Effect of Kernel Learning in Unsupervised Learning for Clustering High Dimensional Databases

Esha Kashyap, S.R.Kannan, Mark Last

Abstract: ***This paper reviews the effectiveness of kernel learning in unsupervised data analysis using clustering. Cluster analysis is an explorative data analysis tool that assists in discovering hidden patterns or natural grouping and has many effective applications in various disciplines. The unison of kernel learning with the objective of unsupervised clustering algorithms facilitates in recognizing non linear structures in high dimensional data containing outliers with heavy noise. The recent kernel clustering methods considered in this paper are the kernelized versions of K-Means, Fuzzy C-Means, Possibilistic C-Means and Intuitionistic Fuzzy C-Means. Computational complexities in kernel based clustering algorithms are quiet prominent and our objective is to understand the performance gains while using kernels in clustering. Experimental studies of this paper substantiate that kernel based clustering algorithms yields significant improvements over their traditional counterparts.***

Index Terms: Unsupervised clustering, Data Analysis, Kernel learning, Partition clustering.

I. INTRODUCTION

Unsupervised clustering techniques of machine learning determine the available structures in data without any information about the objects in the data [4], [5]. Unsupervised learning has no predefined class labels for the data under speculation. Clustering is a renowned unsupervised learning technique, used for exploratory data analysis to unveil hidden patterns in data [25], [27]. Unsupervised learning aims at disclosing the natural groupings in the data [15], [20]. The hard clustering techniques proposed in literature are often not effective with many of real world applications where uncertainty is preferred over sharp boundary between clusters so fuzzy clustering is befitting for such kind of data sets [19]. Ruspini [24] first interjected the notion of fuzzy set theory in clustering. Leading to the incubation of the very first Fuzzy C-Means (FCM) proposed by Dunn [10] and later generalized by Bezdek [2], [3]. Hereafter many variants of fuzzy clustering approach were proposed [16], [17], intended to meliorate the existing algorithm and also address specific issues like noise handling [8], [21], adaption to different data [12], [18] and cluster genre [7], [13]. Kernel learning is employed to deal with the classification of data set that is not linearly separable. Linear separability is an important notion in machine learning research. Kernel based clustering has shown notably better performance rate than their traditional counter parts for almost all the data of real world. The main objective of this study is to observe the quantification in performance rate for kernel based clustering algorithm. The kernel clustering methods considered in this paper are the kernelized versions of K-Means, Fuzzy C-Means, Possibilistic C-Means and Intuitionistic Fuzzy C-Means. In recent years the clustering methods K-Means, Fuzzy C-Means, Possibilistic C-Means and Intuitionistic Fuzzy C-Means, are considered as effective tools in clustering the complicated real life databases [22], [23]. The fundamental difference between K-Means and Fuzzy C-Means is a degree of belonging of a data point to a cluster. The degree of belonging of a data point to a cluster is portrayed by the membership degree in Fuzzy C-Means and it allows the object to belong into all clusters with certain degree of fuzzy memberships. Possibilistic C-Means (PCM) came into consideration to address the challenges connected with the constraints on the membership degree utilized in fuzzy clustering. This paper proves the strength of kernelized version of the proposed unsupervised clustering learning of this paper through the experimental works on noisy ring data, noisy parabolic data, noisy Haberman's survival data, noisy iris plant database, wine data, contraceptive method choice, noisy Wisconsin prognostic breast cancer, SPECT heart data and glass identification database. The paper is arranged as follows. We start section 2 with a brief knowledge of kernels and kernel learning. Section 3 describes various clustering algorithms along with their kernelized versions. The experimental results are given in section 4 followed by the conclusion in the last section.

II. KERNEL METHOD

Kernel learning aids in transforming linearly inseparable data into linearly separable ones. The choice of kernel is decisive in efficiency of an algorithm and must be selected cautiously. In all scenario, the fundamental theoretical properties of any kernel is the same. In this section we define kernels, and to preserve simplicity, we are considering vectors in IR^d rather than C^d.

Definition: Consider a non empty set of n vectors X = $\{x_1, x_2, \ldots x_n\}$, where $x_1 \epsilon$

Revised Manuscript Received on April 24, 2019

Esha Kashyap, Department of Mathematics, Pondicherry University, Puduchery, India.
S.R.Kannan, Department of Mathematics, Pondicherry University, Puduchery, India.
Mark Last, Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Negev, Israel.

IR^d. A function K: X × X → IR is a kernel if and only if there is some implicit non-linear map φ : X → H into a separable Hilbert space H such that

$$K(x,x') = < \varphi(x), \varphi(x') >_H.$$

The kernel learning facilitates in implicit mapping of the data. This approach is known as the kernel trick, it projects a non linear data set into a higher dimension where it is more likely to be linearly separable and becomes accessible for classification task.

