

A performance comparison of using principal component analysis and ant clustering with fuzzy c-means and k-harmonic means

Phichete Julrode

Department of Information Technology
Rangsit University,
Patumthani 12000, Thailand
phichete@pkru.ac.th

Siriporn Supratid

Department of Information Technology
Rangsit University,
Patumthani 12000, Thailand
siri_sup1@hotmail.com

Ureerat Suksawatchon

Department of Information Technology
Burapha University,
Chonburi, 20131, Thailand
ureerat.w@gmail.com

Abstract—Several clustering researches focus on the idea for achieving the optimal initial set of clusters before performing further clustering. This may be accomplished by performing two-level clustering. However, such an idea may possibly not either significantly improve the accuracy rate or well alleviate local traps; contrarily it usually generates abundant runtime consumption. Thereby, one may turn to focus on the relieving the problems of high dimensional, noisy data and hidden outliers. Such difficulties usually occur in real-world environment; and can seriously spoil the computation of several of types of learning, including clustering. This paper proposes a performance comparison using feature reduction based method, principal component analysis (PCA) and ant clustering algorithm combining with two particular fuzzy clustering approaches, fuzzy c-means (FCM) and k-harmonic means (KHM). FCM and KHM are soft clustering algorithms that retain more information from the original data than those of crisp or hard. PCA is employed as preprocess of FCM and KHM for relieving the curse of high-dimensional, noisy data. Ant clustering algorithm is employed as the first level of clustering that supplies the optimal set of initial clusters to those soft clustering methods. Comparison tests among related methods, PCA-FCM, PCA-KHM, ANT-FCM and ANT-KHM are evaluated in terms of clustering objective function, adjusted rand index and time consumption. Seven well-known benchmark real-world data sets are employed in the experiments. Within the scope of this study, the superiority of using PCA for feature reduction over the two-level clustering, ANT-FCM and ANT-KHM is pointed out.

Keywords—component; Principal component analysis; fuzzy c-means; k-harmonic means; ant clustering

I. INTRODUCTION

Many real-world applications are really ambiguous and cannot be exclusively clustered into distinct classes. To compensate for these artifacts, soft clustering methods, fuzzy c-means (FCM) algorithm [1], has recently been used extensively with some success in the fuzzy clustering areas [2-4]. The objective regarding FCM is to group data, based on arithmetic means into set of disjointed clusters. The data within the same clusters are highly similar with one another and dissimilar with those in other clusters. The strength of FCM over traditional clustering such as k-means (KM) is that it allows one piece of data to belong to two or more clusters. Given an input point, FCM yields the degree of membership value in each cluster.

The other presented clustering method, k-harmonic means (KHM) algorithm is an algorithm proposed by Zhang, Hsu, and Dayal [5], Zhang [6] and modified by Hammerly and Elkan [7]. Instead of using an arithmetic mean, k-harmonic clustering uses the harmonic average. Moreover, KHM associates the influence weight of a single data on the cluster center in the following iterations. FCM as well as KHM retain more information from the original data than those of crisp or hard.

There still exists a question on the causes of local optima traps problems, occurring in FCM as well as KHM learning. In order to avoid such local traps, many researches concentrate on optimizing initial set of clusters before performing further clustering [8]. Among these, the research [9] applies two-level clustering based upon ant colony systems and the soft clustering, FCM and KHM. Such clustering algorithm takes the merit of the ability of ant colony algorithm on searching for the good clustering, in order to overcome the chance of getting stuck in local minima. The ants move data objects in 2-D feature space to find a good set of clusters [10]; then such clusters set is supplied to either FCM or KHM for reformulating as well as refining the clustering afterwards. The results clearly demonstrate that both ANT-FCM and ANT-KHM obtain more acceptable results than one-level of soft clustering using FCM or KHM. The later version of ANT-FCM fuzzy [11-12] the ants move the cluster centers, not the data objects to relocates the cluster centroids in the feature space. A particular partition that is associated with optimal set of cluster centers is discovered. The latter algorithm does not use the object merging criterion, which makes it independent of the threshold for merging the data objects. In addition, there are less controlling parameters than the previous ant based clustering algorithms. However, this version of ant clustering uses some unnecessary numbers of memories in the clusters partition selection process.

