© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie

The International Journal of Integrated Engineering

A Novel K-means-based Feature Reduction

Mahboubeh Haghayeghipour[1], Yahya Forghani[1,*] S. Mohammad Hosein Moattar

[1]Islamic Azad University, Mashhad branch,
Ghasemabad St., Mashhad, 9999, IRAN

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.00.00.0000.00.0000
Received 15 June 2018; Accepted 13 January 2019; Available online 30 April 2019

Abstract: The aim of feature reduction is reduction of the size of data file, elimination of irrelevant features, and discovery of the effective data features for data analysis. Irrelevant data features can skew data analysis such as data clustering. Therefore, maintaining the data structure or data clusters must be taken into consideration in feature reduction. In this article, with regard to the success of k-means-based clustering methods, a feature reduction method is presented based on weighted k-means (wk-means). More specifically, firstly, data features are weighted using wk-means method. A feature with a high weight is not a better feature for clustering than a feature with a low weight, necessarily, and the weight of a feature only change feature range for better clustering. Then, by using a novel mathematical model, a group of weighted features with the least effect on data clusters are eliminated and the remaining features are selected. Contrary to sparse k-means method, the number of selected features can be determined explicitly by the user in our proposed method. Experimental results on four real datasets show that the accuracy of clusters obtained by wk-means after feature reduction by the proposed method is better than that of sparse k-means, PCA and LLE.

Keywords: Feature Reduction, Clustering, K-means, Weighted K-means.

1. Introduction

In general, there are two feature reduction methods: feature extraction [1, 2, 3, 4, 5, 6, 7, 8] and feature selection [9, 3, 10, 11, 12, 13, 14, 15, 16, 17,18]. In the former method, a fraction of features is selected as effective features, whereas in the latter technique, a number of effective features, each one as a function of one or more features, are produced. One of the feature extraction methods is principle component analysis (PCA) [1]. In this method, the best extracted feature is a feature with the maximal variance and a feature with a variance close to zero will be regarded as a trivial feature or noise. Another feature extraction method is locally linear embedding (LLE) [4] which consists of two phases: In the first phase, each data is written as a linear combination of other data, and the coefficients of this linear combination are stored. After that, low-dimensional data are produced in a way that each low-dimension data can be written as linear combination of the other low-dimensional data with the same coefficients stored in the previous phase. Simultaneous Orthogonal basis Clustering Feature Selection (SOCFS) [9] is a regularized regression based formulation with a new type of target matrix for unsupervised feature selection method. The mentioned target matrix captures latent cluster centers of the projected data by performing the orthogonal basis clustering, and then leads the projection matrix to select some discriminative features.

The aim of feature reduction is reduction of the size of data file, elimination of irrelevant features, and discovery of the effective data features for data analysis. Irrelevant features can skew data analysis. Those features are effective

**Corresponding author: yahyafor2000@yahoo.com*

features which can properly maintain the structure of the data such as the existing data clusters. Data structure or data clusters must not change during feature reduction. In sparse k-means [10, 15], in order to maintain data clusters as more as possible, those features are selected which decrease intra-cluster distance of the data compared to pair-wise distance between pairs of data. Unfortunately, the number of selected features can not be determined explicitly by the user in this method.

In this article, with regard to the success of k-means-based clustering methods [19, 20, 21], a feature reduction method is presented based on wk-means [22]. More specifically, firstly, data features are weighted using wk-means method. A feature with a high weight is not a better feature for clustering than a feature with a low weight, necessarily, and the weight of a feature only change feature range for better clustering. Then, by using a novel mathematical model, a group of weighted features with the least effect on data clusters are eliminated and the remaining features are selected as the output of the proposed feature reduction method. Contrary to sparse k-means, the number of selected features is explicitly determined by the user in our proposed method. Experimental results on four real datasets show that the accuracy of clustering by wk-means after feature reduction by the proposed method is better than that of sparse k-means, PCA and LLE.

In continue, the prerequisites of the research are explained in section 2. The proposed feature reduction method is presented in section 3. By using four real datasets, the proposed method is compared experimentally with three related feature reduction methods in section 4. Finally, in section 5, conclusion is drawn.