III. KERNALIZATION OF CLUSTERING ALGORITHM

Cluster analysis is a probative data analysis tool that congregate objects into distinct categories based on their resemblance. Clusters can be resort to elucidate the fundamental structure of data, which is beneficial in finer understanding of data. In recent years there have been considerable expansion of kernel based clustering algorithms, kernels demonstrate good generalization performance on a huge number of real world data sets. Researchers in various scientific and engineering areas have an ever growing interest in this area due to its effective learning and reasoning capabilities.

A. K-Means

K-Means is a facile unsupervised learning algorithm for clustering data structure. In this approach, a given data structure is fractionalized into specified number of clusters, which is determined prior. Given a data set X = $\{x_1, x_2, ... x_n\}$ and V = $\{v_1, v_2, ... v_k\}$ prototypes. It minimizes the following objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{\alpha_i} ||x_i - v_j||^2 \quad (1)$$

where $||x_i - v_j||^2$ is the squared Euclidean distance, α_i represents number of data points in the i^{th} cluster and k is the number of clusters.

B. Kernelization of K-Means

In [9] the authors embellished the K-Means clustering algorithm using kernel learning. Generalization of the kernel K-Means was done using a weight $w(a)$ for each point$a$. π_j represents clusters, and the partitioning of data points was designated as $\{\pi_j\}_{j=1}^{k}$. The objective of the weighted K-Means for a non-linear function Ø is as follows :

$$D(\{\pi_j\}_{j=1}^{k}) = \sum_{j=1}^{k} \sum_{a \in \pi_j} w(a) ||\emptyset(a) - m_j||^2 \quad (2)$$

where $m_j = \frac{\sum_{b \in \pi_j} w(a)\emptyset(b)}{\sum_{b \in \pi_j} w(b)}$. The functional in (2) monotonically decreases.

The Euclidean distance measure between data point $\emptyset(a)$ and the prototype value m_j is

$$\emptyset(a) - \frac{\sum_{b \in \pi_j} w(a)\emptyset(b)^2}{\sum_{b \in \pi_j} w(b)} = \emptyset(a)\,\emptyset(b) - 2\frac{\sum_{b \in \pi_j} w(a)w(a)\emptyset(a)\emptyset(b)}{\sum_{b \in \pi_j} w(b)} + \frac{\sum_{b,c \in \pi_j} w(a)w(c)\emptyset(b)\emptyset(c)}{\sum_{b \in \pi_j} w(b)} \quad (3)$$

The dot product $\emptyset(a).\emptyset(b)$ is worked out with the aid of kernel learning.

C. Fuzzy C-Means

An elementary variation of Fuzzy C-Means algorithm (FCM) was proposed by Dunn [10] in 1974, with auxiliary advancements being introduced by Bezdek [2], [3]. In Fuzzy C-Means each data point has stipulated degree of membership for belonging to every cluster. It adopts an iterative technique, in which the cluster centers are refurbished at every step. The agendum of FCM algorithm is partitioning a finite assemblage of n data set, X = $\{x_1, x_2, ... x_n\}$ into c cluster. For a given finite set of data, the algorithm generates an index of c cluster centers V = $\{v_1, v_2, ... v_c\}$, a partition matrix U = $\{u_{ij}\}$, $u_{ij} \in [0,1]$, i = 1,2,... n and j = 1,2,... c where u_{ij} represents the degree to which each element x_i belong to a cluster with center v_j. Fuzzy C-Means is a probabilistic clustering algorithm, this implies that the sum of all membership degrees for each data point over all cluster equals 1. That is

$$\sum_{j=1}^{c} u_{ij} = 1, \forall i = 1,2,..., n \quad (4)$$

Also we have

$$\sum_{i=1}^{n} u_{ij} > 0, \forall j = 1,2,..., c \quad (5)$$

Fuzzy C-Means minimizes the following objective function:

$$J(U,V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m ||x_i - v_j||^2 \quad (6)$$

where $||x_i - v_j||^2$ represent the square of the Euclidean distance between object x_i and cluster center v_j. $m > 1$ represents the degree of fuzzification which lies in (1,∞). Both U and V are upgraded iteratively along the way for optimization of the objective function. The evaluation of cluster centers and the degree of memberships are as follows:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \quad (7)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \frac{||x_i - v_j||^2}{||x_i - v_k||}^{\frac{2}{m-1}}} \quad (8)$$

D. Kernelization Of Fuzzy C-Means Clustering

Kernel based fuzzy clustering has two notable variation: in the first variation the prototype is restricted to the feature space while in the second variation an inverse mapping is carried out to map the prototypes from the kernel space to the feature space. They were abbreviated as KFCM-F and KFCM-K respectively [14].

