International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1669-1678

http://www.ripublication.com

# The New clustering approach using Extreme Learning Machine with k-Medoids Techniques

**Gopinathan.S**
*Department of Computer Science,*
*University of Madras, Chepauk, Chennai, India.*

**Pandiyan.M**
*Department of Computer Science, University of Madras*
*Chepauk, Chennai, India.*

**Thangavel.P**
*Department of Computer Science, University of Madras*
*Chepauk, Chennai, India.*

**Abstract**

In this paper, we proposed a new clustering approach which combines the extreme learning machine and k-Medoids called ELM k-Medoids technique. Extreme learning machine (ELM) is an emerging, efficient and effective learning algorithm for regression and classification applications. The benefits of the ELM and their enhanced performance to solve classification and regression problems infer that clustering in elm high dimensional space would also produce excellent results. Elm was used to project the data object in to the higher dimensional feature space and k-Medoids algorithm used to perform clustering in this feature space. The proposed method was tested on four different data sets and compared with three algorithms like k-Means, k-Medoids and Elm k-Means. The obtained results are shows that the proposed approach gives the best clustering accuracy compared with the other three clustering techniques.

**Keywords:** Extreme Learning Machine (ELM), k-Medoids, k-Means, ELM k-Means.

## I. INTRODUCTION

Clustering techniques are widely used in many application domains such as machine learning, data compression, image analysis, computer vision and so on. The goal of clustering is to assign a set of data object into groups on the basis of a similarity between the objects. In the past decades many clustering techniques were introduced by the researchers. Most widely used techniques are the k-Means algorithm [1]. k - Means fit effectively well in case of ellipsoidal or spherical data distribution that is linearly separable. The inabilities to cluster data that are subjective to arbitrary shapes make k-Means algorithm fail because clustering is nonlinear. To solve such problems, a nonlinear transformation is applied to the original data thus project to high dimensional feature space and then performing clustering. This lead to the introductions of new kernel based clustering. Scholkopf et al. [2] proposed a kernel principal component analysis that requires eigenvalues but problems include local minima and scalability. Local minima was further solved by Tzortzis and Likas [3] which does not depend on cluster initialization and is based on a global kernel k-Means algorithm. The problem of scaling a large kernel matrix was studied by Chitta et al. [4]. To provide high efficiency Zhang et al. [5] integrates the goodness of kernel principal component analysis along with ant based clustering. Recently, ELM based clustering techniques were studied by the researchers because it approximates any continuous target function and also has the capability to classify any disjoint region. The benefits of the ELM and their enhanced performance to solve the classification and regression problems inferred that clustering in elm high dimensional space would also produce excellent results. ELM has the Universal approximation and classification capability, this proficiency results in feature space that is more convenient for clustering after mapping into high dimensional feature space. Qing He et al. [6] proposed a method to solve a clustering problem in elm feature space using k-Means algorithm which is convenient and simpler then the kernel based feature mapping technique because it requires less human intervention and has no necessity to adjust any parameters. Another approach for clustering in extreme learning machine was studied by Alshamiri et al. [7], which integrate Extreme learning machine along with k-Means algorithm. However after transforming the data object in to the elm feature space researchers use the traditional k-Means clustering algorithm in this new space but the k-Means clustering is based in the sum of squared Euclidean distance which is completely dependent on the centroid. In this paper, we combined both extreme learning machine and k-Medoids algorithm [8] for clustering. Partitioning around Medoids or the k-Medoids algorithm is a partitioned clustering algorithm which is slightly modified from the k-Means algorithm. Both the algorithms attempt to minimize the squared-error but the k-Medoids algorithm is more robust to noise and outliers. The ELM method is used to project data objects into high dimensional feature space and the k-Medoids algorithm applied for clustering in the features space.

The rest of this paper is organized as follows: Section 2 describes the basics of the Extreme Learning Machine (ELM). Section 3 introduces k-Medoids algorithm. In Section 4, discuss about ELM k-Medoids algorithm. In Section 5 shows the performance of ELM k-Medoids. Finally, Section 6 concludes this paper.

## II. EXTREME LEARNING MACHINE

Extreme Learning Machine aimed to speed the process of learning in a feed forward network. Traditional algorithm requires training the parameters such as weights and bias of all the layers which was a heuristic task and conventional slow in gradient based learning algorithms such as back propagation algorithms (BP). Thus an ELM for single hidden layer feed forward neural networks (SLFNs) was intended to randomly choose weights of the input and hidden layer, bias and also analytically determine the output weights. An extremely fast learning speed with minor training errors proves the performance of ELM over the conventional methods. It also obtains the smallest norm of weights and is widely used for solving problems based on classification and regression. For formal definition of ELM, we will follow the same notational convention as used in [9].

