

K-means based Clustering Method with a Fixed Number of Cluster Members

Faliu Yi[†], Inkyu Moon^{††}

ABSTRACT

Clustering methods are very useful in many fields such as data mining, classification, and object recognition. Both the supervised and unsupervised grouping approaches can classify a series of sample data with a predefined or automatically assigned cluster number. However, there is no constraint on the number of elements for each cluster. Numbers of cluster members for each cluster obtained from clustering schemes are usually random. Thus, some clusters possess a large number of elements whereas others only have a few members. In some areas such as logistics management, a fixed number of members are preferred for each cluster or logistic center. Consequently, it is necessary to design a clustering method that can automatically adjust the number of group elements. In this paper, a k-means based clustering method with a fixed number of cluster members is proposed. In the proposed method, first, the data samples are clustered using the k-means algorithm. Then, the number of group elements is adjusted by employing a greedy strategy. Experimental results demonstrate that the proposed clustering scheme can classify data samples efficiently for a fixed number of cluster members.

Key words: Clustering Algorithm, K-means Algorithm, Logistic Management, Image Processing

1. INTRODUCTION

Rapid development in information technology in the last few decades has led to increased digital data which are widely used in many fields such as medical science, military, banking, manufacture, and logistics [1-6]. Consequently, automatically processing a large amount of information is essential. Thus far, many information-processing algorithms have been proposed and considerably benefited people [7-8]. Among these algorithms, clustering methods, which are one of popular techniques, are being widely used in data mining, object classification, pattern recognition, machine learning, image analysis, information retrieval and bioinformatics [9-15]. There are two types of clus-

tering approaches: the supervised algorithm and the unsupervised algorithm [16]. For supervised clustering, the category of each sample should be known in advance, whereas this is not required for the unsupervised clustering scheme.

However, to the best of our knowledge, the number of cluster members are random for all existing clustering methods, i.e, the number of cluster members is dependent on the clustering criteria such as statistical inference, similarity measurement, and density [17-18]. There is no adjustment for the number of cluster elements. Indeed, adjustments for cluster elements are not necessary in most of the cases. Nevertheless, in some fields such as logistical management, it would be necessary to adjust cluster members to ensure that the

※ Corresponding Author : Inkyu Moon, Address: (501-759) 309 Pilmum-daero, Dong-gu, Gwangju, South Korea, TEL : +82-62-230-6033, FAX : +82-62-230-6033, E-mail : inkyu.moon@chosun.ac.kr

Receipt date : June 18, 2014, Revision date : Aug. 13, 2014
Approval date : Aug. 29, 2014

[†] Department of Computer Engineering, Chosun University, 309 Pilmum-daero, Dong-gu, Gwangju 501-759, South Korea
(E-mail : yifaliu@hotmail.com)

^{††} Department of Computer Engineering, Chosun University, 309 Pilmum-daero, Dong-gu, Gwangju 501-759, South Korea

number of cluster elements satisfies the predefined requirement. For example, when a logistics company intends to schedule its logistics center and demands each logistics center to be responsible for a fixed number of cities, the conventional clustering algorithm cannot achieve its purpose. On the other hand, a clustering method with a fixed number of cluster members would be suitable to satisfy this type of requirement.

In this paper, we propose a k-means based clustering method with fixed number of cluster members. First, the data samples or patterns extracted from objects are grouped into k categories using the k-means clustering algorithm [17-18]. Then, the number of cluster members is adjusted to satisfy the required number using the greedy strategy [19]. The adjustment procedure is the key step for the proposed clustering method and is explained in detail in section 3. In addition, we show improvement of the work present in [27]. The biggest difference between the proposed method and that in [27] is the adjustment order for the original clusters obtained with k-means algorithm. In [27], the cluster with minimum members is firstly selected to adjust its members at each iteration. In this paper, however, we adjust the cluster that its cluster center point is on the convex hull which contain all of the unadjusted cluster center points at each iteration. When there are only two clusters, the two methods would lead to the same results. But, the proposed method will be superior to that in [27] when more than two clusters are existed. We demonstrate that our method can achieve good results when the number of cluster members is pre-defined and can result in better performance than that in [27].

This paper is organized as follows. In Section 2, we describe the conventional k-means clustering algorithm. In Section 3, we describe the procedure for the proposed clustering method with a fixed number of cluster elements. In Section 4, we present the experimental results. The conclusion is

presented in Section 5.

2. K-MEANS CLUSTERING ALGORITHM

K-means [17-19] is a robust clustering method and has been successfully used in image segmentation, object classification, pattern recognition, data mining and logistics management [20-24]. Hugo Steinhaus first presented the idea of k-means in 1957 and the standard k-means algorithm was proposed by Stuart Lloyd in 1957. However, the term “k-means” was first used by James MacQueen in 1967 [25-27]. It is one of the simplest yet robust deterministic clustering algorithms, which aims to partition N observations or data samples into k user-defined clusters where each observation belongs to the cluster with the nearest mean that is regarded as a prototype of the corresponding cluster [27-28]. The flowchart of the well known k-means approach achieved by Stuart Lloyd is presented in Fig. 1.