KFCM-F Algorithm:

In KFCM-F, the prototype is retained to the feature space at the time of clustering. F stands for feature space. It minimizes the following objective function:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ik}^m ||\emptyset(x_k) - \emptyset(v_i)||^2 \quad (9)$$

subject to the following constraints,

$$u_{ik} \in [0,1], \forall i, k \quad (10)$$

$$0 < \sum_{k=1}^{n} u_{ik} < n, \forall i \quad (11)$$

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k \quad (12)$$

KFCM-F clustering algorithm serves the benefit that the data points are implicitly mapped to the kernel space via kernel learning. The optimization procedure follows Picard iteration technique. Restraining ourselves to Gaussian kernel, we have

K(x, x) = 1 and procure the following membership value:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}(\frac{1-K(x_k,v_i)}{1-K(x_k,v_j)})^{\frac{1}{m-1}}} \quad (13)$$

for i = 1,2,..., c and k = 1,2,..., n

Derivation of prototype depends exclusively on choice of kernel learning. Prototype value evaluated using Gaussian kernel is as follows :

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} K(x_k,v_i) x_k}{\sum_{k=1}^{n} u_{ik}^{m} K(x_k,v_i)} \quad (14)$$

The KFCM-F algorithm is analogous to the original FCM algorithm apart from the role played by the kernel learning in the former.

KFCM-K Algorithm:

The KFCM-K does not constrain the prototype in the feature space. It minimizes the following algorithm:

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} ||\emptyset(x_j) - v_i||^2 \quad (15)$$

The estimation of the partition matrix for i = 1,2,..., c and j = 1,2,..., n involving the constraints (10), (11), (12) generates the following partition matrix:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}(\frac{||\emptyset(x_k)-v_i||}{||\emptyset(x_k)-v_j||})^{\frac{2}{m-1}}} \quad (16)$$

Optimization of J taking Euclidean distance into consideration with respect to v_i, we obtain

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} \emptyset(x_k)}{\sum_{k=1}^{n} u_{ik}^{m}} \quad (17)$$

Value of prototype when Gaussian and polynomial kernel was considered i = 1,2,..., c is

$$\overline{v_i} = \frac{\sum_{k=1}^{n} u_{ik}^{m} K(x_k,\sigma_i) x_k}{\sum_{k=1}^{n} u_{ik}^{m} K(x_k,\sigma_i)} \quad (18)$$

$$\overline{v_i} = \frac{\sum_{k=1}^{n} u_{ik}^{m} (x_k^T \sigma_i + \theta) x_k}{(\sigma_i^T \sigma_i + \theta)^{p-1} \sum_{k=1}^{n} u_{ik}^{m}} \quad (19)$$

E. Possibilistic C-Means

Each component produced in Possibilistic C-Means (PCM) [22] harmonizes with a dense area in the feature space.

As iteration proceeds, the cluster prototypes are consequently drawn into these dense regions in feature space. All clusters are independent of the other cluster in PCM approach. Given a feature space X = $\{x_1, x_2, ...x_n\} \subset IR^p$. PCM aims at partitioning the data set X into c clusters by minimization of the following functional:

$$J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d^2(x_j,v_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n}(1-u_{ij})^m \quad (20)$$

subject to the conditions

$$u_{ij} \in [0,1], \forall\, i,j \quad (21)$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \quad (22)$$

where $d(x_j,v_i) = \|x_j - v_i\|$, v_i represents the prototype associated with the i^{th} cluster, u_{ij} is the typicality values which depends on all data, η_i is the bandwidth or scale or resolution parameter, $m > 1$ is the weighting exponent known as the possibilistic parameter. The first expression demands that the distance from data points to the prototype be as less as possible, whereas the second expression compels the membership matrix, U_{ij} to be large. The elements of U meet the above conditions. Unlike FCM, PCM relaxes the column sum constraint of FCM so that sum of each column satisfies a looser constraint

$$0 < \sum_{i=1}^{c} u_{ij} \leq c \quad (23)$$

The membership matrix U evaluated from PCM does not correspond to a partition matrix for this very reason. η_i must be particularized beforehand. Krishnapuram and Kellar [26] recommended the value of η_i as follows

$$\eta_i = \beta \frac{\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^2}{\sum_{j=1}^{n} u_{ij}^{m}}, \text{ for } 1 \leq i \leq c \quad (24)$$

represents the weighted mean of the intra cluster distance of the i^{th} cluster while the β is usually set as one.

The minimization of equation (20) with respect to u_{ij} leads to

$$u_{ij} = \frac{1}{1+\frac{d_{ij}}{\eta_i}^{\frac{1}{m-1}}}, \forall\, i,j \quad (25)$$

And the updates of prototype is as follows

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}}, \forall i \quad (26)$$

F. Kernelization Of Possibilistic C-Means Clustering

The original PCM delineated in (20) uses the Euclidean distance metric as its distance measure, but the Euclidean distance metric considers each feature of a data in the database crucial and not dependent on any other data point, which may not be appropriate for real world application, essentially while handling high dimensional data. A legitimate distance metric should have the ability to identify meaningful features and segregate admissible and inadmissible features. Furnishing a distance metric of this kind is highly specific of the problem in hand and decides the effectiveness of a learning algorithm. The notion of Kernel Gaussian learning is introduced [26] to compute the distance measure of PCM. The refurbished objective function is as follows:

$$J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} ||\emptyset(x_j) - \emptyset(v_i)||^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n}(1-u_{ij})^m \quad (27)$$