For instance, we are given a set of training examples.

$Z = \{(x_j, t_j) \mid x_j \in \Re^d, t_j \in \Re^m, j = 1,2,...,N\}$, Single Layer Feed Forward Networks(SLFNs) with V hidden neurons and activation function G(x), can be represented as [9]:

$$\sum_{i=1}^{K} \lambda_i G(w_i . x_j + b_j) = Y_j,\ j = 1,2,\ldots,N. \qquad (1)$$

Where, $\lambda_i = [\lambda_{i1},...,\lambda_{im}]^T$ is the weight vector connecting the i$^{th}$ hidden node and the output nodes, $W_i = [w_{i1},\ldots,w_{id}]^T$ is the weight vector connecting the i$^{th}$ hidden node and the input nodes, and $b_i$ is the bias of the i$^{th}$ hidden node. $w_i . x_j$ is the inner product of $w_i$ and $x_j$. Parameters $\lambda_i$, i =1,...,V can be estimated such that,

$$\sum_{i=1}^{L} \lambda_i G(w_i . x_j + b_j) = t_j,\ j = 1,2,\ldots,N. \qquad (2)$$

Equation (2) can be written as in [9]:

$$H\lambda = T \qquad (3)$$

Where,

$$H=\begin{bmatrix} G(w_1.x_1+b_1)...G(w_V.x_1+b_V) \\ \vdots \qquad \cdots\vdots \\ G(w_1.x_N+b_1)...G(w_V.x_N+b_V) \end{bmatrix}_{N\times V} \quad (4)$$

$$\lambda=\begin{bmatrix} \lambda_1^T \\ \vdots \\ \lambda_V^T \end{bmatrix}_{V\times m} \text{ and } T=\begin{bmatrix} t_1^T \\ \vdots \\ t_V^T \end{bmatrix}_{V\times m} \quad (5)$$

H is called the hidden layer output matrix of the neural network [10].The ELM algorithm consists of three steps and can be summarized as follows [11]:

**Algorithm 1:**

**Input:** A training dataset with N tuples $Z=\{(x_j,t_j)\,|\,x_j\in\Re^d, t_j\in\Re^m, j=1,2,...,N\}$,

**Output:** The output weights $\lambda$

Steps:

1. Randomly generate input weight $W_i$ and bias $b_i$, i=1,…, V.
2. Compute the hidden layer output matrix H
3. Calculate the output weigh $\lambda:\lambda=H^{\dagger}T$ ,where $H^{\dagger}$ is the Moore-Penrose generalized inverse [12] ofthe hidden layer output matrix H and T = $\left[t_{i,\cdots,}t_V\right]^T$.

Theoretically, the number of neurons in the hidden layer of the ELM should be large enough to achieve good generalization performance. A detailed discussion on hidden nodes selection in particular and ELM in general can be found in [9-11].

## III.K – MEDOIDS ALGORITHM

k-Medoids is an unsupervised learning algorithm. In k-Medoids clustering, it partitions a group of data objects into k number clusters. The functioning of k-Medoids [8] clustering algorithm is like k-Means clustering [1]. It uses Medoids to represent the cluster rather than using centroids in k-Means clustering. The algorithm randomly selects k representative data item as initial Medoids to form k number of clusters, and all other remaining data objects are placed closest to Medoids in a cluster. Then the new Medoids are determined and it shows the enhanced cluster. In this process all the remaining data objects are over again assigned to the cluster having neighboring Medoids. The position of Medoids change consequently with each iteration and it reduces the sum of the dissimilarity between the data objects along with the Medoids. This process is repeated until no Medoids change, as it marks the end of the process that results in the final clusters along with their representative Medoids.

**Algorithm 2: k-Medoids Clustering:**

**Input:**

k: number of clusters

P: the data set containing n items

**Output:**

A set of k clusters

Steps:

1. Initially select k random points as the Medoids from the given P data points of the data set.
2. Associate each data point to the closest Medoids by using Manhattan distance metrics.
3. For each pair of non-selected object *m* and selected object *p*, calculate the total swapping cost (TSC) $TSC_{mp.}$

   If TSC < 0, *p* is replaced by *m*
4. Repeat the steps 2 to 3 until there is no change of the Medoids.

The main objective of k-Medoids algorithm is that minimizes the sum of the dissimilarities of all the objects to their nearest Medoids

$$\chi = \sum_{l=1}^{n} \sum_{p \in C_l} |p - m_l| \qquad (6)$$

$\chi$ : Sum of absolute error for all items in the dataset.

P: Data point in the space representing a data item

$m_l$ : Medoids of cluster $C_l$.