Nevertheless, such two-level clustering, either ANT-FCM or ANT-KHM may not significantly improve clustering precision; contrarily it usually generates abundant runtime consumption. Several real-world applications have usually been suffered from the problems of high dimensional, noisy data and outliers. Such problems seriously spoil the computation of several types of learning, including clustering. Irrelevant dimensional features deteriorate the generalization

performance of clustering. A linear feature extraction method, principal component analysis (PCA) is one of the important tools for coping with such dimensionality problems [13]. In order to perform dimension reduction, PCA maps the original predicting features into smaller numbers of features. Thereby, applying PCA [14] as preprocess for dimension reduction would lead to the improvement of the clustering efficiency. Previously, the aforementioned two-level clustering and feature reduction in clustering areas have been separately explored. The comparison study among them has not yet been fulfilled.

This paper presents a performance comparison of the approaches using principal component analysis (PCA) and those using ant clustering algorithm with fuzzy clustering. The performance is evaluated in terms of clustering objective function, adjusted rand index (ARI) [15] and time consumption. The consequences of PCA-FCM and PCA-KHM are compared with ANT-KHM and ANT-FCM. Here, the algorithms in the ant clustering parts of ANT-FCM and ANT-KHM does not need to consume the unnecessary memories for the clusters partition selection process; instead it employs Boltzmann probability [16] distribution to alleviate such cost of memories. The rest of the paper is organized as follows. Section 2 introduces FCM and KHM clustering. Section 3 and 4 respectively describe ant clustering technique and PCA preprocessing that supply the data in a new dimensional space to FCM and KHM. Then, experimental results are determined in section 5. Finally, conclusions are made in section 6.

II. FUZZY C-MEANS (FCM) AND K-HARMONIC MEANS CLUSTERING (KHM)

Fuzzy C-Means (FCM) is a clustering method that allows a data point to belong to two or more clusters with different degrees of membership; FCM is based on minimization of the following objective function:

$$\sum_{i=1}^N \sum_{c=1}^C u_{ic}^m \|x_i - \bar{x}_c\|^2 \quad (1)$$

where, $\mathbf{X} = \{x_1, \dots, x_i, \dots, x_N\}$, x_i is the i^{th} of d -dimensional measured data; is a set of data to be clustered \bar{x}_c is a c^{th} cluster centers, where $c = 1, 2, \dots, C$. m , fuzziness degree controls the extent of membership sharing between fuzzy clusters; here it equals 2, u_{ic} is the degree of membership of input x_i in the cluster c . $\|\cdot\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown in (1). The update of membership u_{ij} and the cluster centers \bar{x}_c follow (2) and (3) consecutively:

$$u_{ij} = \left(\sum_{c=1}^C \frac{\|x_i - \bar{x}_j\|^{\frac{2}{m-1}}}{\|x_i - \bar{x}_c\|} \right)^{-1} \quad (2)$$

$$\bar{x}_j = \frac{\sum_{c=1}^C u_{nj}^m x_i}{\sum_{n=1}^N u_{nj}} \quad (3)$$

This iteration will stop when:

$$\max\{|u_{ic}^{iter+1} - u_{ic}^{iter}|\} < \varepsilon$$

where, ε is a termination criterion ranged between 0 and 1 and superscript $iter$ is the iteration number. However, the problem of getting into local optima still exists in FCM learning.

The other soft clustering which is used in this work is K-harmonic means clustering (KHM). KHM also applies degrees of membership to allow each data point to belong to two or more clusters [7] [17-19], similar to FCM. However, in KHM, the arithmetic mean of distance from a data point to the centers, used in FCM, is replaced by the harmonic mean. The harmonic means gives a good (low) score for each data point when that data point is close to any one center. This is a property of the harmonic means, which is a smooth differentiable function. The following notations are used to formulate the KHM algorithm.

$p(c/x_i)$: the membership function defining the proportion of data point x_i that belongs to center c .

$w(x_i)$: the weight function defining how much influence data point x_i has in re-computing the center parameters in the next iteration.