Minimization of the above function results in

$$u_{ij} = [1 + \frac{(2-2K(x_j,v_i))^{\frac{1}{m-1}}}{\eta_i}]^{-1}, \forall\, i,j \quad (28)$$

$$\emptyset(v_i) = \frac{\sum_{j=1}^{n} u_{ij}^{m} \emptyset(x_j)}{\sum_{j=1}^{n} u_{ij}^{m}}, \forall i \quad (29)$$

The prototypes in the kernel space could not be directly computed using (29), so $\emptyset(x_k)^T$ was right multiplied on both sides rewriting equation (29) thus

$$K(x_k, v_i) = \frac{\sum_{j=1}^{n} u_{ij}^{m} K(x_j,x_k)}{\sum_{j=1}^{n} u_{ij}^{m}}, \forall\, i,j \quad (30)$$

Correspondingly the updated η_i value in the high dimensional kernel space is

$$\eta_i = \beta \frac{\sum_{j=1}^{n} u_{ij}^{m} (2-2K(x_j,v_i))}{\sum_{j=1}^{n} u_{ij}^{m}}, \beta > 0 \quad (31)$$

Customarily β was chosen to be 1.

G. Intuitionistic Fuzzy C-Means

The abstraction [1] of intuitionistic fuzzy sets (IFS) and intuitionistic L-fuzzy sets (ILFS) were established as a result of the generalization of the concepts of fuzzy sets (FS). For a fixed set E, an intuitionistic L-fuzzy set (ILFS) A^* in E is an entity having the form

$$A^* = \{(x, \mu_A(x), v_A(x)) \mid x \in E\} \quad (32)$$

where the functions $\mu_A : E \to L$ and $\upsilon_A : E \to L$ defines the degree of membership and non-membership of an element $x \in E$ and $A^* \subset E$. The functions μ_A and υ_A should satisfy the following condition:

$$\forall x \in E, \mu_A(x) \leq N(\upsilon_A(x)) \quad (33)$$

where $N : L \to L$ is an involutive order reversing operation in a lattice $< L, \leq >$. When L = [0,1], the entity A^* is an intuitionistic fuzzy set (IFS), and the following condition holds :

$$\forall x \in E, 0 \leq \mu_A(x) + \upsilon_A(x) \leq 1 \quad (34)$$

In IFS, we can also define another function $\pi_A : E \to [0,1]$ by

$$\pi_A(x) = 1 - \mu_A(x) - \upsilon_A(x) \quad (35)$$

which corresponds to the hesitation degree of an element $x \in E$. Allude the fact that if there is no hesitation degree, i.e. $\pi_A(x) = 0$ or $\upsilon_A(x) = 1 - \mu_A(x)$ then A^* reduces to a fuzzy set (FS). In addition if $\upsilon_A(x) = 0$, A^* reduces to a crisp set.

H. Kernelization Of Intuitionistic Fuzzy C-Means Clustering

The proposed IFS can deal efficiently the uncertainty and vagueness correlated with real world data as it takes advantage of non-membership degree and hesitation degree along with membership degree for representing the real world data. Further the proposed intuitionistic fuzzy based clustering algorithm is more accurate and potent to noise and it converges rapidly in comparison to conventional FCM. Introduced Evolutionary Kernel IFCM clustering algorithm (EKIFCM) [23] considered as alternative to the fuzzy clustering method. It constitutes of two phases: Kernel Intuitionistic Fuzzy C-Means (KIFCM) and Parameter selection of KIFCM via GA. The objective function of KIFCM is as follows :

$$J(U,V) = \sum_{j=1}^{n} \sum_{i=1}^{N} (u_{ij}^{(l)})^{(\omega)} \times ||\emptyset(x_i), \emptyset(\theta_j^{(l)})||^2 \quad (36)$$

where $||\emptyset(x_i), \emptyset(\theta_j^{(l)})||$ is the Euclidean distance between a data point x_i and prototype in the kernel space.

Choosing Gaussian kernel as the kernel learning, the above equation gets reduced to the following EKIFCM associates with the minimized functional :

$$J(U,\theta) = 2\sum_{j=1}^{c} \sum_{i=1}^{N} (u_{ij}^{l})^{\omega} \times (1 - K(x_i, \theta_j^{(l)})) \quad (37)$$

Minimization leads to

$$u_{ij}^{(l+1)} = \left(\frac{1}{1-K(x_i,\theta_j^{(l+1)})}\right)^{\frac{1}{\omega-1}} \quad (38)$$