2. Prerequisites

2.1 K-means

Consider training data set $X = \{x_1, x_2, \dots, x_n\}$ which must be grouped into c clusters, where $x_i \in R^m$, and m is data dimension. K-means clustering model is as follows:

$$\min_{u,z} \sum_{k=1}^{c} \sum_{i=1}^{n} u_{ik} \|x_i - z_k\|_2^2$$
$$\text{subject to} \begin{cases} \sum_{k=1}^{c} u_{ik} = 1, \quad i = 1,2,\dots,n; \\ u_{ik} \in \{0,1\}, \quad i = 1,2,\dots,n; k = 1,2,\dots,c. \end{cases} \tag{1}$$

where z_k is k-th cluster center, u_{ik} is the membership degree of i-th data to k-th cluster, and $\|x_i - z_k\|_2^2 = \sum_{j=1}^{m} (x_{ij} - z_{kj})^2$. The aim of the k-means model is determination of cluster centers in a way that the summation of distance between each data cluster and its corresponding cluster center is minimized. The first constraint of the k-means model states that each data must belong only to one cluster. Clustering is NP-Complete. Algorithm 1 is an iterative algorithm which tries to obtain a local optimum of the k-means model.

Algorithm 1: k-means algorithm.

1. Initialize cluster centers (z).
2. Fix cluster centers (z), and then obtain the optimal membership values (u) of k-means model.
3. Fix membership values (u), and then obtain the optimal cluster centers (z) of k-means model.
4. Repeat steps 2 and 3 until convergence condition is met.

2.2 Wk-Means

The wk-means model is as follows:

$$\min_{u,z,w} \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} w_{kj}^{\beta} (x_{ij} - z_{kj})^2$$
$$\text{subject to} \begin{cases} \sum_{k=1}^{c} u_{ik} = 1, \quad i = 1,2,\dots,n; \\ \sum_{j=1}^{m} w_{kj} = 1, \quad k = 1,2,\dots,c; \\ u_{ik} \in \{0,1\}, \quad i = 1,2,\dots,n; k = 1,2,\dots,c; \\ w_{kj} \geq 0, \quad k = 1,2,\dots,c; j = 1,2,\dots,m. \end{cases} \tag{2}$$

where w_{kj} is the weight of j-th feature of k-th cluster which is determined during the clustering or optimization process. Indeed, the optimization model (2) assigns a higher weight to a feature with smaller range. In other words, this model changes the feature ranges or the clusters shapes by feature weighting such that data to be clustered better. The hyper-parameter β controls the relation between feature weight w_{kj} and the term $(x_{ij} - z_{kj})^2$. If $\beta = 1$, only the weight of one feature of each data cluster becomes non-zero. Algorithm 2 is an iterative algorithm which tries to obtain a local optimum of the wk-means model.

Algorithm 2: wk-means algorithm.

1. Initialize cluster centers (z) and weights (w).
2. Fix cluster centers (z) and weights (w), and then obtain the optimal membership values (u) of wk-means model by using the following equation:
$$u_{ik} = \begin{cases} 1 & if\ \forall l: \sum_{j=1}^{m} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2 \le \sum_{j=1}^{m} w_{lj}^{\beta} \left(x_{ij} - z_{lj}\right)^2 \\ 0 & otherwise. \end{cases}$$
3. Fix cluster centers (z) and membership values (u), and then obtain the optimal weights (w) of wk-means model by using the following equation:
$$w_{kj} = \begin{cases} \frac{1}{m_i} & if \sum_{i=1}^{n} u_{ik}\left(x_{ij} - z_{kj}\right)^2 = 0\ and\ m_i = \left|\left\{t: \sum_{i=1}^{n} u_{ik}(x_{it} - z_{kt})^2 = 0\right\}\right|; \\ 0 & if \sum_{i=1}^{n} u_{ik}\left(x_{ij} - z_{kj}\right)^2 \ne 0\ and \sum_{i=1}^{n} u_{ik}(x_{it} - z_{kt})^2 = 0\ for\ some\ t; \\ \frac{1}{\sum_{t=1}^{m}\left[\frac{\sum_{i=1}^{n} u_{ik}\left(x_{ij} - z_{kj}\right)^2}{\sum_{i=1}^{n} u_{ik}(x_{it} - z_{kt})^2}\right]^{\frac{1}{\beta-1}}} & if \sum_{i=1}^{n} u_{ik}(x_{it} - z_{kt})^2 = 0\ for\ each\ t. \end{cases}$$
4. Fix membership values (u) and weights (w), and then obtain the optimal cluster centers (z) of wk-means model by using the following equation:
$$z_{kj} = \frac{\sum_{i=1}^{n} u_{ik} x_{ij}}{\sum_{i=1}^{n} u_{ik}}.$$
5. Repeat steps 2-4 until convergence condition is met.

