feb; 7(2): 189–98.

35. Nagaraj R, Thiagarasu V. Correlation similarity measure based document clustering with directed ridge regression. Indian Journal of Science and Technology. 2014 May; 7(5): 692–7.
36. Das Gupta B, Desai D. On a connection between small set expansions and modularity clustering. Information Processing Letters. 2014; 114(7): 349–52.
37. UCI Machine Learning Repository. 2012. Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html

APPENDIX-I

Table 2. Comparison of expansion techniques through Dunn's index

	DUNN-INDEX (Bigger value better Cluster)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.000397	0.000341	0.000544	0.000205	2.74E-05	2.34E-05	2.34E-05	3.53E-05	8.78E-05	2.54E-05	0.000109	3.95E-05	0.000471	0.000111	0.000152	0.000111
Wine	2.51E-06	2.7E-06	1.7E-06	2.25E-06	2.36E-07	2.19E-07	2.57E-07	2.46E-07	2.42E-07	2.19E-07	2.53E-07	2.31E-07	2.98E-07	2.98E-07	2.98E-07	2.98E-07
Breast Cancer Data	2.26E-08	1.9E-08	1.89E-08	2.52E-08	1.9E-09	7.8E-12	2.05E-09	2.1E-09	1.81E-09	9.89E-12	2.46E-09	6.74E-12	1.59E-09	1.37E-09	1.8E-09	5.37E-12
WDBC	3.99E-10	2.92E-10	4.88E-10	2.56E-10	5.39E-11	5.39E-11	5.39E-11	2.96E-11	4.74E-11	3.74E-11	3.61E-11	2.69E-11	2.82E-11	2.82E-11	3.39E-11	8.99E-12
Sonar, Mines vs. Rocks	3.89E-08	3.16E-08	3.62E-08	3.74E-08	3.65E-09	3.75E-09	3.58E-09	3.58E-09	3.45E-09	3.24E-09	3.32E-09	3.03E-09	1.42E-09	1.2E-09	1.36E-09	9.9E-10
Parkinsons Disease	1.62E-10	1.15E-10	1.19E-10	1.36E-10	1.45E-11	1.24E-11	1.14E-11	1.13E-12	1.43E-11	1.23E-11	1.23E-11	1.06E-12	1.14E-11	1.21E-11	1.24E-11	1.08E-12

Table 3. Comparison of expansion techniques through Davies-Bouldin index

	DB-INDEX (Smaller value better Cluster)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.083394	0.090315	0.092462	1.025567	0.080717	0.111379	0.126377	0.104145	0.072303	0.095226	0.069899	0.071304	0.051438	0.05311	0.05402	0.065403
Wine	0.068371	0.075929	0.056856	0.081039	0.055985	0.099355	0.098612	0.071878	0.054221	0.10223	0.044092	0.04507	0.006697	0.006912	0.006047	0.005994
Breast Cancer Data	0.105149	0.189541	0.121907	0.121687	0.043219	0.064424	0.04863	0.047835	0.035633	0.067389	0.045657	0	0.010907	0.020716	0.015164	0
WDBC	0.025269	0.02631	0.025058	0.05144	0.003168	0.00315	0.003522	0	0.00281	0.002993	0.002658	0	0.001865	0.001903	0.00203	0
Sonar, Mines vs. Rocks	0.015189	0.018288	0.019575	0.014084	0.003385	0.00375	0.004609	0	0.002421	0.004612	0.003394	0	0.000697	0.000702	0.000658	0
Parkinsons Disease	0.034985	0.044884	0.04446	0.056182	0.012876	0.017144	0.017701	0	0.013318	0.014214	0.013869	0	0.006871	0.007375	0.006746	0