The intuitionistic membership value is obtained as

$$u_{ij}^{(l+1)^*} = u_{ij}^{(l+1)} + \pi_{ij}^{(l+1)} \quad (39)$$

and the prototype value is

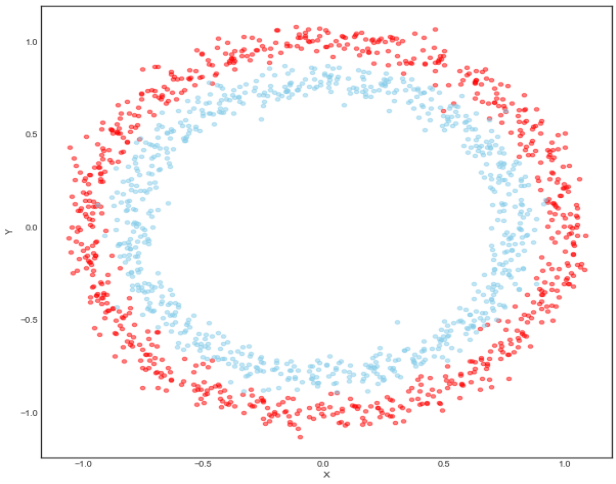
$$\theta_j^{(l+1)^*} = \frac{\sum_{i=1}^{N} u_{ij}^{(l+1)} K(x_i, \theta_j^{(l+1)}) x_i}{\sum_{i=1}^{N} u_{ij}^{(l+1)} K(x_i, \theta_j^{(l+1)})} \quad (40)$$

. IV. EXPERIMENTAL STUDIES

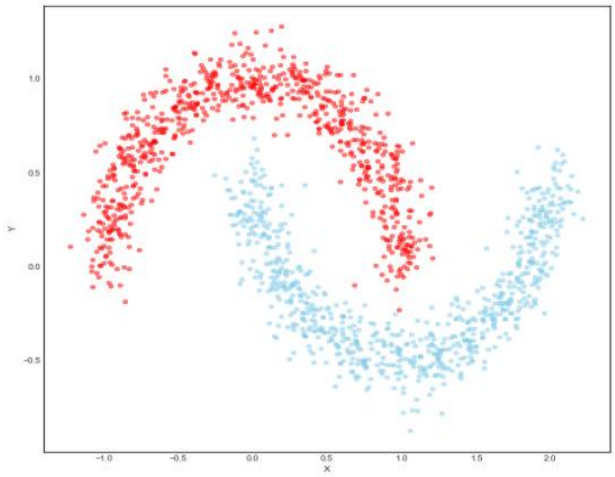
In this segment, a series of experiments on various data sets using standard K-Means, Fuzzy C-Means, Intuitionistic Fuzzy C-Means and their kernelized version given in [14] and [23] were taken into consideration. The main objective of this comprehensive analysis of experiments is to examine the performance gains with kernel based learning in K-Means, Fuzzy C-Means, and intuitionistic through classification rate.

A. Experimental Studies Using Synthetic Data Sets

Two synthetic data sets was considered in this study : Ring and Parabolic data sets. Experiments were performed using standard FCM and its kernelized versions in [14].



(a)



(b)

Fig 1. Synthetic data sets: (a) Ring data set (b) Parabolic data set

Clustering performance was observed over different value of parameters, which includes fuzzification coefficient m, cluster number k and kernel parameters m was varied over {1.2, 1.4, 1.7, 2, 2.5}. The Gaussian kernel parameter σ^2 was varied over the values {0.05, 1, 4, 8, 30, 100}. The polynomial kernel parameter θ and the value of d were restricted to {2, 10, 15, 20, 50} and {2, 4, 8, 12, 16} respectively. The clustering results are listed in tables 1 and 2 to observe the significant raise in classification rate on synthetic datasets.

The ring data set produced satisfactory performance gains with kernelized versions of generic clustering algorithms. Both KFCM-F and KFCM-K performed exceptionally better than FCM, and was able to categorize 100% of the ring data, though the performance

of KFCM-F was a bit discrepant. Marginal increment in classification rate was observed for parabolic data set.

B. Experimental Studies Using Machine Learning Database

Machine learning data sets [23] were taken into consideration includes: Haberman's Survival data, Iris Plant data, Wine data, Contraceptive Method Choice, SPECT heart data, Wisconsin Prognostic Breast Cancer and Glass identification database.

Table I. Clustering result for Parabolic data set

	Method	Parameter	Classification Rate(%)
Parabolic data set	FCM	k=2,m=2.5	87.4±0.00
	KFCM-F(G)	k=2,m=1.4,σ^2=1	88.2±0.00
	KFCM-K(G)	k=2,m=2,σ^2=1	89.0±0.00
	KFCM-K(P)	k=2,m=2.5,θ=10, d=12	87.8±0.00
	FCM	k=3,m=1.2	86.9±0.30
	KFCM-F(G)	k=3,m=1.4,σ^2=100	86.6±0.90
	KFCM-K(G)	k=3,m=2.5,σ^2=1	89.0±0.10
	KFCM-K(P)	k=3,m=1.4,θ=10, d=2	86.2±0.90
	FCM	k=4,m=1.2	88.1±0.00
	KFCM-F(G)	k=4,m=1.4,σ^2=30	87.9±0.00
	KFCM-K(G)	k=4,m=2.5,σ^2=1	89.0±0.10
	KFCM-K(P)	k=4,m=2.5,θ=20, d=16	88.5±0.00

