Pattern Recognition 45 (2012) 1092–1103

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Frameworks for multivariate m-mediods based modeling and classification in Euclidean and general feature spaces

Shehzad Khalid [a],*, Shahid Razzaq [b]

[a] Department of Computer Science and Engineering, Bahria University, Islamabad 44000, Pakistan
[b] Department of Computing, SEECS, National University of Science and Technology, Islamabad 44000, Pakistan

ARTICLE INFO

Article history:
Received 5 February 2011
Received in revised form
4 August 2011
Accepted 18 August 2011
Available online 31 August 2011

Keywords:
Multivariate m-mediods
Classification
Anomaly detection
Data mining
Dynamic modeling

ABSTRACT

This paper presents an extension of m-mediods based modeling technique to cater for multimodal distributions of sample within a pattern. The classification of new samples and anomaly detection is performed using a novel classification algorithm which can handle patterns with underlying multivariate probability distributions. We have proposed two frameworks, namely MMC-ES and MMC-GFS, to enable our proposed multivarite m-mediods based modeling and classification approach workable for any feature space with a computable distance metric. MMC-ES framework is specialized for finite dimensional features in Euclidean space whereas MMC-GFS works on any feature space with a computable distance metric. Experimental results using simulated and complex real life dataset show that multivariate m-mediods based frameworks are effective and give superior performance than competitive modeling and classification techniques especially when the patterns exhibit multivariate probability density functions.

1. Introduction

In recent research, there has been a growth of research attention aimed at the development of sophisticated approaches for pattern modeling and data classification. Detecting anomalous events is an important ability of any good classification system. Classification of unseen samples and anomaly detection require building models of normality. Once the models of normal classes are learnt, these can then be used for classifying new unseen trajectory data as normal (i.e. belonging to one of the modeled classes) or anomalous (not lying in the normality region of modeled classes).

Various machine learning techniques have been proposed for modeling of normal patterns and performing classification using the generated model of normality. Statistical approaches dealing with classification and anomaly detection are based on approximating the density of training data and rejecting test patterns that fall in regions of low density. Khalid and Naftel [9] and Hu et al. [7,8] models normal motion patterns by estimating single multivariate Gaussian for each class. Khalid and Naftel performs classification using Mahalanobis classifier and anomaly detection using Hotelling's test. In [7], the probability of a sample belonging to each pattern is calculated and the sample is classified to the pattern with the highest probability. However, if the probability of association of sample to the closest pattern is less then a threshold, the sample is deemed anomalous. Approaches using GMM to model normality distribution have also been proposed [1–3]. Various techniques [4–6] based on hidden Markov models (HMM) have also been presented for modeling and classification of temporal data. Owen and Hunter [10] use Self-Organizing Feature Maps (SOFM) to learn normal patterns. While classifying unseen samples, if the distance of the sample to its allocated class exceeds a threshold value, the trajectory is identified as anomalous. Marsland et al. [11] propose a novelty filter referred to as Grow When Required (GWR) network that uses SOM, based on habituation, to learn the environment and to discover novel features. The proposed approach is suitable for online use. GWR can add and delete nodes whenever the network in its current state does not sufficiently match the evolving pattern.

Approaches using support vector machines (SVM) have also been proposed [12–14]. These approaches are based on the principal of separating data belonging to different classes by identifying an optimal hyper plane between them. SVM based approaches involve computation of pairwise distances and time-consuming optimizations. Zhang et al. [12] propose a hybrid approach using SVM and nearest neighbor classifier for content based image recognition with the multiclass setting. Rasch et al. [13] perform classification and anomaly detection using one-class SVM. Various Mahalanobis distance metric learning approaches have also been applied for data clustering and classification [15–19]. Certain discriminative learning methods such as Fisher Discriminative Learning (FDA) [20] have been proposed for improving the performance of classifier. FDA is a supervised dimensionality reduction mechanism which computes a transformation matrix to maximize inter-class and minimize intra-class

* Corresponding author.
E-mail address: shehzad_khalid@hotmail.com (S. Khalid).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.08.021

scatter. FDA caters for the global distribution of the pattern and gives poor performance in the presence of multivariate distribution of samples within a pattern. An extension of FDA, referred to as local FDA (LFDA) [21] has been proposed to work in the presence of multivariate distribution of samples within a pattern by preserving the local structure of data. LFDA based supervised dimensionality reduction has been combined with well known classifiers such as GMM to improve classifier's performance [22]. However, LFDA can only be applied in feature spaces with a calculable mean and is not applicable to general feature spaces.