The basic algorithm for KHM clustering is shown as follows:

1. Randomly choose the initial C centers.
2. Calculate the objective function value according to

$$KHM(\mathbf{X}, C) = \sum_{i=1}^N \frac{C}{\sum_{c=1}^C \frac{1}{\|x_i - \bar{x}_c\|^m}}, \quad (4)$$

where m is equivalent to 2, like that in FCM.

3. For each data point x_i , compute its membership $p(c/x_i)$ in each center cluster c according to (5)

$$p(c/x_i) = \frac{\|x_i - \bar{x}_c\|^{-m-2}}{\sum_{c=1}^C \|x_i - \bar{x}_c\|^{-m-2}}, \quad (5)$$

4. For each data point x_i , compute its weight $w(x_i)$ according to (6)

$$w(x_i) = \frac{\sum_{c=1}^C \|x_i - \bar{x}_c\|^{-m-2}}{(\sum_{c=1}^C \|x_i - \bar{x}_c\|^{-m})^2}, \quad (6)$$

5. For each center cluster c , re-compute its location from all data points x_i according to their memberships and weights as show in (7)

$$\bar{x}_c = \frac{\sum_{i=1}^N p(c/x_i) w(x_i) x_i}{\sum_{i=1}^N p(c/x_i) w(x_i)}, \quad (7)$$

6. Repeat steps 2-5 until reaching a predefined number of iterations or until $KHM(\mathbf{X}, C)$ does not change significantly.

7. Assign a data point x_i to a cluster c with the biggest $p(c/x_i)$.

III. USING ANT CLUSTERING WITH FCM AND KHM (ANT-FCM AND ANT-KHM)

The ant clustering used in ANT-FCM and ANT-KHM, in this comparative study are based upon the former one, proposed by Kanade and Hall [12]. The ants co-ordinate to move cluster centers in feature space to search for optimal cluster centers. Initially, the feature values are normalized between 0 and 1. Each ant is assigned to a particular feature of a cluster in a partition. The ants never change the feature, cluster or partition assigned to them. After randomly moving the cluster centers for a fixed number of iterations, called an epoch, the quality of the partition is evaluated. According to the former ant clustering, particular clusters partition is selected as the best one based on a certain cascading rules of probability. A certain additional memories are exploited for storing some numbers of good clusters. However, the ant clustering used in this paper selection does not need the additional memories to store more than one good cluster centers; instead, it applies the heuristic technique, based on Boltzmann probability [16]. Such technique supports the exploration of the good clusters. Similar to the former ant clustering, there are two directions for the random movement of the ant. The positive direction is when the ant is moving in the feature space from 0 to 1, and the negative direction is when the ant is moving in the feature space from 1 to 0. If during the random movement the ant reaches the end of the feature space the ant reverses the direction. After a fixed number of epochs or if there exists less than 0.001 distance between previous and current objective values, the ants stop. The cluster centers, finally obtained are then used as the initial cluster centers for FCM or KHM.

IV. USING PRINCIPAL COMPONENT ANALYSIS WITH FCM AND KHM FOR DIMENSION REDUCTION (PCA-FCM AND PCA-KHM)

PCA is an orthogonal basic transformation (Chen, Billings & Luo, 1989). Given a data set: $\{\mathbf{x}_i \in \mathbf{R}^D | i = 1, \dots, N\}$, where D is the number of dimensions, N refers to the samples size. $\mathbf{Y} = (y_1, \dots, y_n)$ is given as a centered matrix; $y_i = \mathbf{x}_i - \bar{\mathbf{x}}$, where $\bar{\mathbf{x}} = \sum_{i=1}^N \mathbf{x}_i / N$. After data centralization, The covariance matrix is found by $\mathbf{M} = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \mathbf{Y}\mathbf{Y}^T$. In order to find the principal components in the new feature space, where noise and irrelevant data is eliminated, one has to find eigenvalues λ_p and eigenvectors \mathbf{v}_p of \mathbf{M} , satisfying $\lambda_p \mathbf{v}_p = \mathbf{M}\mathbf{v}_p$. \mathbf{v}_p represent the orthogonal vectors in noise-free feature on which \mathbf{Y} would be projected. The eigenvalues λ_p refer to diagonal elements in a diagonal matrix. All the other values in such matrix are zeros. The size of eigenvalues equals the amount of variance in the direction of the corresponding eigenvectors \mathbf{v}_p where $p = 1, 2, \dots, P, P \leq D$. The directions of the first eigenvectors, \mathbf{v}_p corresponding to the biggest eigenvalue, λ_1 that covers as much variance as possible. The basis principal eigenvectors $\mathbf{v}_p / \sqrt{\lambda_p}$ of \mathbf{M} are the principal directions of \mathbf{Y}^T . The principal eigenvectors \mathbf{u}_p of \mathbf{M} are called principal components if the entries of each \mathbf{u}_p are the