The flowchart in Fig. 1 can be explained in detail as follows: First, the number of clusters k has to be defined. Then, k points or observations are randomly chosen as the centroid points of the k

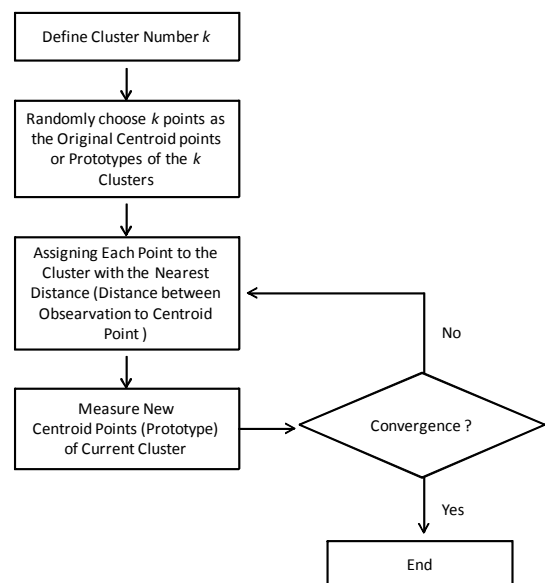


Fig. 1. Flowchart of the conventional k-means algorithm.

clusters. Each centroid point is viewed as the prototype of one cluster. The final clustering results correspond to the choice of the initial k centroid points. In general, it is better to select k centroid points that are far away from each other [29]. Further, every sample point is classified into the cluster with the nearest distance, i.e, comparing each point with k centroid points and assigning that point to the centroid point representing one cluster with the minimum distance. The distance can be measured using the Minkowski distance, Euclidean distance or Cityblock distance [18]. When all observations have been assigned to their corresponding clusters, one round of clustering is completed. The next step is to update the centroid point of each cluster. The centroid point is calculated using the cluster members at the current cluster as achieved from previous clustering rounds. Consequently, each point can be assigned to a new centroid point as per the shortest distance criterion. This iteration is continued until the centroid points converge.

In fact, the above-mentioned k-means algorithm attempts to minimize the within-cluster sum of squares. In other words, the k-means technique seeks to minimize the following term [27-29]:

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} d^2(x_j, \mu_i) \quad (1)$$

where k is the cluster number, S_i is the set including the corresponding cluster member of the i th cluster, μ_i is the mean value of observations in the i th cluster and is used to denote the prototype of the i th cluster, and $d^2(x_j, \mu_i)$ is the distance between the sample x_j and centroid point μ_i . Because k-means algorithms are highly dependent on the initial cluster centroids, Eq. (1) will always converge to a local minimum value. In order to achieve a better minimum result, the k-means algorithm can be conducted repeatedly and the minimum within-cluster sum of squares can be chosen. Moreover, the third and fourth step in Fig. 1 can be expressed as Eq. (2) and Eq. (3) respectively

[27-28]:

$$S_i^{(t)} = x_p : D(x_p, \mu_i^{(t)}) \leq d(x_p, \mu_i^{(t)}) \forall 1 \leq j \leq k, \quad (2)$$

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j, \quad (3)$$

where (t) denotes the t th iteration and $S_i^{(t)}$ represents the number of members in the i th set. Here, each point x_p is only allotted to exactly one cluster S_j even if the point x_p achieves the same distance as that of more than one centroid points in the k clusters.

The computational time complexity of a general k-means algorithm is linearly proportional to the number of samples, the number of clusters, the number of observation dimensions, and the number of iterations resulting in convergence. One of the drawbacks of the traditional k-means algorithm is that clustering results can be affected by noise or outlier observations. However, this problem can be mitigated or avoided by employing outlier analysis methods such as Random Sample Consensus (RANSAC) [30-31] on the sample data before implementing the k-means algorithm. Thus far, many algorithms based on the k-means approach have been proposed and widely used in the area of image processing, computer vision and data mining [20-23].