Table II. Clustering result for Ring data set

	Method	Parameter	Classification Rate(%)
Ring data set	FCM	k=2,m=1.2	51.5±0.60
	KFCM-F(G)	k=2,m=2.5,σ^2=0.05	76.3±16.90
	KFCM-K(G)	k=2,m=2.5,σ^2=8	100.0±0.00
	KFCM-K(P)	k=2,m=2,θ=15, d=8	100.0±0.00
	FCM	k=3,m=2	62.5±2.70
	KFCM-F(G)	k=3,m=2,σ^2=0.05	88.2±12.60
	KFCM-K(G)	k=3,m=1.4,σ^2=4	100.0±0.00
	KFCM-K(P)	k=3,m=2,θ=2, d=4	100.0±0.00
	FCM	k=4,m=1.7	100.0±0.10
	KFCM-F(G)	k=4,m=1.4,σ^2=8	100.0±0.00
	KFCM-K(G)	k=4,m=1.4,σ^2=8	100.0±0.00
	KFCM-K(P)	k=4,m=1.4,θ=50, d=12	100.0±0.00

B. Experimental Studies Using Machine Learning Database

Machine learning data sets [23] were taken into consideration includes: Haberman's Survival data, Iris Plant data, Wine data, Contraceptive Method Choice, SPECT heart data, Wisconsin Prognostic Breast Cancer and Glass identification database. The performance gains while using kernel in clustering machine learning data sets produced heterogeneous outcome, and is encapsulated in Table 3. There is significant increment in classification rate by 96.4% for Glass Identification data when kernelized version of Intuitionistic Fuzzy C-Means was used in comparison to K-Means and an increment of 9.7% was observed for Iris data set. The kernel algorithms provides better enhancement in classification rate for Wine and Wisconsin Prognostic Breast Cancer data sets and considerable increment in classification rate was perceived for Haberman's Survival and SPECT Heart data sets. Kernel parameters plays a decisive role so that kernel based clustering algorithms can produce desired outcomes.

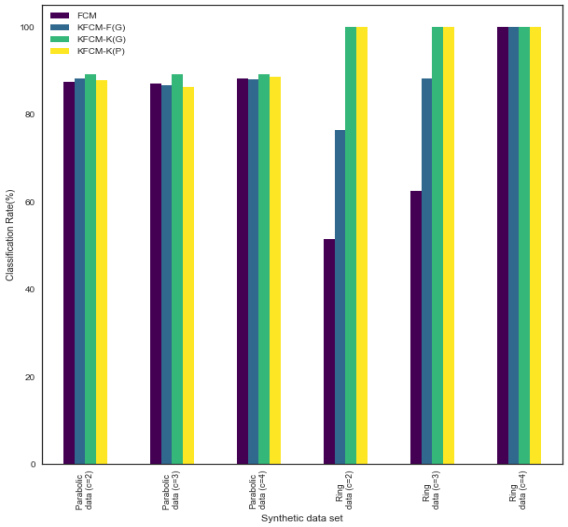


Fig 2. Graphical representation of the difference in classification rate for parabolic and ring data sets

Table III. Clustering result for Machine Learning data set

	Method	Parameter	Classification Rate(%)
Haberman's Survival Data	K-Means	k=2	0.5140±0.009
	FCM	k=2,m=2	0.5098±0.00
	IFCM	k=2,m=2,α=2	0.5196 ±0.00
	KFCM(G)	k=2,m=2.5,σ=0.5	0.7283±0.008
	EKIFCM	k=2,m=2.11, σ=49.04, α=11.01	0.7581±0.00
Iris	K-Means	k=3	0.8933±0.00
	FCM	k=3,m=2	0.8933±0.00
	IFCM	k=3,m=2,α=2	0.9000±0.00
	KFCM(G)	k=2,m=2.5,σ=100	0.8733
	EKIFCM	k=3,m=4.69, σ=3.92, α=2.89	0.9800±0.00
Wine	K-Means	k=3	0.7022±0.00
	FCM	k=3,m=2	0.6853±0.00
	IFCM	k=3,m=2,α=2	0.6910±0.00
	KFCM(G)	k=2,m=2,σ=150	0.7140
	EKIFCM	k=3,m=2.67, σ=188.13,α=13.59	0.7584±0.00
Contraceptive Method Choice	K-Means	-	0.3078±0.001
	FCM	k=3,m=2	0.3069±0.00
	IFCM	k=3,m=2,α=2	0.3021±0.00
	KFCM(G)	m=1.7,σ=1.1	0.3397±0.032
	EKIFCM	k=3,m=2.50, σ=2.16,α=3.36	0.4535±0.00
Wisconsin Prognostic Breast Cancer	K-Means	-	0.9585±0.00
	FCM	k=2,m=2	0.9542±0.00
	IFCM	k=2,m=2,α=2	0.9685±0.00
	KFCM(G)	k=2,m=2,σ=150	0.9709±0.001
	EKIFCM	k=2,m=4.42, σ=155.06, α=1	0.9742±0.00
SPECT heart data	K-Means	-	0.5131±0.068
	FCM	k=2,m=2	0.6030±0.00
	IFCM	k=2,m=2,α=2	0.6105±0.00
	KFCM(G)	k=2,m=2,σ=150	0.7254±0.002
	EKIFCM	k=2,m=19.42, σ=196.10, α=6.39	0.7940±0.00
Glass Identification Database	K-Means	-	0.4710±0.079
	FCM	k=6,m=2	0.5280±0.00
	IFCM	k=6,m=2,α=2	0.4953±0.00
	KFCM(G)	k=6,m=2,σ=150	0.4785±0.131
	EKIFCM	k=6,m=1.5, σ=50.79, α=6.06	0.9252±0.00