projected values of data points on the principal direction \mathbf{v}_p . The principal components \mathbf{u}_p are calculated based on the basis principal eigenvectors $\mathbf{v}_p / \sqrt{\lambda_p}$ via the coordinates in the eigenvector basis is called *principal components*. In PCA, one has to find eigenvalues λ_p and eigenvectors \mathbf{v}_p of \mathbf{M} , satisfying $\lambda_p \mathbf{v}_p = \mathbf{M}\mathbf{v}_p$. The size of each eigenvalue λ_p equals the amount of variance in the direction of the corresponding eigenvectors \mathbf{v}_p where $p = 1, 2, \dots, P, P \leq D$. The directions of the first eigenvectors corresponding to the biggest eigenvalues cover as much variance as possible by P orthogonal directions. The principal eigenvectors \mathbf{v}_p of \mathbf{M} are the principal directions of $(\mathbf{x}_i - \bar{\mathbf{x}})$. Let \mathbf{u}_p of \mathbf{M} are the principal components; Entries of each \mathbf{u}_p are the projected values of data points on the principal direction \mathbf{v}_p . \mathbf{v}_p and \mathbf{u}_p are related via (8) and Fig. 1.

$$\mathbf{u}_p = \mathbf{Y}^T \mathbf{v}_p / \sqrt{\lambda_p} \quad (8)$$

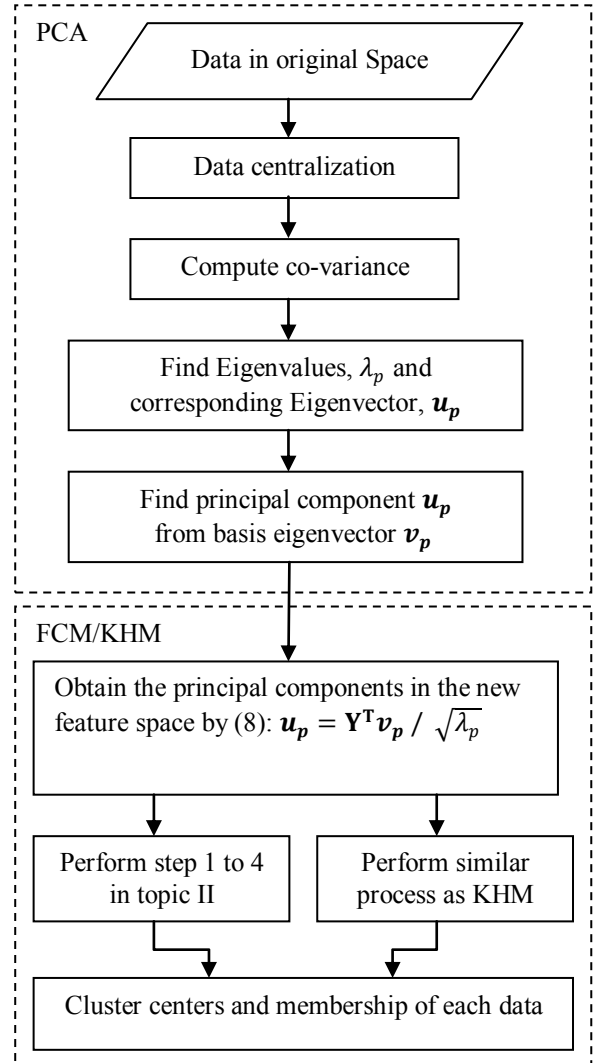


Figure 1. Flow chart of PCA-FCM and PCA-KHM

Through such a PCA method, the main P of D dimensions are extracted; whilst noisy and irrelevant dimensional features that could seriously deteriorate the generalization performance of clustering are eliminated. The PCA is utilized in this paper as a preprocess that generates dimension reduction data with noisy decreased. Such refined data is later employed in FCM and KHM learning process.