3. PROCEDURES OF THE PROPOSED CLUSTERING METHOD

Based on description of the k-means algorithm in section 2, note that the number of members in each cluster is not fixed but random after implementing k-means. However, in some cases, a fixed number of elements in each category are expected. For example, when k-means is used to group N observations into two clusters and each cluster contains $N/2$ members, the conventional k-means fails because it cannot adjust the cluster elements as per the user's requirement. In this section, a k-means based clustering method with a

fixed number of cluster elements is proposed. This method will be useful in fields such as logistics management when the number of group elements needs to be adjusted to satisfy a predefined number. The procedure for the proposed clustering method is described as follows:

step 1: Classify N observations into k clusters using the conventional k-means algorithm. Denote the resulting k centroid points as $\mu_1, \mu_2, \dots, \mu_k$.

step 2: Calculate the center point of set $\mu_1, \mu_2, \dots, \mu_k$. The computed center point is expressed as C_μ . Here, $\mu_1, \mu_2, \dots, \mu_k$ are prototypes of k clusters. C_μ and $\mu_1, \mu_2, \dots, \mu_k$ can be used to obtain the corresponding distance between C_μ and μ_i ($0 < i \leq k$) as $d_i = d(C_\mu, \mu_i)$. This step is used to decide which cluster member should be adjusted initially. In [27], the cluster with the minimum cluster member size is selected to adjust its elements. However, this method results in errors when the cluster with the minimum number of elements is not located at the margin. Here, the margin means the cluster center point that is on the convex hull containing all of the cluster center points. This phenomenon is illustrated in the experiment.

step 3: Choose cluster j from the k clusters that have the maximum distance to C_μ , i.e., $d_j = \max(d_i = d(C_\mu, \mu_i))$ where $0 < i \leq k$. Then, adjust the element in the j th cluster. Assume that the required number of elements for the k clusters is a_1, a_2, \dots, a_k where $a_1 + a_2 + \dots + a_k = N$. The number of elements in the k clusters obtained from k-means algorithm in step 1 is assumed to be n_1, n_2, \dots, n_k . Once the cluster used to adjust its element is determined, the number of elements can be denoted by n_j (here, the j th cluster is selected). However, the desired number of elements for this cluster is unknown, and it should be chosen from the set of (a_1, a_2, \dots, a_k) . The greedy strategy is applied to select the desired number of elements of the j th cluster. The value that is most similar to n_j in (a_1, a_2, \dots, a_k) is selected as the desired number of members for the j th cluster because this value can make

adjustments with the minimum element. Assume that the obtained desired number is a_j for the j th cluster.

step 4: The current number of elements in the j th cluster is n_j and the desired number of elements is a_j . In this step, we have to determine whether this cluster needs to recruit new members or it has to discard redundant elements. This decision can be taken easily by comparing n_j with a_j . If $n_j - a_j > 0$, the j th cluster needs to discard $n_j - a_j$ elements. If $n_j - a_j < 0$, $a_j - n_j$ new members should be recruited into the j th cluster. Of course, when $n_j - a_j = 0$, it is unnecessary to adjust the cluster element. Here, the recruitment and discard process is designed as follows. Both the recruitment and discard process represent the greedy strategy because they attempt to minimize the within-cluster sum of squares while adjusting the cluster element at the same time.

Recruitment process: Recruit $a_j - n_j$ members from the $N - n_j$ observation samples (excluding the n_j points that are already in the j th cluster) with the shortest distance criterion. In this paper, the Euclidean distance is employed although other distance measurement methods such as the Minkowski and Cityblock distance can also be adopted. Thus, $a_j - n_j$ points that are nearest to μ_j are recruited into the j th cluster from other clusters to satisfy the desired number of cluster members. Here, μ_j is the centroid point of the j th cluster.

Discard process: The $n_j - a_j$ elements in the j th cluster will be discarded with the same shortest distance criterion. However, the $n_j - a_j$ elements that are nearest to the centroid points of other clusters are discarded. It is confusing because one may assume that it is better to discard the $n_j - a_j$ elements that are farthest to μ_j which is the centroid point of the current cluster. Nevertheless, this will result in incorrect adjustments and considerably increase the within-cluster sum of squares. This is demonstrated in Fig. 2. When one point needs to be discarded from cluster 2, it is noted that discarding

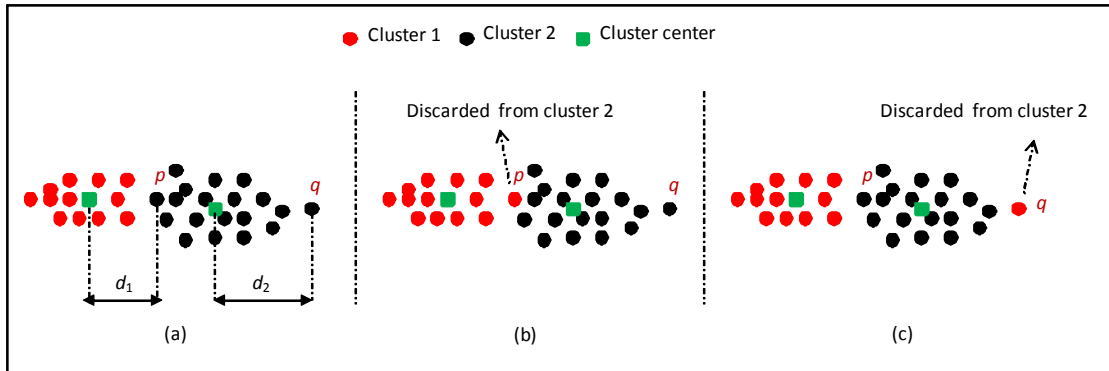


Fig. 2. Illustration of the discard Process. (a) Clustering result with k-means. (b) Adjustment results by discarding point p . (c) Adjustment result by discarding point q .