## IV. PROPOSED METHOD

This section describes the proposed method which combines both the ELM and k-Medoids algorithms. ELM projects the input data objects into a high-dimensional feature space. This nonlinear data transformation into some high dimensional feature space increases the possibility of linear separability of the patterns. The assimilation of ELM facilitates the k-Means algorithm to explore the inherent data structure in the new space. In ELM, the hidden layer maps the data from the input space $R^J$ to the high-dimensional feature space $R^Q (Q >> J)$ where the data clustering is performed. After transforming the data into the ELM feature space, k-Medoids clustering algorithm can be applied directly in ELM feature space as we called ELM k-Medoids algorithm.

The ELM k-Medoids algorithm consists of seven steps and can be summarized follows:

**Algorithm 3**. ELM k-Medoids algorithm.

**Input:**

k: Number of clusters,

V: Number of the hidden-layer nodes,

Z: Data set containing D objects.

**Output:**

A set of k clusters.

**Method:**

1: Mapping the original data objects in Z into the ELM feature space H using $G(X)=[g_1(x),\ldots,g_i(x),\ldots,g_V(x)]^{T;}$

2: Arbitrarily choose k data objects from H as the initial Medoids.

3: Assign each remaining data objects to a cluster with the adjoining Medoids.

4. Randomly select a non-Medoid data objects and calculate the total cost of Swapping old Medoid data item with the currently selected non-Medoid data item.

5. If the total cost of swapping is less than zero, then perform the swap operation to generate the new set of k-Medoids.

6. Repeat steps 2 to 5 till the Medoids stabilize their locations or reached the maximal iteration number limit.

7. Return a set of k Clusters.

## V. EXPERIMENTAL RESULTS

In this paper, four benchmark data sets are used to analyze the performance of the ELM k-Medoids algorithm. These data sets can be downloaded from the UCI Machine Learning Repository. Table 1 include the following information; data sets, number of patterns, number of attribute features and number of classes.

TABLE I: DATA SET SPECIFICATION

| Data set | Patterns | Attributes | Classes |
|---|---|---|---|
| Iris | 150 | 4 | 3 |
| Wine | 178 | 13 | 3 |
| Glass | 214 | 9 | 6 |
| Dermatology | 358 | 34 | 6 |

### *A. Data Sets*

The data sets considered in this work can be described briefly as follows. Fisher Iris data is the most popular data sets to test the performance of novel methods in pattern recognition and machine learning. There are three classes in the data set (Setosa, Versicolor and Virginica), each having 50 patterns with four features (sepal length, sepal width, petal length and petal width). Petal length, Petal width and iris type contains a linear relationship, while sepal length and iris type have a nonlinear relationship. Wine data set consist of 178 patterns of 13 attributes. Different kinds of wine are identified with 3 classes. In these 178 patterns, 1-59, 60-130 and 131-178 belong to the first, second and third wine category respectively. The glass data type consists of 214 patterns of 9 attributes and 6 glass types. The glass types includes float processed building windows, non-float processed building windows, vehicle windows, containers, tableware and head lamps. Dermatology data set aims to determine the type of Eryhemato-Squamous Disease. It contains 366 patterns. After the removal of missing values, the data set consists of 358 patterns and 34 features belonging to 6 different classes.

### *B. Results and Discussion*

For each data set, we report the Accuracy of clustering as follows,

$$Accuracy = \frac{\sum_{i=1}^{N} \eta(o_i, t_i)}{N} \times 100$$

Where, N is the number of instances of a dataset, $o_i$ and $t_i$ are real category labels and predicted label respectively. $o_i$ and $t_i$ are compared. If the prediction is true $t_i$ is set to 1 otherwise set to 0. In this work the number of hidden neurons is selected based on the highest mean of accuracy using trial and error method. For experimental comparison of k-Means, k-Medoids, ELM k-Means and ELM k-Medoids are all iterated for an average of 20 times. The following tables (II-V) depict the performance of these algorithms with different cluster numbers for each data set and their Mean and Standard deviation accuracy values.