V. EXPERIMENTAL AND RESULTS

PCA-FCM and PCA-KHM are tested on seven benchmark medical data sets obtained from the URL, “<http://archive.ics.uci.edu/ml/datasets.html>”. The characteristics of the tested data sets are summarized in Table I.

TABLE I. CHARACTERISTICS OF DATA SETS CONSIDERED.

Name of data set	No. of classes	No. of features	Size of data set (<i>size of classes in parentheses</i>)
Pima Indians Diabetes	2	8	768 (500, 268)
Parkinson	2	22	195 (48, 147)
Lymphography	4	18	148 (2, 67, 46, 33)
Hepatitis	2	19	155 (32, 123)
Dermatology	6	34	366 (112, 61, 72, 49, 52, 20)
Contraceptive	3	9	1473 (629, 333, 511)
Breast Tissue	6	9	106 (21, 15, 18, 16, 14, 22)

The performance comparison is performed on such two algorithms and ANT-FCM and ANT-KHM as well as other related: FCM and KHM alone. According to PCA feature reduction, the number of orthonormal eigenvectors, corresponding to the first 90% (threshold of selection) of accumulation of the largest eigenvalues of the covariance matrix would be used in the feature space. For example, if the first, second and third largest percentages of variance yield 56%, 42% and 2%; only the first and second would be selected by the reason of the accumulation of 56% and 42% > 90%. Such a selection criterion of eigenvector is applied for all data sets and related methods. The quality of the respective clustering are compared, where the quality is measured by two criteria. One is the objective function FCM and KHM. Such functions refer to the sum. The other one, ARI is calculated based on the following procedures: suppose T is the true clustering of a data set based on domain knowledge and R a clustering result given by a clustering algorithm. Let a , b , c , and d , respectively, denote the number of pairs belonging to the same cluster in both T and R , the number of pairs belonging to the same cluster in T but to different clusters in R , the number of pairs belonging to different clusters in T but to the same cluster in R and the number of pairs belonging to different clusters in both T and R . The ARI $ARI(T, R)$ is then defined as follows:

$$ARI(T, R) = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)} \quad (9)$$

The value of $ARI(T, R)$ lies between zero and one and higher value indicates that R is more similar to T . In addition, $ARI(T, T) = 1$.

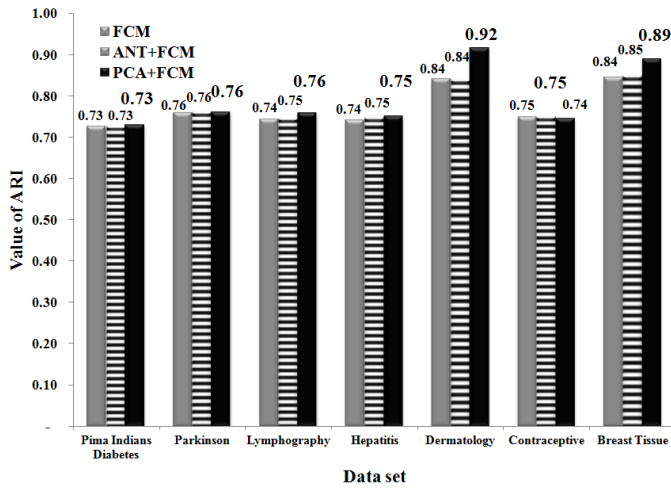
In Fig 2 (a) and (b), ANT-FCM as well as ANT-KHM seems to generate competitive results with PCA-KHM and PCA-FCM consecutively in terms of ARI-based measure. Nevertheless, the average ARI values yielded by PCA-KHM and PCA-FCM are denoted as 0.79 and 0.79, compared to 0.77 and 0.77 of ANT-KHM and ANT-FCM. Moreover, a visual analysis of runtime comparisons can be determined in a form of bar graphs in Fig. 2 (c) and (d). The comparison shows the unrivaled improvement of runtime, yielded by PCA-FCM and PCA-KHM. The prominent decrease of runtime is explicitly caused by the dimension reduction as a result of PCA-FCM and PCA-KHM. Contrarily, an abundant multiplication of runtimes is produced by ANT-FCM and ANT-KHM. Table II demonstrates very close results among