one point (point p in Fig. 2(a)) that is nearest to the other cluster's (cluster 1 in Fig. 2) centroid points is better than discarding the point (point q) that is farthest to the current cluster's (cluster 2 in Fig. 2) centroid point. Otherwise, the result will affect the other cluster as showed in Fig. 2(c). It is obvious that the discarded point q in Fig. 2(c) is recruited by cluster 1, which results in poor element distribution in cluster 1 because the elements are considerably dispersive and pass through the elements in cluster 1. On the other hand, discarding point p from cluster 2 as shown in Fig. 2 (b) makes the adjustment reasonable.

step 5: Set $k=k-1$, $N=N-a_j$, remove a_j from set $\{a_1, a_2, \dots, a_k\}$, and determine the updated set denoted as $\{a_1, a_2, \dots, a_{k-1}\}$. If $k \neq 1$, go to step 1 and continue the adjustment process. Otherwise, terminate the iteration and update all centroid points for new cluster members with a criterion of minimizing the following term [27-28]:

$$\min_{x_j \in S_i, 0 < i \leq k} \|x_j - \mu_i\|^2 \quad (4)$$

where x_j is the element in a new set S_i of the i th cluster and μ_i is the updated centroid point of the i th cluster.

The procedure for the proposed clustering algorithm indicates that this method can be implemented recursively thus making the implementation convenient and efficient. The pseudo

code for the k-means based clustering method with fixed cluster members is presented in Fig. 3. In Fig. 3, $N=N-a_j$ means the remaining unadjusted data. That is, the recursion does k-means only on the remaining data after taking out the data points in clusters already adjusted.

```

N: total number of observations
k: number of cluster
(a1,a2,...,ak): set of the required number of cluster member
nj: the number of element of jth cluster achieved from k-means algorithm
aj: the desired number of element of jth cluster

Start
1 k-means(N, k) % apply k-means algorithm to the N observations
2 if k=1
3   update the centroid point of each new cluster
4   return.
5 else
6   calculate the center point of the k centroid points
7   adjust the element of jth cluster that is the farthest to the calculated center point
8   recruit aj-nj element from other clusters or discard nj-aj element from current cluster
9   N=N-aj
10  k=k-1;
11  goto line 1
12 end
end
    
```

Fig. 3. Pseudo code for the proposed k-means based clustering method.

4. EXPERIMENTAL RESULTS

In this section, a computer simulation is conducted to demonstrate the feasibility of the k-means based clustering method with a fixed number of cluster members. Concurrently, we compare our results with that present in [27] leading to arbitrary results in some cases. The main difference

between the proposed k-means based method and the one in [27] is that the manner in which the first cluster is selected where the cluster member should be initially adjusted in each iteration. For clustering with only two clusters, the two approaches will achieve the same clustering results. However, when the number of clusters is more

than two, the method in [27] produces problems in some cases. Consequently, the simulation in this paper is executed with two types of clusters: One with two clusters and the other with three clusters. Fig. 4 shows the clustering results with two clusters. The number of cluster members is 20 and 40 which are obtained using the conventional