TABLE II WINE DATASET

| No of Cluster | k-Means | k-Medoids | Elm K-Means | Elm k-Mediods |
|---|---|---|---|---|
| 3 | 49.27±0.88 | 50.45±1.68 | 58.09±8.44 | **66.35±8.44** |
| 4 | 48.51±0.61 | 48.65±0.70 | 55.51±8.20 | 57.36±5.97 |
| 5 | 45.17±1.44 | 44.78±2.80 | 52.42±8.02 | 52.05±5.09 |
| 6 | 42.17±2.84 | 42.34±3.18 | 46.10±6.82 | 47.36±4.11 |

TABLE III DERMATOLOGY DATASET

| No of Cluster | k-Means | k-Medoids | Elm K-Means | Elm k-Mediods |
|---|---|---|---|---|
| 3 | 25.82±1.39 | 25.35±0.12 | 46.17±6.87 | 49.78±3.57 |
| 4 | 28.48±0.72 | 28.09±1.16 | 54.33±6.28 | 60.84±5.65 |
| 5 | 28.74±1.86 | 29.57±0.54 | 56.49±8.17 | 61.15±6.72 |
| 6 | 29.37±1.88 | 30.01±0.69 | 60.35±5.73 | **62.92±5.70** |
| 7 | 28.70±2.17 | 30.75±1.33 | 56.96±6.74 | 61.28±7.70 |

TABLE IV GLASS DATASET

| No of Cluster | k-Means | k-Medoids | Elm K-Means | Elm k-Mediods |
|---|---|---|---|---|
| 3 | 48.62±1.96 | 49.07±0.01 | 49.02± 1.94 | 49.72± 0.65 |
| 4 | 51.12±3.67 | 50.00±0.00 | 51.57±2.56 | 52.15±1.58 |
| 5 | 52.13±1.97 | 52.69±0.86 | 52.52±1.97 | 52.59±1.72 |
| 6 | 52.36±3.20 | 53.53±0.24 | 52.41±1.90 | **53.93±1.37** |
| 7 | 52.13±2.73 | 53.57±2.06 | 52.29±2.91 | 53.18±2.11 |

TABLE V IRIS DATASET

| No of Cluster | k-Means | k-Medoids | Elm K-Means | Elm k-Mediods |
|---|---|---|---|---|
| 3 | 80.13±15.77 | 89.53±0.65 | 94.93± 10.12 | **96.23± 0.997** |
| 4 | 71.07±1.54 | 72.67± 5.47 | 81.97±8.62 | 81.03±2.39 |
| 5 | 62.97±6.13 | 64.00± 6.27 | 70.47±7.10 | 70.07±2.95 |
| 6 | 59.33± 6.17 | 57.83±7.28 | 62.87±7.83 | 59.77± 4.699 |

## VI. CONCLUSION

A new clustering technique is thus formulated that combines both ELM and k-Medoids. ELM is used in this article to project data in to the higher dimensional feature space which in turn is clustered using the k-Medoids approach. The proposed technique suggested in this paper was applied to four different data sets with comparison among three algorithms like k-Means, k-Medoids and Elm k-Means. The experimental results clearly depict the efficiency of the proposed method over the other clustering approaches. In future, we aim to propose hybrid clustering algorithms that fuses well with extreme learning machine to improve the efficiency.

## REFERENCE

[1] Ng, M.K.: A note on constrained k-Means algorithms. Pattern Recognit. 33, 515– 519 (2000)

[2] Scholkopf, B., Smola, A., Muller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. 10(5), 1299–1319 (1998)

[3] Tzortzis, G.F., Likas, A.C.: The global kernel k-Means algorithm for clustering in feature space. IEEE Trans. Neural Networks 20(7), 1181–1194 (2009)

[4] Chitta, R.,Jin, R.,Havens, T.C.,Jain, A.K.: Approximate kernel k-Means: solution to large scale kernel clustering. In: Proceedings of 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, USA, PP.895-903 (2011)

[5] Zhang, L., Cao, Q.: A novel ant-based clustering algorithm using the kernel method. Inf. Sci. 181, 4658–4672 (2011)

[6] Qing He., Xin Jin.,Changyingu Du.,Fuzhen Zhuang.,Zhongzhi Shi.: Clustering in extreme learning machine feature space Neurocomputing 128, 88-95 (2014)

[7] A.K.Alsharmiri., B.R. Surampudi., Alok Sing.:A Novel ELM k-Means Algorithm for Clustering. Springer SEMCCO 2014, LNCS 8947, 212-222 (2015)

[8] Kaufman, L. and Rousseeuw, P.J., Clustering by means of Medoids, in

Statistical Data Analysis Based on the L1 - Norm and Related Methods, edited by Y. Dodge, North- Holland, 405–416 (1987)

[9] Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing 70, 489–501 (2006)

[10] Huang, G.B.: Learning capability and storage capacity of two- hidden-layer feedforward networks. IEEE Trans. Neural Networks 14(2), 274–281 (2003)

[11] Lan, Y., Soh, Y.C., Huang, G.B.: Constructive hidden nodes selection of extreme learning machine for regression. Neurocomputing 73, 3191–3199 (2010)

[12] Serre, D.: Matrices: Theory and Applications. Springer-Verlag Inc., New York (2002)