KHM and ANT-KHM, and FCM and ANT-FCM. The highest difference between PCA-KHM and ANT-KHM is approximately 10^{10} times ANT-KHM; while that between PCA-FCM and ANT-FCM is about 10^8 times ANT-FCM. On the other hand, the lowest difference between PCA-KHM and ANT-KHM is approximately 100 times ANT-FCM; while that PCA-FCM between PCA-FCM and ANT-FCM is about 3 times ANT-FCM.

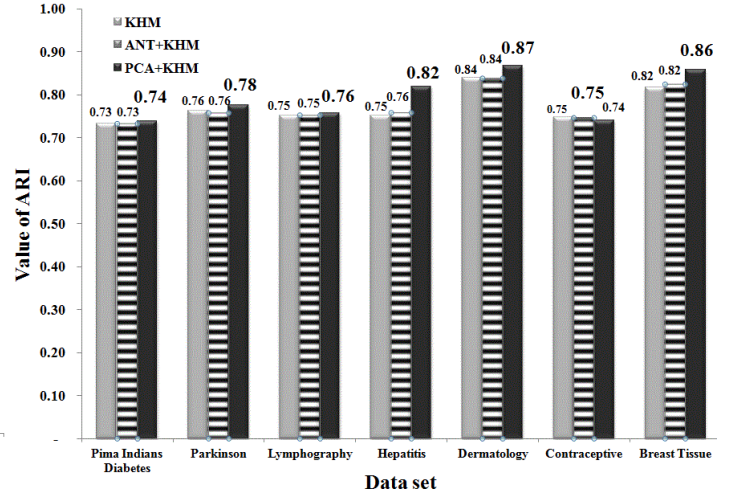
In Fig. 3, the dimension reduction is visualized in a form of bar graph. The gray bars represent the original number of data; the original number of dimensions is illustrated above such gray bars. The dark bars indicate the reduced data dimension, as consequence of PCA. The ratio (%) of dimensions after reduction to those before reduction for each data set is noted above the dark bars as well. The graphs indicate the outstanding of dimension reduction. For example, 94.74% ($100\% - 5.26\%$) is the highest reduction PCA regarding Hepatitis; and 82.38 is the average one. Albeit the high rate of dimension reduction, the superior results still maintain.

CONCLUSION

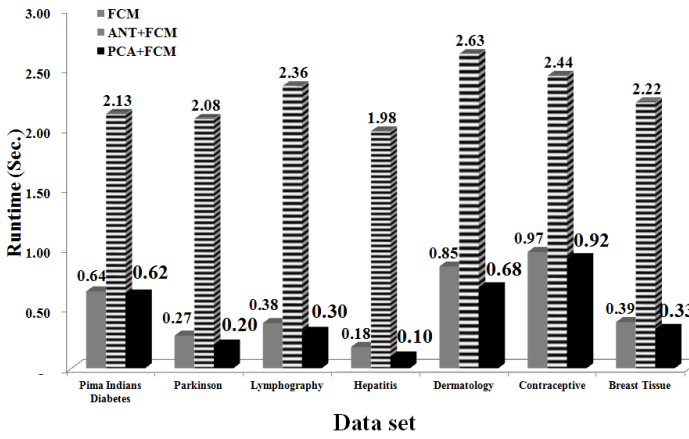
This paper proposes a performance comparison using feature reduction based method, principal component analysis (PCA) and ant clustering algorithm with fuzzy clustering, fuzzy c-means (FCM) and K-harmonic means (KHM). In the comparison, the results of PCA-FCM, PCA-KHM and those of ANT-FCM, ANT-KHM as well as FCM and KHM alone are examined. The tests are performed on 7 well-known benchmark real-world data sets. The performance measurements criteria FCM and KHM objective function, ARI and runtime are calculated over 10 averages independent runs. The results based on runtime and clustering objective function indicate prominent superiority of the PCA-FCM and PCA-KHM over ANT-FCM and ANT-KHM, although the ARI values are not much different. It is also seen that ANT-FCM and ANT-KHM are more complicate than the other algorithms; in addition, the obvious dimension reduction, which is the consequence of performing PCA causes the decrease number of iterative clustering process. Nevertheless, the algorithms require a priori known number of clusters; thus, are not applicable when the number of clusters is unknown. In order to improve the flexibility of the clustering approaches, the future