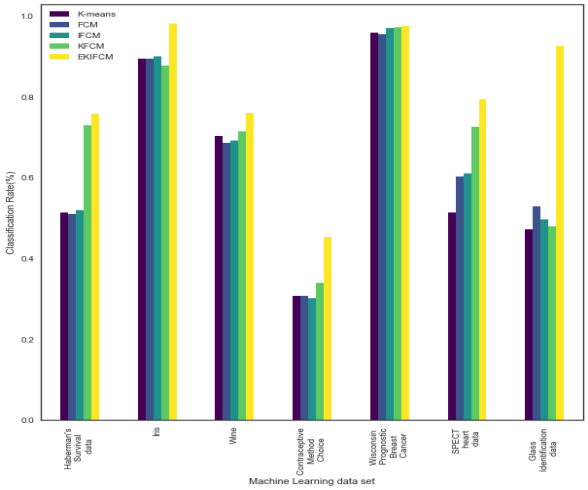


Fig 3. Graphical representation of the difference in classification rate for machine learning data sets

V. CONCLUSIONS

This paper examined the importance of kernel based learning in unsupervised clustering techniques of K-Means, Fuzzy C-Means, Possibilistic C-Means and Intuitionistic Fuzzy C-Means in clustering high dimensional databases with heavy noise. In order to analyze the advantages of kernel based learning, this paper performed the clustering with Ring data, Parabolic data set, Haberman's Survival data, Iris Plant database, Wine data, Contraceptive Method Choice, Wisconsin Prognostic Breast Cancer, SPECT heart data and Glass Identification Database. Once this paper implements the kernel based learning on proposed unsupervised clustering techniques, the increment in clustering accuracy was observed from the experiment results and the paper expressed well the technical reasons to understand the effect of kernel learning in clustering the different noisy databases.

ACKNOWLEDGEMENTS

This work was financially supported by DST India and MOST Israel.

REFERENCES

[1] Atanassov, Krassimir T. "Intuitionistic fuzzy sets." In *Intuitionistic fuzzy sets*, pp. 1-137. Physica, Heidelberg, 1999.
[2] James C Bezdek. Objective function clustering. In Pattern recognition with fuzzy objective function algorithms, pages 43-93. Springer, 1981.
[3] James Christian Bezdek. Fuzzy mathematics in pattern classification. Ph. D. Dissertation, Applied Mathematics, Cornell University, 1973.
[4] Battista Biggio and Fabio Roli. Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84:317-331, 2018.
[5] Jonathan H Chen and Steven M Asch. Machine learning and prediction in medicine beyond the peak of inated expectations. The New England journal of medicine, 376(26):2507, 2017.
[6] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu, and Tzong-Jer Chen. Fuzzy c-means clustering with spatial information for image segmentation. computerized medical imaging and graphics, 30(1):9-15, 2006.
[7] Rajesh N Dave. Fuzzy shell-clustering and applications to circle detection in digital images. International Journal Of General System, 16(4):343-355, 1990.
[8] Rajesh N Dave. Robust fuzzy clustering algorithms. In Fuzzy Systems, 1993., Second IEEE International Conference on, pages 1281-1286. IEEE, 1993.
[9] Inderjit S Dhillon, Yuqiang Guan, and Brian Kulis. Kernel k-means: spectral clustering and normalized cuts. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 551-556. ACM, 2004.
[10] Joseph C Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 1973.
[11] Hichem Frigui and Raghu Krishnapuram. Clustering by competitive agglomeration. Pattern recognition, 30(7):1109-1119, 1997.
[12] Guojun Gan, Jianhong Wu, and Zijiang Yang. A fuzzy subspace algorithm for clustering high dimensional data. In International Conference on Advanced Data Mining and Applications, pages 271-278. Springer, 2006.
[13] Isak Gath and Dan Hoory. Fuzzy clustering of elliptic ring-shaped clusters. Pattern recognition letters, 16(7):727-741, 1995.
[14] Daniel Graves and Witold Pedrycz. Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. Fuzzy sets and systems, 161(4): 522-543, 2010.
[15] Michael Gregory, Philip Kohn, J Shane Kippenhan, Enock Teefe, Jacob Morse, Venkata Mattay, Daniel Weinberger, Joseph Callicott, and Karen Berman. S215. Aggregating genetic and brain networks associated with risk for schizophrenia via spectral clustering of working memory activation and pgc2 loci. Biological Psychiatry, 83(9):S431, 2018.
[16] Donald E Gustafson and William C Kessel. Fuzzy clustering with a fuzzy covariance matrix. In Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on, pages 761-766. IEEE, 1979.
[17] Richard J Hathaway and James C Bezdek. Nerf c-means: Non-euclidean relational fuzzy clustering. Pattern recognition, 27(3):429-437, 1994.
[18] Richard J Hathaway and James C Bezdek. Fuzzy c-means clustering of incomplete data. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31(5):735-744, 2001.
[19] Ching-Wen Huang, Kuo-Ping Lin, Ming-Chang Wu, Kuo-Chen Hung, Gia-Shie Liu, and Chih-Hung Jen. Intuitionistic fuzzy c-means clustering algorithm with neighborhood attraction in segmenting medical image. Soft Computing, 19(2):459-470, 2015.
[20] Paraskevas Iatropoulos, Erica Daina, Manuela Curreri, Rossella Piras, Elisabetta Valoti, Caterina Mele, Elena Bresin, Sara Gamba, Marta Alberti, Matteo Breno, et al. Cluster analysis identi_es distinct pathogenetic patterns in c3 glomerulopathies/ immune complex-mediated membranoproliferative gn. J Am Soc Nephrol, 29:283-294, 2018.
[21] Stelios Krinidis and Vassilios Chatzis. A robust fuzzy local information c-means clustering algorithm. IEEE transactions on image processing, 19(5):1328-1337, 2010.
[22] Raghuram Krishnapuram and James M Keller. The possibilistic c-means algorithm: insights and recommendations. IEEE transactions on Fuzzy Systems, 4 (3):385-393, 1996.
[23] Kuo-Ping Lin. A novel evolutionary kernel intuitionistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy systems, 22(5):1074-1087, 2014.
[24] Enrique H Ruspini. A new approach to clustering. Information and control, 15 (1):22-32, 1969.
[25] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.
[26] Xiao-hongWu. A possibilistic c-means clustering algorithm based on kernel methods. In Communications, Circuits and Systems Proceedings, 2006 International Conference on, volume 3, pages 2062{2066. IEEE, 2006.
[27] Zhiwen Yu, Peinan Luo, Jane You, Hau-San Wong, Hareton Leung, Si Wu, Jun Zhang, and Guoqiang Han. Incremental semi-supervised clustering ensemble for high dimensional data clustering. IEEE Transactions on Knowledge and Data Engineering, 28(3):701-714, 2016.