3. Our proposed method

Our proposed feature reduction method has two phases. In the first phase, the data are clustered using wk-means method. Wk-means assigns a different weight to different cluster features. Therefore, it changes the range of cluster features or the cluster shapes for better clustering. In the second phase and after determining the optimal weight of each cluster feature, S features of data clusters are selected using the following model in a way that the minimum amount of change occurs in clustering or in the objective function of wk-means model:

$$\min_{u,z,\widetilde{w}} F = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2$$
$$\text{subject to} \begin{cases} \sum_{j=1}^{m} \widetilde{w}_{kj} = S, \ k = 1,2,\dots,c; \\ \sum_{k=1}^{c} u_{ik} = 1, \ i = 1,2,\dots,n; \\ \widetilde{w}_{kj} \in \{0,1\}, \ k = 1,2,\dots,c; j = 1,2,\dots,m; \\ u_{ik} \in \{0,1\}, \ i = 1,2,\dots,n; \ k = 1,2,\dots,c. \end{cases} \quad (3)$$

where w_{kj} is the j-th feature of k-th cluster which was previously obtained by wk-means model. $\widetilde{w}_{kj}$ is a feature selection variable. If $\widetilde{w}_{kj} = 1$, j-th feature of k-th cluster is selected, otherwise it is eliminated. Algorithm 3 is proposed for solving the model (3).

Algorithm 3: our proposed algorithm.

1. Initialize cluster centers (z) and feature selection variables ($\widetilde{w}$).
2. Fix cluster centers (z) and feature selection variables ($\widetilde{w}$), and then obtain the optimal membership values (u) of the model (3) by using the following equation:
$$u_{ik} = \begin{cases} 1 & if\ \forall l: g_{ik} \le g_{il}, \\ 0 & otherwise. \end{cases} \quad (4)$$
where
$$g_{il} = \sum_{j=1}^{m} \widetilde{w}_{lj} w_{lj}^{\beta} (x_{ij} - z_{lj})^2. \quad (5)$$

3. Fix cluster centers (z) and membership values (u), and then obtain the optimal feature selection variables ($\widetilde{w}$) of the model (3) by using the following equation:

$$\widetilde{w}_{kj} = \begin{cases} 1 & f_{kj} \leq S\min_{l}\{f_{kl}\}; \\ 0 & otherwise. \end{cases} \tag{6}$$

where

$$f_{kl} = \sum_{i=1}^{n} u_i w_{ik}^{\beta} (x_{il} - z_{kl})^2, \tag{7}$$

and Smin is S-th smallest value.

4. Fix feature selection variables ($\widetilde{w}$) and membership values (u), and then obtain the optimal cluster centers (z) of the model (3) by using the following equation:

$$z_{kj} = \frac{\sum_{i=1}^{n} u_{ik} x_{ij}}{\sum_{i=1}^{n} u_{ik}}. \tag{8}$$

5. Repeat steps 2-4 until convergence condition is met.

Each of the algorithms 1-3 terminates when cluster centers do not change. In continue, the correctness of each step of algorithm 3 is proved.

Theorem 1. If cluster centers (z) and feature selection variables ($\widetilde{w}$) are fixed, the optimal solution of the model (3), i.e. the optimal membership values (u), can be obtained by using Eq. (4).

Proof.