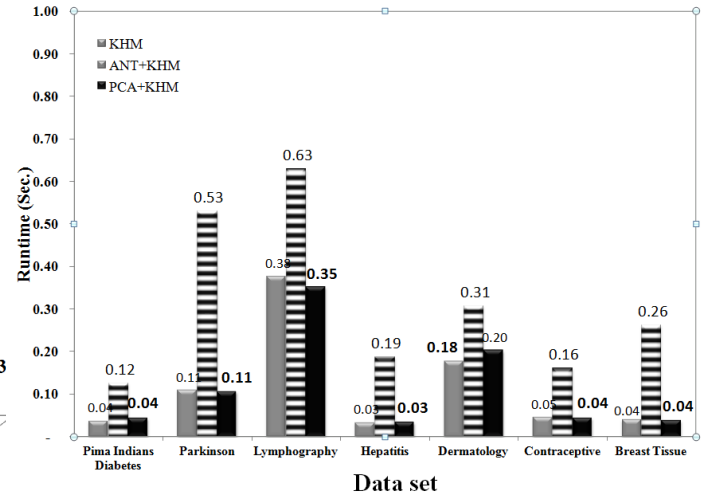
(a)



(b)



(c)



(d)

Figure 2. The measurement (a) ARI value of FCM, ANT-FCM and PCA-FCM, (b) ARI value of KHM, ANT-KHM and PCA-KHM. The measurement (c) Runtime of FCM, ANT-FCM and PCA-FCM, and (d) Runtime of KHM, ANT-KHM and PCA-KHM.

TABLE II. OBJECTIVE VALUES OF FCM AND KHM

Source	KHM Values			FCM Objective function		
	<i>KHM</i>	<i>ANT+KHM</i>	<i>PCA+KHM</i>	<i>FCM</i>	<i>ANT+FCM</i>	<i>PCA+FCM</i>
Pima Indians Diabetes	2.03E+06	2.04E+06	7.40E+00	1.77E+06	1.76E+06	9.47E+01
Parkinson	4.35E+04	4.35E+04	5.63E+01	3.78E+05	3.79E+05	6.17E+01
Lymphography	5.35E+04	5.36E+04	4.43E+02	6.11E+01	6.11E+01	1.85E+01
Hepatitis	2.73E+10	2.75E+10	5.01E+00	3.83E+05	3.84E+05	2.13E+01
Dermatology	8.24E+06	8.15E+06	3.77E+01	7.64E+02	7.64E+02	1.14E+01
Contraceptive	1.94E+06	1.98E+06	2.15E+01	5.03E+03	5.01E+03	3.79E+02
Breast Tissue	1.85E+03	1.85E+03	4.41E+01	3.61E+08	3.68E+08	1.72E+00

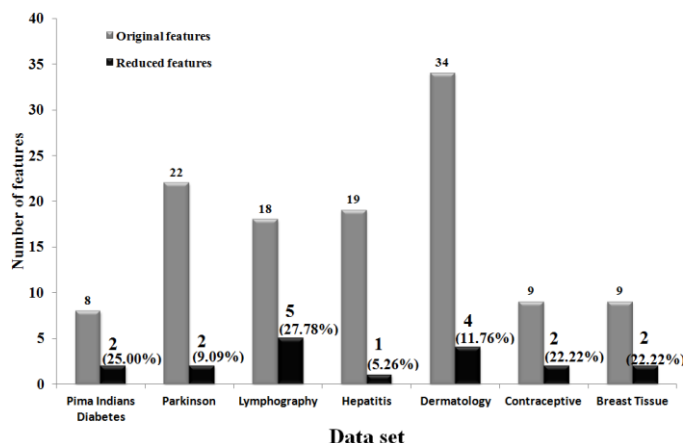


Figure 3. The dimension reduction on seven medical data sets, as a result of PCA preprocessing on FCM and KHM. The gray and dark bars represent the original number of data dimensions and those of the reduced. The percentages of dimension reduction for each data set are pointed out on the dark bars.

works may apply the combination with some types of hierarchical clustering methods such as those of agglomerative or divisive.