Table 4. Comparison of expansion techniques through $GK_{modified}$ index

	GK-modified (Bigger value better Cluster) (Between 0 to 1)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.964014	0.970289	0.950084	0.840451	0.846461	0.635876	0.664076	0.975945	0.933376	0.843699	0.968837	0.975671	0.928557	0.976661	0.937784	0.99149
Wine	0.953199	0.93389	0.961277	0.933833	0.783126	0.756997	0.728502	0.926194	0.923517	0.815691	0.87516	0.912724	0.976016	0.931006	0.967159	0.972019
Breast Cancer Data	0.83423	0.705329	0.768516	0.769587	0.661687	0.559034	0.638338	0.847254	0.661075	0.565263	0.721925	1	0.849165	0.692234	0.724757	1
WDBC	0.921765	0.908165	0.894913	0.710634	0.898477	0.888169	0.813933	1	0.891919	0.883486	0.80812	1	0.930708	0.933428	0.923976	1
Sonar, Mines vs. Rocks	0.914169	0.903288	0.875147	0.94944	0.902068	0.860326	0.847662	1	0.967544	0.829558	0.892031	1	0.883308	0.931148	0.91492	1
Parkinsons Disease	0.958763	0.929491	0.94199	0.940922	0.971049	0.92917	0.928209	1	0.938144	0.915496	0.92682	1	0.930239	0.93088	0.911436	1

Table 5. Comparison of expansion techniques through Normalized Mutual Information

	NMI (Bigger value better Cluster) (Between 0 to 1)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.734654	0.737888	0.743317	0.602369	0.380793	0.17601	0.253872	0.864186	0.754372	0.758206	0.791886	0.830835	0.771104	0.813866	0.735613	0.899695
Wine	0.847295	0.734631	0.821558	0.708585	0.424939	0.322663	0.201308	0.704769	0.661265	0.414339	0.396941	0.772761	0.815433	0.786475	0.815036	0.817782
Breast Cancer Data	0.465026	0.258994	0.202026	0.376431	0.254109	0.033109	0.237477	0.405801	0.337558	0.082483	0.154673	NA	0.376392	0.329539	0.234021	NA
WDBC	0.625173	0.688157	0.621534	0.192315	0.404712	0.359157	0.435136	NA	0.431593	0.549517	0.241059	NA	0.592977	0.599616	0.679029	NA
Sonar, Mines vs. Rocks	0.007421	0.011988	2.29E-07	7.85E-05	2.74E-05	0.020004	0.000969	NA	0.046946	0.001274	0.020537	NA	0.010955	0.004524	0.007006	NA
Parkinsons Disease	0.118936	0.202293	0.239146	0.06633	0.097798	0.213691	0.216494	NA	0.149591	0.185238	0.225242	NA	0.260233	0.216494	0.260123	NA

Table 6. Comparison of expansion techniques through Normalized Variation of Information

	NVI (Bigger value better Cluster) (Between 0 to 1)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.419441	0.415599	0.408516	0.56901	0.765108	0.903503	0.854609	0.239149	0.394862	0.389466	0.344527	0.289377	0.372792	0.313913	0.418206	0.182325
Wine	0.264959	0.419435	0.302857	0.451315	0.730226	0.807636	0.888089	0.455884	0.506061	0.738714	0.75338	0.370335	0.311621	0.351922	0.312187	0.308265
Breast Cancer Data	0.697377	0.851633	0.887875	0.768755	0.854721	0.9832	0.865582	0.746249	0.796957	0.957081	0.916408	NA	0.768667	0.803139	0.867821	NA
WDBC	0.545624	0.475531	0.54915	0.893616	0.746685	0.781449	0.721984	NA	0.724857	0.621185	0.86299	NA	0.579636	0.572629	0.486378	NA
Sonar, Mines vs. Rocks	0.996276	0.993971	1	0.999961	0.999986	0.98991	0.999515	NA	0.976354	0.999363	0.989625	NA	0.994494	0.997733	0.996485	NA
Parkinsons Disease	0.9369	0.888182	0.865007	0.965727	0.949049	0.880759	0.879359	NA	0.919159	0.898584	0.873888	NA	0.851241	0.879359	0.851476	NA