If cluster centers (z) and feature selection variables ($\widetilde{w}$) are fixed, the model (3) is transformed into the following model:

$$\min_{u} \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2$$
$$\text{subject to} \begin{cases} \sum_{k=1}^{c} u_{ik} = 1, \ i = 1,2,\dots,n; \\ u_{ik} \in \{0,1\}, \ \ i = 1,2,\dots,n; \ k = 1,2,\dots,c. \end{cases} \tag{9}$$

The model (9) can be written as the summation of n sub-models, i.e. as follows:

$$\sum_{i=1}^{n} \left(\begin{array}{l} \min_{u} \sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2 \\ \text{subject to} \begin{cases} \sum_{k=1}^{c} u_{ik} = 1; \\ u_{ik} \in \{0,1\}, \ \ k = 1,2,\dots,c. \end{cases} \end{array} \right) \tag{10}$$

Consider the i-th sub-model:

$$\min_{u} \sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2$$
$$\text{subject to} \begin{cases} \sum_{k=1}^{c} u_{ik} = 1; \\ u_{ik} \in \{0,1\}, \ \ k = 1,2,\dots,c. \end{cases} \tag{11}$$

According to the constraints of this sub-model, only and only one member of the set $\{u_{ik}\}_{k=1}^{c}$ is equal to 1 and all others are zero. Therefore, in order to minimize the objective function of the model (11), $u_{ik} = 1$ if its coefficient in the objective function of the model, i.e. $g_{ik} = \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{1j} - z_{kj}\right)^2$, is the minimum of the set $\{g_{il}\}_{l=1}^{c}$.

End of proof.

Theorem 2. If cluster centers (z) and membership values (u) are fixed, the optimal solution of the model (3), i.e. the optimal values of feature selection variables ($\widetilde{w}$), can be obtained by using Eq. (6).

Proof.

If cluster centers (z) and membership values (u) are fixed, the model (3) is transformed into the following model:

$$\min_{\widetilde{w}} \sum_{k=1}^{c} \sum_{j=1}^{m} \widetilde{w}_{kj} \sum_{i=1}^{n} u_{ik} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2$$
$$\text{subject to} \begin{cases} \sum_{j=1}^{m} \widetilde{w}_{kj} = S, \ \ k = 1,2,\dots,c; \\ \widetilde{w}_{kj} \in \{0,1\}, \ \ k = 1,2,\dots,c; j = 1,2,\dots,m; \end{cases} \tag{12}$$

The model (12) can be written as the summation of c sub-models, i.e. as follows:

$$\sum_{k=1}^{c} \left(\begin{array}{l} \min_{\widetilde{w}} \sum_{j=1}^{m} \widetilde{w}_{kj} \sum_{i=1}^{n} u_{ik} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2 \\ \text{subject to} \begin{cases} \sum_{j=1}^{m} \widetilde{w}_{kj} = S; \\ \widetilde{w}_{kj} \in \{0,1\}, \ \ j = 1,2,\dots,m; \end{cases} \end{array} \right) \tag{13}$$

Consider the k-th sub-model:

$$\min_{\widetilde{w}} \sum_{j=1}^{m} \widetilde{w}_{kj} \sum_{i=1}^{n} u_{ik} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2$$
$$\text{subject to} \begin{cases} \sum_{j=1}^{m} \widetilde{w}_{kj} = S; \\ \widetilde{w}_{kj} \in \{0,1\}, \ \ j = 1,2,\dots,m; \end{cases} \tag{14}$$

According to the constraints of this sub-model, only and only S member of the set $\{\widetilde{w}_{kj}\}_{j=1}^{m}$ is equal to 1 and all others are zero. Therefore, in order to minimize the objective function of the model (14), $\widetilde{w}_{kj} = 1$ if its coefficient in the objective function of the model, i.e. $f_{kj} = \sum_{i=1}^{n} u_i w_{ik}^{\beta}(x_{ij} - z_{kj})^2$, is one of the S minimum values of the set $\{f_{kl}\}_{l=1}^{m}$.
End of Proof.

Theorem 3. If feature selection variables ($\widetilde{w}$) and membership values (u) are fixed, the optimal solution of the model (3), i.e. the optimal cluster centers (z), can be obtained by using Eq. (8).
Proof.