REFERENCES

- [1] J. C. Bezdek, R. Ehrlich and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 3 no. 10, pp. 191-203, 1984.
- [2] M. A. Balafar, A. R. Ramli, M. I. Saripan, R. Mahmud and S. Mashohor, "Medical image segmentation using fuzzy c-mean (FCM), learning vector quantization (LVQ) and user interaction," *Advanced Intelligent Computing Theories and Applications With Aspects of Contemporary Intelligent Computing Techniques*, vol. 15, no. 5, pp. 177-184, 2008.
- [3] W. Chen, M. L. Giger and U. Bick, "A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images," *Academic Radiology*, vol. 13, no. 1, pp. 63-72, 2006.
- [4] S. R. Kannan, S. Ramathilagam, A. Sathya and R. Pandiyarajan, "Effective fuzzy c-means based kernel function in segmenting medical images," *Computers in Biology and Medicine*, vol. 40, no. 6, pp. 572-579, 2010.
- [5] B. Zhang, M. Hsu and U. Dayal, "K- harmonic means -A data clustering algorithm," *Hewlett Packard Research Laboratory Technical Report PL1999124*, Citeseer, 1999.
- [6] B. Zhang, "Generalized k-harmonic means- Boosting in unsupervised learning," *Technical Report TR HPL-2000-137*, Hewlett Packard Labs, Palo Alto, CA. 2000.
- [7] G. Hamerly and C. Elkan, "Alternatives to the k-means algorithm that find better clusterings," *Proceedings of the eleventh international conference on Information and knowledge management CIKM 02*, pp. 600-607, 2002.
- [8] Y. Kao, J. Lin and S. Huang, "Fuzzy clustering by differential evolution. In Intelligent Systems Design and Applications," *ISDA '08. Eighth International Conference*, vol. 1, pp. 246-250, 2008.
- [9] J. Hua, Y. Shenghe, L. Jing, Y. Fengqin and H. Xin, "Ant clustering algorithm with K-harmonic means clustering," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8679-8684, 2010.
- [10] P. M. Kanade and L. O. Hall, "Fuzzy ants as a clustering concept," *Fuzzy Information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American*, pp. 227- 232, 2003.
- [11] P. M. Kanade and L. O. Hall, "Fuzzy ant clustering by centroid positioning," *In IEEE International Conference on Fuzzy Systems, Budapest, 2004. IEEE Computer Society*, pp. 371-376, 2004.
- [12] P. M. Kanade and L. O. Hall, "Fuzzy Ants and Clustering," *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, vol. 37, no. 5, pp. 758-769, 2007.
- [13] M. Andrzej and R. Waldemar, "Principal Components Analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303-342, 1993.
- [14] I. T. Jolliffe, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37-52, 1986.
- [15] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Patter Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841-847, Aug. 1991.
- [16] C. Y. Lee, "Entropy-Boltzmann selection in the genetic algorithms," *IEEE transactions on systems man and cybernetics Part B Cybernetics a publication of the IEEE Systems Man and Cybernetics Society*, vol. 33, no. 1, pp. 138-149, 2003.
- [17] Z. Gungor and A. Unler, "K-harmonic means data clustering with simulated annealing heuristic," *Applied Mathematics and Computation*, vol. 184, no. 2, pp. 199-209, 2007.
- [18] Z. Gungor and A. Unler, "K-harmonic means data clustering with tabu-search method," *Applied Mathematical Modelling*, vol. 32, no. 6, pp. 1115-1125, 2008.
- [19] S. Chen, S. A. Billings and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873-1896, 1989.