Table 7. Comparison of expansion techniques through Adjusted Random Index

	ARI (Bigger value better Cluster) (Between 0 to 1)															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0.704547	0.672893	0.729555	0.529866	0.351311	0.139094	0.215447	0.886508	0.664414	0.731041	0.80247	0.851608	0.696625	0.774382	0.728655	0.904886
Wine	0.864168	0.713336	0.835464	0.711731	0.402074	0.292816	0.16058	0.687681	0.641985	0.308013	0.334715	0.773141	0.847805	0.790504	0.846859	0.847056
Breast Cancer Data	0.530662	0.252518	0.18166	0.348819	0.283514	0.002672	0.249457	0.409993	0.395551	0.072587	0.159116	0.001853	0.441766	0.262308	0.258235	0.001853
WDBC	0.711473	0.78591	0.736598	0.282325	0.505112	0.460161	0.535731	1.94E-05	0.554308	0.671101	0.305081	1.94E-05	0.634049	0.657484	0.747982	1.94E-05
Sonar, Mines vs. Rocks	0.004459	0.006426	-0.00473	-0.00472	-0.00472	0.010905	-0.00335	0.000139	0.013485	-0.00444	0.022017	0.000139	0.013395	-0.00147	0.004458	0.000139
Parkinsons Disease	-0.09358	0.128252	0.118828	-0.07977	-0.09769	0.014965	0.119717	0.000195	-0.07128	0.152676	0.135538	0.000195	0.094434	0.119717	0.233793	0.000195

APPENDIX-II

Table 8. Summary of comparison of expansion techniques through Dunn's index

	DUNN-INDEX															
	k-means				*k*-medoids				*c*-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	ʘ	↓	↑	↓	ʘ	↓	↓	↑	ʘ	↓	↑	↓	ʘ	↓	↓	↓
Wine	ʘ	↑	↓	↓	ʘ	↓	↑	↑	ʘ	↓	↑	↓	ʘ	↑	↑	↑
Breast Cancer Data	ʘ	↓	↓	↑	ʘ	↓	↑	↑	ʘ	↓	↑	↓	ʘ	↓	↑	↓
WDBC	ʘ	↓	↑	↓	ʘ	↑	↑	↓	ʘ	↓	↓	↓	ʘ	↑	↑	↓
Sonar, Mines vs. Rocks	ʘ	↓	↓	↓	ʘ	↑	↓	↓	ʘ	↓	↓	↓	ʘ	↓	↓	↓
Parkinsons Disease	ʘ	↓	↓	↓	ʘ	↓	↓	↓	ʘ	↓	↓	↓	ʘ	↑	↑	↓

Table 9. Summary of comparison of expansion techniques through Davies-Bouldin index

	DB-INDEX															
	k-means				k-medoids				c-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	Θ	↓	↓	↓	Θ	↓	↓	↓	Θ	↓	↑	↑	Θ	↓	↓	↓
Wine	Θ	↓	↑	↓	Θ	↓	↓	↓	Θ	↑	↑	↑	Θ	↓	↑	↑
Breast Cancer Data	Θ	↓	↓	↓	Θ	↓	↓	↓	Θ	↓	↓	↑	Θ	↓	↓	↑
WDBC	Θ	↓	↑	↓	Θ	↓	↓	↑	Θ	↓	↑	↑	Θ	↓	↓	↑
Sonar, Mines vs. Rocks	Θ	↓	↓	↑	Θ	↓	↓	↑	Θ	↓	↓	↑	Θ	↓	↑	↑
Parkinsons Disease	Θ	↓	↓	↓	Θ	↓	↓	↑	Θ	↓	↓	↑	Θ	↓	↑	↑

Table 10. Summary of comparison of expansion techniques through $GK_{modified}$ index

	GK-modified-INDEX															
	k-means				*k*-medoids				*c*-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	ʘ	↑	↓	↓	ʘ	↓	↓	↑	ʘ	↓	↑	↑	ʘ	↑	↑	↑
Wine	ʘ	↓	↑	↓	ʘ	↓	↓	↑	ʘ	↓	↓	↓	ʘ	↓	↓	↓
Breast Cancer Data	ʘ	↓	↓	↓	ʘ	↓	↓	↑	ʘ	↓	↑	↑	ʘ	↓	↓	↑
WDBC	ʘ	↓	↓	↓	ʘ	↓	↓	↑	ʘ	↓	↓	↑	ʘ	↑	↓	↑
Sonar, Mines vs. Rocks	ʘ	↓	↓	↑	ʘ	↓	↓	↑	ʘ	↓	↓	↑	ʘ	↑	↑	↑
Parkinsons Disease	ʘ	↓	↓	↓	ʘ	↓	↓	↑	ʘ	↓	↓	↑	ʘ	↑	↓	↑