AUTHORS PROFILE

Esha Kashyap studied Bachelor of Science from Burdwan University, West Bengal, India in 2013. She pursued Master degree from Pondicherry University in 2015. Ms Esha is currently pursuing her Ph.D. in Pondicherry University and she is working under the guidance of Dr. S. R. Kannan, Professor, Pondicherry University. She has two years of research experience. Her main research work focuses on Unsupervised Clustering in Data Analysis. She has knowledge in Mathematica, Python, Linux, Scilab, Fortran and Latex.

S.R. Kannan received his Ph.D. from Indian Institute of Technology (www.iitm.ac.in), Madras, India and PDF at DISI (www.disi.unige.it), University of Genova, Genova, Italy. Recently he has received post doctoral fellowship from department of Electrical Engineering, National Cheng Kung University (web.ncku.edu.tw), Taiwan. Presently, Dr. S.R. Kannan is working as Professor in Department of Mathematics, Pondicherry University (A Central University of India), India. He had been awarded a grant in the framework of a joint agreement between the Direzione Generale per la Cooperazione allo Sviluppo of the Italian Ministry of Foreign Affairs and the ICTP Programme for Training and Research in Italian Laboratories (www.ictp.it). He had been invited by Director General, National Agriculture Research Center, Tsukuba, Japan, for joint research work on remote sensing data to estimate rice yield (http://narc.naro.affrc.go.jp/narc-e/index.html). He has received two major research grants from UGC India, major research grant from CSIR India, bilateral research grants from DST India & NSC Taiwan for joint collaborative research project Indo Taiwan, and bilateral research grants from DST India & MOST Israel for joint collaborative research project Indo Israel.

Mark Last is a Full Professor at the Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Israel and the Head of the Data Science Research Center at Ben-Gurion University. Prof. Last obtained his Ph.D. degree from Tel Aviv University, Israel in 2000. Prior to starting his appointment at Ben-Gurion University in March 2001, Mark Last was a Visiting Assistant Professor at the Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA (1999-2001). Between the years 2009-2012, Prof. Last has served as the Head of the Software Engineering Program at Ben-Gurion University. He has published over 200 peer-reviewed papers and 11 books on data mining, text mining, and cyber security. Prof. Last is a Senior Member of the IEEE Computer Society and a Professional Member of the Association for Computing Machinery (ACM). He currently serves as an Associate Editor of *IEEE Transactions on Cybernetics* and an Editorial Board Member of *Data Mining and Knowledge Discovery*. From 2007 to 2016, he has served as an Associate Editor of *Pattern Analysis and Applications*. His main research interests are focused on data mining, cross-lingual text mining, soft computing, cyber intelligence, and medical informatics.