If feature selection variables ($\widetilde{w}$) and membership values (u) are fixed, the model (3) is transformed into the following unconstrained model:

$$\min_{u} F = \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik} \sum_{j=1}^{m} \widetilde{w}_{kj} w_{kj}^{\beta} \left(x_{ij} - z_{kj}\right)^2 \qquad (15)$$

We have at the optimal solution of the model (15):

$$\frac{\partial F}{\partial z_{kj}} = 0 \rightarrow \sum_{i=1}^{n} -2u_{ik} w_{kj}^{\beta} \widetilde{w}_{kj}\left(x_{ij} - z_{kj}\right) = 0;$$

$$\rightarrow 2w_{kj}^{\beta}\widetilde{w}_{kj} \sum_{i=1}^{n} u_{ik} x_{ij} = 2w_{kj}^{\beta}\widetilde{w}_{kj} z_{kj} \sum_{i=1}^{n} u_{ik};$$

$$\rightarrow z_{kj} = \frac{\sum_{i=1}^{n} u_{ik} x_{ij}}{\sum_{i=1}^{n} u_{ik}}.$$

End of Proof.

4. Experimental results

In this section, the proposed feature reduction method is compared with sparse k-means, LLE and PCA by using the following real dataset of UCI repository:

- Wine: contains 3 clusters, 178 data and 13 features.
- Vertebral 2: contains 2 clusters, 310 data and 6 features.
- Vertebral 3: contains 3 clusters, 310 data and 6 features.
- Parkinson: contains 2 clusters, 195 data and 22 features.

In each experiment, one of the mentioned feature reduction methods was applied on a real datasets, then the obtained dataset was clustered using wk-means method, and finally its accuracy was reported. The hyper-parameter β of wk-means model was set to 6.

Clustering and clustering based feature reduction results depend severely to initial cluster centers which are selected randomly. Therefore, each experiment was repeated 20 times, and mean and standard deviation of accuracy was reported in Table 1. The best feature reduction method for each dataset was bolded in Table 1. Table 2 shows the clustering accuracy without the feature reduction phase.

Table 1 Accuracy of clustering by using wk-means after feature reduction by using LLE, PCA, sparse K-means, and the proposed method (%).

#of features after feature reduction	Feature reduction algorithm	Wine	Parkinsons	Vertebral3	Vertebral2
D = 2	Proposed	**0.6510±0.055**	**0.7153±0.050**	0.4935±0.008	**0.6260±0.030**
	Sparse K-means [15]	0.5618±0.010	0.6153±0.010	**0.6294±0.010**	0.5355±0.060
	LLE [4]	0.5438±0.010	0.6213±0.010	0.5416±0.013	0.6202±0.022
	PCA	0.6492±0.023	0.4821±0.011	0.4895±0.034	0.5648±0.010
D = 3	Proposed	0.6089±0.086	**0.7476±0.023**	**0.5335±0.030**	0.6106±0.016
	Sparse K-means [15]	0.4965±0.010	0.7133±0.020	0.3957±0.020	0.6710±0.020
	LLE [4]	0.4045±0.011	0.7228±0.010	0.3960±0.010	0.6883±0.042
	PCA	**0.8202±0.021**	0.7430±0.011	0.3383±0.023	**0.6961±0.030**
D = 4	Proposed	**0.6915±0.053**	**0.7676±0.020**	**0.5580±0.015**	0.6334±0.011
	Sparse K-means [15]	0.6740±0.042	0.7133±0.020	0.5330±0.030	**0.6870±0.030**
	LLE [4]	0.6817±0.043	0.7535±0.014	0.5065±0.010	0.6625±0.020
	PCA	0.6792±0.023	0.7430±0.009	0.5230±0.013	0.6355±0.020
D = 5	Proposed	0.6314±0.093	**0.7917±0.034**	0.5183±0.018	0.6464±0.011
	Sparse K-means [15]	**0.7542±0.020**	0.7233±0.010	**0.5577±0.010**	**0. 6990±0.030**

	LLE [4]	0.6604±0.010	0.7635±0.020	0.5065±0.010	0.6645±0.020
	PCA	0.6567±0.020	0.7438±0.010	0.5452±0.010	0.6355±0.020
D = 6	Proposed	**0.6679±0.054**	**0.7184±0.033**	-	-
	Sparse K-means [15]	0.6516±0.026	0.6390±0.018	-	-
	LLE [4]	0.4921±0.035	0.6902±0.037	-	-
	PCA	0.5430±0.050	0.5530±0.035	-	-

Table 2 Accuracy of clustering by using wk-means with no feature reduction (%).