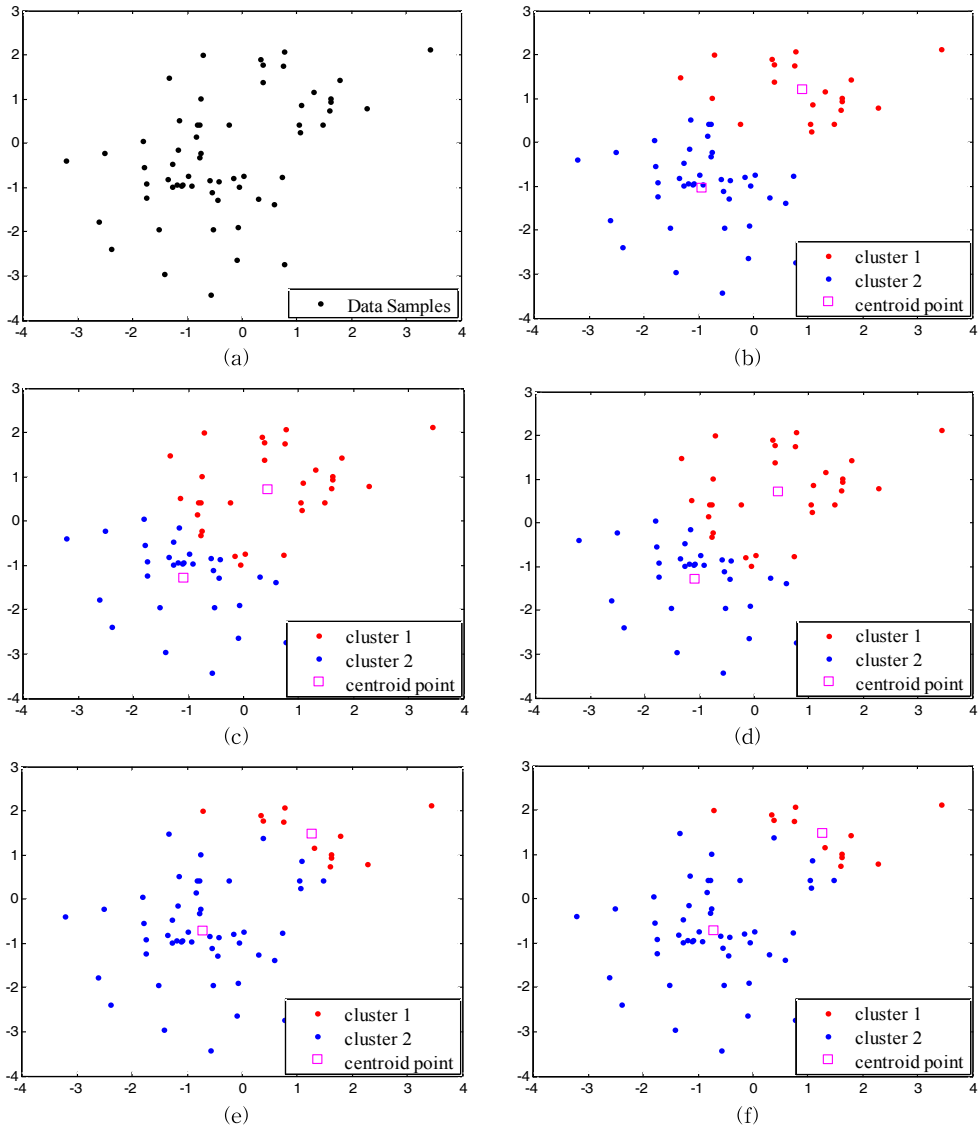


Fig. 4. Clustering results with two clusters. (a) data samples, (b) clustering result using the conventional k-means algorithm, (c) and (d) clustering results using the method in [27] and the proposed approach when the number of cluster elements is assigned to be 50% and 50% of the total samples, respectively, (e) and (f) clustering results using the method in [27] and the proposed approach when the number of cluster elements is assigned to be 20% and 80% of the total samples, respectively.

k-means algorithm whereas the total number of data samples is 60, which are randomly generated. Here, the sample data are generated from two groups that follow Gaussian distribution. Fig. 4(c) and 4(d) show the clustering results when the number of members for the two clusters is adjusted to occupy 50% and 50% of the total number of observations with the method in [27] and the proposed method respectively. Fig. 4(e) and 4(f) are the adjustment results when the number of members in the two clusters are adjusted to possess 20% and 80% with the method in [27] and the proposed method, respectively. It is noted that for the adjustment of cluster members with two clusters, the method in [27] and the proposed approach achieve the same acceptable performance. In order to numerically view the results, the sum of within-cluster sum of square [see Eq. (1)] for the two methods [method in [27] and method in this paper] with two clusters are given in Table 1. It is noted from Table 1 that the sums of within-cluster sum of square for the two methods are the same in condition of two clusters.

Fig. 5 presents the clustering results with three clusters. 200 samples are randomly generated and classified with conventional k-means, the clustering approach in [27], and the proposed clustering algorithm. Also, these sample data are generated from three groups that follow Gaussian distribution. The number of elements in the three clusters obtained by the conventional k-means algorithm is 86, 48, and 66 as shown in Fig. 5(b). Fig. 5(c) and 5(d) are the clustering results with method in [27] and the proposed method, respectively, whereas the number of cluster members is adjusted to be 33%,

33%, and 34% of the total data samples. Fig. 5(e) and 5(f) are the clustering results with method in [27] and the proposed method, respectively when the number of cluster elements is adjusted to be 20%, 40%, and 40% of the total data samples. It is apparently shown that the cluster members for cluster 3 in Figs. 5(c) and 5(e) are considerably poor when they are adjusted by starting from cluster 1. On the other hand, the proposed method in this paper can make up the deficiency in [27]. The experimental results discover that this proposed algorithm can achieve better adjustment performance than that in [27] when multiple clusters are considered. In other words, this means that the choice of the first cluster for updating the corresponding element is very important and affects the final clustering results. Here, we have to note that the method in [27] can also adjust multiple cluster members well when the data samples are not long and narrow as shown in the experimental part in [27]. Also, the sum of within-cluster sum of square [see Eq. (1)] for the two methods [method in [27] and method in this paper] with multiple clusters are provided in Table 2. It is proved that the new method proposed in this paper can achieve much smaller sum of within-cluster sum of square when it is compared with method in [27] under the condition of multiple clusters.