Table 11. Summary of comparison of expansion techniques through Normalized Mutual Information

	NMI															
	k-means				*k*-medoids				*c*-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansionv (Pow-5)	With Differential Expansion
Iris	Θ	↑	↑	↓	Θ	↓	↓	↑	Θ	↑	↑	↑	Θ	↑	↓	↑
Wine	Θ	↓	↓	↓	Θ	↓	↓	↑	Θ	↓	↓	↑	Θ	↓	↓	↑
Breast Cancer Data	Θ	↓	↓	↓	Θ	↓	↓	↑	Θ	↓	↓	↓	Θ	↓	↓	↓
WDBC	Θ	↑	↓	↓	Θ	↓	↑	↓	Θ	↑	↓	↓	Θ	↑	↑	↓
Sonar, Mines vs. Rocks	Θ	↑	↓	↓	Θ	↑	↑	↓	Θ	↓	↓	↓	Θ	↓	↓	↓
Parkinsons Disease	Θ	↑	↑	↓	Θ	↑	↑	↓	Θ	↑	↑	↓	Θ	↓	↓	↓

Table 12. Summary of comparison of expansion techniques through Normalized Variation of Information

	NVI															
	k-means				*k*-medoids				*c*-means				Expectation Maximization			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	⊙	↓	↓	↑	⊙	↑	↑	↓	⊙	↓	↓	↓	⊙	↓	↑	↓
Wine	⊙	↑	↑	↑	⊙	↑	↑	↓	⊙	↑	↑	↓	⊙	↑	↑	↓
Breast Cancer Data	⊙	↑	↑	↑	⊙	↑	↑	↓	⊙	↑	↑	↓	⊙	↑	↑	↓
WDBC	⊙	↓	↑	↑	⊙	↑	↓	↓	⊙	↓	↑	↓	⊙	↓	↓	↓
Sonar, Mines vs. Rocks	⊙	↓	↑	↑	⊙	↓	↓	↓	⊙	↑	↑	↓	⊙	↑	↑	↓
Parkinsons Disease	⊙	↓	↓	↑	⊙	↓	↓	↓	⊙	↓	↓	↓	⊙	↑	↑	↓
Wine	⊙	↓	↓	↓	⊙	↓	↓	↑	⊙	↓	↓	↑	⊙	↓	↓	↓
Breast Cancer Data	⊙	↓	↓	↓	⊙	↓	↓	↑	⊙	↓	↓	↓	⊙	↓	↓	↓
WDBC	⊙	↑	↑	↓	⊙	↓	↓	↓	⊙	↑	↓	↓	⊙	↑	↑	↓
Sonar, Mines vs. Rocks	⊙	↑	↓	↓	⊙	↓	↑	↑	⊙	↓	↑	↓	⊙	↓	↓	↓
Parkinsons Disease	⊙	↑	↑	↓	⊙	↑	↑	↑	⊙	↑	↑	↑	⊙	↑	↑	↓

Table 13. Summary of comparison of expansion techniques through Adjusted Random Index

	ARI															
	***k*-means**				***k*-medoids**				***c*-means**				**Expectation Maximization**			
	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion	Without Expansion	With Trigonometric Expansion (Pow-5)	With Exponential Expansion (Pow-5)	With Differential Expansion
Iris	0	↓	↑	↓	0	↓	↓	↑	0	↑	↑	↑	0	↑	↑	↑
Wine	0	↓	↓	↓	0	↓	↓	↑	0	↓	↓	↑	0	↓	↓	↓
Breast Cancer Data	0	↓	↓	↓	0	↓	↓	↑	0	↓	↓	↓	0	↓	↓	↓
WDBC	0	↑	↑	↓	0	↓	↓	↓	0	↑	↓	↓	0	↑	↑	↓
Sonar, Mines vs. Rocks	0	↑	↓	↓	0	↓	↑	↑	0	↓	↑	↓	0	↓	↓	↓
Parkinsons Disease	0	↑	↑	↓	0	↑	↑	↑	0	↑	↑	↑	0	↑	↑	↓