Wine	Parkinsons	Vertebral3	Vertebral2
0.8998±0.030	0.6615±0.010	0.4859±0.030	0.6537±0.010

Table 3 Running time of different feature reduction methods (s).

#of features after feature reduction	Feature reduction algorithm	Wine	Parkinsons	Vertebral3	Vertebral2
D = 2	Proposed	0.058	0.077	0.048	0.040
	Sparse K-means [15]	0.025	0.036	0.038	0.027
	LLE [4]	0.080	0.168	0.111	0.094
	PCA	**0.004**	**0.004**	**0.004**	**0.004**
D = 3	Proposed	0.055	0.077	0.045	0.040
	Sparse K-means [15]	0.021	0.036	0.031	0.028
	LLE [4]	0.080	0.153	0.111	0.111
	PCA	**0.004**	**0.003**	**0.006**	**0.005**
D = 4	Proposed	0.061	0.094	0.048	0.040
	Sparse K-means [15]	0.025	0.036	0.036	0.033
	LLE [4]	0.075	0.170	0.0108	0.111
	PCA	**0.004**	**0.003**	**0.006**	**0.005**
D = 5	Proposed	0.057	0.097	0.049	0.040
	Sparse K-means [15]	0.026	0.045	0.039	0.035
	LLE [4]	0.070	0.172	0.096	0.116
	PCA	**0.006**	**0.001**	**0.007**	**0.008**
D = 6	Proposed	0.064	0.094	-	-
	Sparse K-means [15]	0.027	0.045	-	-
	LLE [4]	0.073	0.172	-	-
	PCA	**0.006**	**0.004**	**-**	-

According to Table 1, the clustering accuracy depends on the number of selected features, feature reduction method, and dataset. According to Table 1 and 2, feature reduction could sometimes enhance the clustering accuracy of Parkinson, Verteberal2 and Verteberal3 datasets. As mentioned earlier, feature reduction can enhance data analysis if there are irrelevant feature in data. Using the mean of clustering accuracies of different datasets as a criterion for comparison of different data mining methods is incorrect [23], and the ranking criterion must be used for comparison [24] as follows:
According to Table 1, the accuracy of the proposed method is better than PCA, LLE and sparse k-means for 13, 13 and 12 cases out of 18 cases, respectively. Therefore, the proposed method is more accurate than the three other methods.
The mentioned experiments were performed using 64-bit 2.5GHz Core i5 CPU with 6GB RAM. Table 3 shows the running time of the mentioned feature reduction methods. According to Table 3, it can be stated that the speed of the proposed method is better than that of LLE, whereas it is worse than that of PCA and sparse k-means.

5. Conclusion

In this article, with regard to the success of k-means-based clustering methods, a feature reduction method is presented based on wk-means. Contrary to related method such as sparse k-means and sparse fuzzy c-means methods, the number of selected features is explicitly determined by the user in our proposed method. Experimental results on four real datasets showed that:

- Feature reduction could sometimes enhance the clustering accuracy of Parkinson, Verteberal2 and Verteberal3 datasets. Indeed, feature reduction can enhance data analysis if there are irrelevant feature in data.

- The accuracy of the proposed method is better than that of PCA, LLE and sparse k-means for 13, 13 and 12 scenarios out of 18 scenarios, respectively. Therefore, the proposed method is more accurate than the three other methods.
- The speed of the proposed method is better than that of LLE, whereas it is worse than that of PCA and sparse k-means.

Our proposed model is based on k-means clustering not a fuzzy clustering model. In future, the fuzzy version of our proposed model is studied.

References

[1] I. Jolliffe, Principal Component Analysis, 2nd ed., NY: Springer, 2002.

[2] C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and k-means clustering," in *Proceedings of the 24th international conference on*, Corvallis, 2007.

[3] B. Kim, J. Shah and F. Doshi-Velez, "Mind the Gap: A Generative Approach to Interpretable Feature Selection and Extraction," in *Neural Information Processing Systems Conference*, Massachusetts Institute of Technology Cambridge, 2015.

[4] S. T. Rovis and L. K. Saul, "Nonlinear Dimentionality Reduction by Locally Linear Embedding," *Sience,* vol. 290, pp. 2323-2326, 2000.