In this simulation, each point can be viewed as a city while the centroid point of each cluster can be considered as the logistics center. Sometimes, logistics company need each logistics center to handle business with a fixed number of cities. Thus, the proposed method in this paper can be applied to choose the cities (cluster members) for

Table 1. Sum of within-cluster sum of square for method in [27] and this new method

	Pre-defined percentage for each cluster elements	Method in [27]	New method in this paper
Sum of within-cluster sum of square	50%, 50% [see Fig. 4(c) and (d)]	109.81	109.81
	20%, 80% [see Fig. 4(e) and (f)]	97.97	97.97

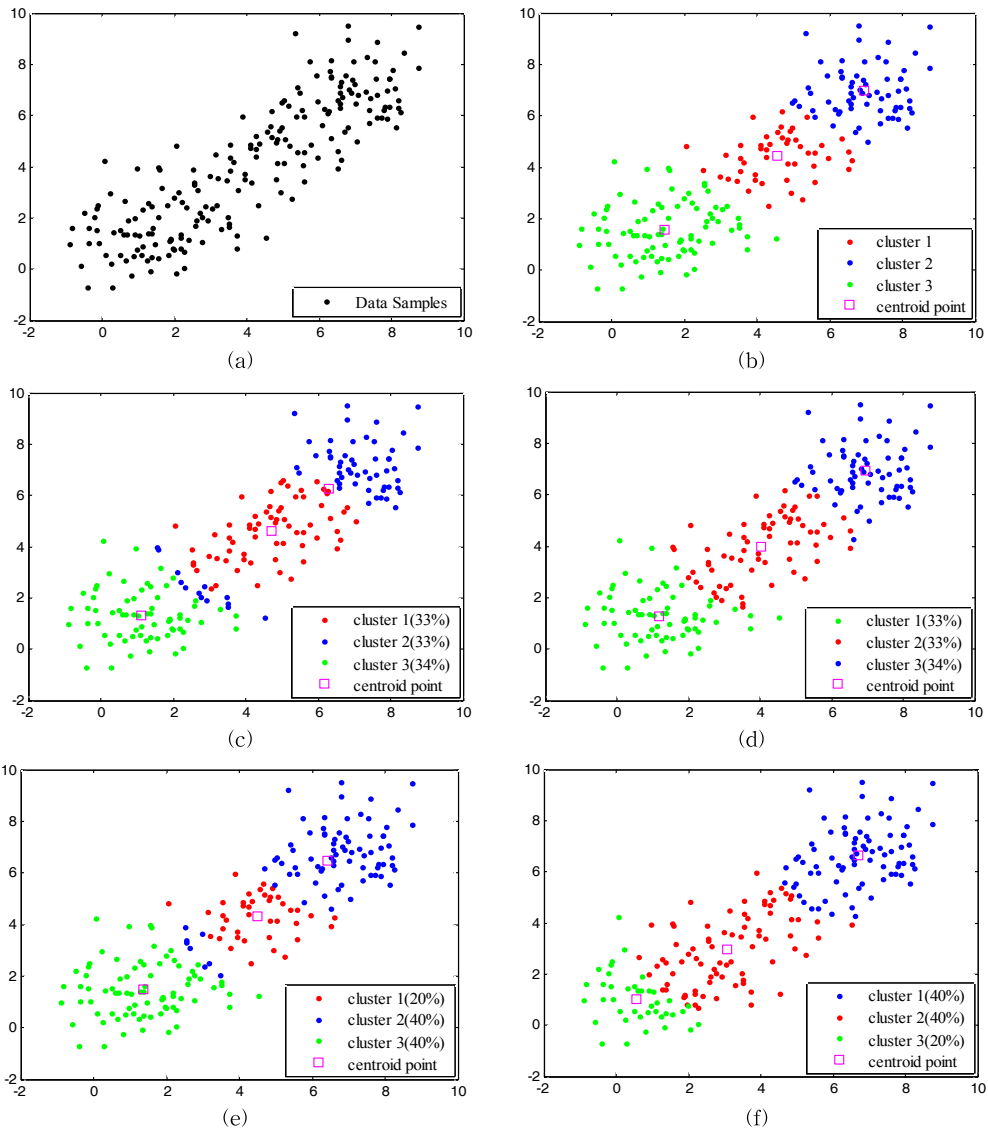


Fig. 5. Clustering results with three clusters. (a) data samples, (b) clustering result by conventional k-means algorithm, (c) and (d) clustering results with method in [27] and the proposed approach in this paper when the number of cluster elements are assigned to be 33%, 33%, and 34% of the total samples respectively. (e) and (f) clustering results with the method in [27] and the proposed approach when the number of cluster elements is assigned to be 20%, 40%, and 40% of the total samples, respectively.

Table 2. Within-cluster sum of square for method in [27] and this new method

	Pre-defined percentage for each cluster elements	Method in [27]	New method in this paper
Sum of within-cluster sum of square	33%, 33%, 34% [see Fig. 5(c) and (d)]	628.36	460.66
	20%, 40%, 40% [see Fig. 5(e) and (f)]	626.76	466.53

each cluster. In other words, the number of cities will be random when a simple clustering algorithm such as k-means is applied to all of the cities. Thus, each cluster members need to be adjusted in order to satisfy with the predefined number of cities which are managed by each logistics center. In addition, the proposed method can be helpful to image segmentation, object classification and pattern recognition when some priori knowledges such as the occupation rate of each class of object are known.

5. CONCLUSIONS

In this paper, a k-means based clustering method with a fixed number of cluster members was proposed and demonstrated. Experimental results show that the proposed algorithm worked well when the number of cluster elements had to satisfy some predefined values. Further, simulation results revealed that the proposed clustering method is superior to the conventional k-means and previously proposed clustering algorithm for a fixed number of cluster members. This algorithm is suitable for data sets with two or more clusters (more than two clusters). The proposed method is achieved based on the greedy strategy and can be implemented recursively to make the algorithm convenient and efficient. Moreover, the proposed method can be applied to other clustering methods and is not limited to the k-means algorithm. This method would be useful in image segmentation, object classification, pattern recognition and logistics management. Moreover, some outlier analysis approaches can be applied to the data samples before performing the proposed clustering algorithm to make the algorithm robust against noise.

REFERENCE

- [1] J. Maintz and M. Viergever, "A Survey of Medical Image Registration," *Medical image analysis*, Vol. 2, Issue 1, pp. 1-36, 1998.
- [2] J. Kim, "A Comparative Study on Classification Methods of Sleep Stages by Using EEG," *Journal of Korea Multimedia Society*, Vol. 17, Issue 2, pp. 113-123, 2014.
- [3] V. Gudivada and V. Raghavan, "Content based Image Retrieval Systems," *Computer*, Vol. 28, Issue 9, pp. 18-22, 1995.
- [4] M. Liberatore and D. Breem, "Adoption and Implementation of Digital-imaging Technology in the Banking and Insurance Industries," *IEEE Transactions on Engineering Management*, Vol. 44, Issue 4, pp. 367-377, 1997.
- [5] B. Wang and S. He, "Robust Optimization Model and Algorithm for Logistics Center Location and Allocation under Uncertain Environment," *Journal of Transportation Systems Engineering and Information Technology*, Vol. 9, Issue 2, pp. 69-74, 2009.
- [6] R. Chen and Y. Li, "Logistics Center Locating Based on TOPSIS method," *Journal of Anhui University of Technology*, Vol. 23, Issue 2, pp. 221-224, 2006.
- [7] R. Näätänen, "The Role of Attention in Auditory Information Processing as Revealed by Event-related Potentials and other Brain Measures of Cognitive Function," *Behavioral and Brain Sciences*, Vol. 13, Issue 2, pp. 201-233, 1990.
- [8] D. LaBerge and S. Samuels, "Toward a Theory of Automatic Information Processing in Reading," *Cognitive Psychology*, Vol. 6, Issue 2, pp. 293-323, 1974.
- [9] C. Wu, B. Shen, and Q. Chen, "Vehicle Classification in Pan-tilt-zoom Videos Via Sparse Learning," *Journal of Electronic Imaging*, Vol. 22, Issue 4, pp. 041102-041102, 2013.
- [10] W.R. Schwartz, F.R. de Siqueira, and H. Pedrini, "Evaluation of Feature Descriptors for Texture Classification," *Journal of Electronic Imaging*, Vol. 21, Issue 2, pp. 023016-