[5] N. H. Adnan, "Feature Selection for Human Grasping Activity Using Pearson's Correlation Techniques," *International Journal of Integrated Engineering,* vol. 8, no. 1, pp. 5-10, 2016.

[6] S. Z. A. N. A. B. H. M. K. M. N. M. Siti Armiza Mohd Aris, "Statistical Feature Analysis of EEG Signals for Calmness," *INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING,* vol. 10, no. 7, pp. 22-33, 2018.

[7] J. C. M. T. S. M. N. M. N. Hannah Sofian, "Calcification Detection of Coronary Artery Disease in Intravascular Ultrasound Image: Deep Feature Learning Approach," *International Journal of Integrated Engineering* , vol. 10, no. 7, pp. 43-57, 2018.

[8] N. M. N. O. M. R. R. M. K. A. Y. Joel Than Chia Ming, "Lung Disease Classification using GLCM and Deep Features from Different Deep Learning Architectures with Principal Component Analysis," *INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING,* vol. 10, no. 7, pp. 76-89, 2018.

[9] D. Han and J. Kim, "Unsupervised Simultaneous Orthogonal Basis Clustering Feature Selection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, KAIST, South Korea, 2015.

[10] X. Qiu, Y. Qiu, G. Feng and P. Li, "A sparse fuzzy c-means algorithm based on sparse clustering frame work," *Neurocomputing,* vol. 157, pp. 290-295, 2015.

[11] K. K. Shin, "Ensemble Clustering Based On Feature Selection Approch To Learning Relation Data," Faculty of computing and informatics university Malaysia sabah, 2015.

[12] S. Wang, W. Pedrycz, Q. Zhu and W. Zhu, "Subspace learning for unsupervised feature selection via matrix factorization," *Journal of Elsevier on Pattern Recognition,* vol. 48, pp. 10-19, 2015.

[13] Y. Kong and Y. Deng, "Discriminative Clustering and Feature Selection for Brain MRI Segmentation," *Jornal of IEEE on SIGNAL PROCESSING LETTERS,* vol. 22, no. 5, pp. 573-578, 2015.

[14] P. Maji and S. Roy, "Rough-Fuzzy Clustering and Unsupervised Feature Selection for Wavelet Based MR Image Segmentation," *Journal of Plos One,* 2015.

[15] R. Tibshirani and D. M. Witten, "A framework for feature selection in clustering," *J Am Stat Assoc.,* vol. 105, no. 490, p. 713–726, 2010.

[16] M. F. N. X. C. Y. Y. A. G. H. Q. Z. Luo, "Adaptive unsupervised feature selection with structure regularization," *IEEE transactions on neural networks and learning systems* , vol. 29, no. 4, pp. 944-956, 2018.

[17] W. X. Z. G. W. Y. Z. H. Y. J. G. Zheng, "Unsupervised feature selection by self-paced learning regularization," *Pattern Recognition Letters,* 2018.

[18] X. S. Z. R. H. Y. Z. Zhu, "Local and global structure preservation for robust unsupervised spectral feature selection," *IEEE Transactions on Knowledge and Data Engineering,* vol. 30, no. 3, pp. 517-529, 2018.

[19] X. Huang, Y. Ye and H. Zhang, "Extensions of Kmeans-Type Algorithms: A New Clustering Framework by Integrating Intracluster Compactness and Intercluster Separation," *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,* vol. 25, no. 8, pp. 1433 - 1446, 2014.

[20] K. C. F. C. a. S. W. Z. Deng, "Enhanced soft subspace clustering integrating within-cluster and between-cluster information," *Pattern Recognition,* vol. 43, p. 767–781, 2010.

[21] A. S. S. A. D. Gujar, "A new framework of Kmeans algorithm by combining the dispersions of clusters," *International Journal of Advance Research in Computer Science and Management Studies,* vol. 4, pp. 73-78, 2016.

[22] I. W. T. a. J. T. K. K. Zhang, "Automated variable weighting in k-means type clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* p. 657–668, 2005.

[23] J. Demsar, "Statistical comparison of classifier over multiple data sets," *Journal of Machine Learning,* vol. 7, pp. 1-30, 2006.

[24] S. García, A. Fernández and J. Luengo, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Information Sciences,* vol. 180, pp. 2044-2064, 2010.