- 17, 2012.
- [11] A. Nadas, R. Mercer, L. Bahl, R. Bakis, P. Cohen, A. Cole, F. Jelinek, and B. Lewis, "Continuous Speech Recognition with Automatically Selected Acoustic Prototypes Obtained by either Bootstrapping or Clustering," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 113-1155, 1981.
- [12] C. Gu, S. Zhang, and Y. Sun, "Realtime Encrypted Traffic Identification using Machine Learning," *Journal of Software*, Vol. 6, Issue 6, pp. 1009-1016, 2011.
- [13] R.A. Patil, V. Sahula, and A.S. Mandal, "Features Classification using Support Vector Machine for a Facial Expression Recognition System," *Journal of Electronic Imaging*, Vol. 21, Issue 4, pp. 043003-1, 2012.
- [14] R. Rangayyan and F. Oloumi, "Fractal Analysis and Classification of Breast Masses using the Power Spectra of Signatures of Contours," *Journal of Electronic Imaging*, Vol. 21, Issue 2, pp. 023018-1, 2012.
- [15] K. Bemis., L. Eberlin, C. Ferreira, R. Cooks, and O. Vitek, "Spatial Segmentation and Feature Selection for Desi Imaging Mass Spectrometry Data with Spatially-aware Sparse Clustering," *BMC Bioinformatics*, Vol. 13, pp. A8, 2012.
- [16] L. Guerra, L. McGarry, V. Robles, C. Bielza1, P. Larranaga, and R. Yusteet, "Comparison between Supervised and Unsupervised Classifications of Neuronal Cell Types: A Case Study," *Developmental Neurobiology*, Vol. 71, Issue 1, pp. 71-82, 2011.
- [17] R. Gonzalez and R.Woods, *Digital Imaging Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 2002.
- [18] E. Gose and R. Johnsonbaugh, *Pattern Recognition and Image Analysis*, Prentice Hall, Upper Saddle River, NJ, USA, 1996.
- [19] J. Kleinberg and E. Tardos, *Algorithm Design*, Pearson Education, England, 2005.
- [20] T. Paul and S. Bandhyopadhyay, "Segmentation of Brain Tumor from Brain MRI Images Reintroducing K-Means with Advanced Dual Localization Method," *International Journal of Engineering Research and Applications*, Vol. 2, Issue 3, pp. 226-231, 2012.
- [21] B. Ramamurthy and K. Chandran, "CBMIR: Shape-based Image Retrieval using Canny Edge Detection and K-means Clustering Algorithms for Medical Images," *International Journal of Engineering Science and Technology*, Vol. 3, Issue 3, pp. 209-212, 2011.
- [22] E. Ngai, L. Xiu, and D. Chau, "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification," *Expert Systems with Applications*, Vol. 36, Issue 2, pp. 2592-2602, 2009.
- [23] Y. He and Q. Zhen, "Logistics Customer Segmentation Modeling on Attribute Reduction and K-Means Clustering," *Journal of Communication and Computer*, Vol. 10, pp. 1114-1119, 2013.
- [24] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun Database: Large-scale Scene Recognition from Abbey to Zoo," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3485-3492, 2010.
- [25] H. Steinhaus, "Sur La Division Des Corps Matériels En Parties," *Bulletin of the Polish Academy of Science*, Vol. 4, Issue 12, pp. 801-804, 1957.
- [26] S. Lloyd, "Least Square Quantization in PCM," *IEEE Transactions on Information Theory*, Vol. 28, Issue 2, pp. 129-137, 1982.
- [27] F. Yi, and I. Moon, "Extended K-Means Algorithm," *IEEE International Conference on Intelligent Human-Machine Systems and Cybernetics*, pp. 263-266, 2013.
- [28] L. Costantini, L. Capodiferro, M. Carli, and A. Neri, "Texture Segmentation based on

Laguerre Gauss functions and K-means Algorithm Driven by Kullback-Leibler Divergence," *Journal of Electronic Imaging*, Vol. 22, Issue 4, pp. 043015-043015, 2013.

- [29] R. Liu, D. Dey, D. Boss, P. Marquet, and B. Javidi, "Recognition and Classification Red Blood Cells using Digital Holographic Microscopy and Data Clustering with Discriminant Analysis," *Journal of the Optical Society of America A*, Vol. 28, Issue 6, pp. 1204-1210, 2011.
- [30] S. Fang, Q. Shi, and Y. Cao, "Adaptive Removal of Real Noise from a Single Image," *Journal of Electronic Imaging*, Vol. 22, Issue 3, pp. 033014-033014, 2013.
- [31] C. Hui and R. Wang, "The Application of Random Sample Consensus in Photogrammetric Relative Orientation," *IEEE International Conference on Information Science and Engineering*, pp. 4113-4116, 2010.



Faliu Yi

received the B.E. degree from Yunnan University, Kunming, China, in 2008 and the M.E. degree in computer engineering from Chosun University, Gwangju, Korea, in 2012, and is currently working toward the Ph.D.

degree in computer engineering from the same university. His research interests include 3D image processing, computer vision, image segmentation, object tracking, pattern recognition and parallel computing.



Inkyu Moon

received the B.S. degree in electronics engineering from Sung KyunKwan University, Korea, in 1996, and the Ph.D. degree in electrical and computer engineering from University of Connecticut in 2007. He joined

Chosun University, Korea, in 2009, and is currently an associate professor at the School of Computer Engineering. His research interests include digital holography, bio-medical imaging, and optical information processing. Dr. Moon is member of IEEE, OSA and SPIE.