In our previous work [28], we have proposed *m*-mediods based Modeling and Classification approach for features in Euclidean Space (MC-ES). *m*-mediods based approach models a pattern by a set of cluster centers of mutually disjunctive sub-classes (referred to as mediods) within the pattern. The modeling mechanism is influenced by HSACT-LVQ based clustering mechanism as proposed in our previous work [9]. It has been shown that hierarchical semi-agglomerative approach using a neural network, such as HSACT-LVQ, outperforms hard clustering techniques such as *k*-Means. *k*-Means is initialized with the number of cluster centers that are equivalent to the expected number of groupings in the dataset. The cluster centers itself is normally initialized to a randomly picked sample from dataset. This type of hard clustering does not guarantee that the network will identify and distinguish all major groupings. *k*-Means may organize the cluster centers to represent variations within one major grouping of the data by allocating more than one cluster center to that group. This may be caused by the initialization of more than one cluster center close to or within the distribution of a single pattern. Allocation of more than one output neuron to a pattern will result in having a single cluster center representing multiple patterns. *k*-Means, therefore, produce poor clustering and classification results due to poor initialization. On the other hand, our HSACT-LVQ algorithm avoids this problem by initializing itself with greater number of cluster centers than the number of groupings to be identified in the dataset. Finer clusters are then merged, based on their similarities, to generate coarse clusters representing the desired number of sub-classes (mediods). The modeling technique, proposed in [28] referred to as *m*-mediods modeling, models the class containing n members with *m*-mediods known *a-priori*. Once the *m*-mediods model for all the classes have been learnt, the MC-ES approach performs classification of new samples and anomaly detection by checking the closeness of said samples to the models of different classes using hierarchical classifier. The anomaly detection module required specification of threshold which is used globally for all the patterns. However, this approach had unaddressed issues like manual specification of threshold for anomaly detection, identification of appropriate value of threshold for anomaly detection and anomaly detection of patterns with different scale and orientation which is used globally for all the patterns. These issues are addressed by a localized *m*-mediods based approach (LMC-ES) as proposed in [23] which enables us to automatically select a local significance parameter for each pattern taking into consideration the distribution of individual patterns.

LMC-ES can effectively handle patterns with different orientation and scale and has been shown to give superior performance than competitors including GMM, HMM and SVM based classifiers. However, there are still open issues (i) Modeling, classification and anomaly detection in the presence of multivariate distribution of samples within a pattern, (ii) Soft classification in the presence of multimodal pattern distribution to minimize misclassification, (iii) Modeling and classification in feature spaces for which we can not compute mean.

The contribution of this paper is to present an extension of *m*-mediods based modeling approach, wherein the multimodal distribution of samples in each pattern is represented using multivariate *m*-mediods. An approach for multivariate model-based classification and anomaly detection is also presented. The proposed mechanism is based on a soft classification approach which enables the proposed multivariate classifier to adapt to the multimodal distribution of samples within different patterns. We have proposed two frameworks for multivariate *m*-mediods based modeling and classification applicable to two different feature spaces:

1. Finite dimensional features in Euclidean space
2. General feature spaces with a computable pairwise similarity measure

This enables our multivariate *m*-mediods based approach to be used for classification and anomaly detection in any feature space with a given distance function.

The remainder of the paper is organized as follows. In Section 2, an overview of the general working of proposed multivariate *m*-mediods based modeling and classification approach is presented. Section 3 presents a framework of multivariate modeling and classification for finite dimensional features in Euclidean space with a calculable mean. In Section 4, a modification of multivariate modeling and classification framework to operate in any feature space with a computable similarity function is presented. Comparative evaluation of proposed multivariate *m*-mediods and previously proposed localized *m*-mediods [23] based frameworks is presented in Section 5. Experiments have been performed to show the effectiveness of proposed system for modeling, classification and anomaly detection in the presence of multimodal distribution of samples within a pattern, as compared to competitors. These experiments are reported in Section 6. The last section summarizes the paper.

2. Overview of our classification approach

Classification and anomaly detection in the presence of multivariate distributions of sample within a pattern is a challenging task. Fig. 1 gives an overview of our general multivariate modeling and classification framework to effectively cope with this challenge. The proposed classifier, like any other classifier is composed of two main modules: construction of multivariate *m*-mediods based model to cater for variation in distribution of samples belonging to a particular pattern and using the generated model for classification of unseen samples and anomaly detection. The module for construction of *m*-mediods based model is composed of three steps. In step 1, we model a pattern using a set of *m*-mediods representing mutually disjunctive sub-classes, possessing different probability densities. The proposed approach is motivated by the observation that any distribution of samples within a pattern can be represented by well separated and distributed representative samples (mediods). The resulting model can be visualized as a bag of quantized sub-samples belonging to the pattern. In multivariate settings, there is a variation in density of samples belonging to a pattern. Our approach tends to identify mediods in a fashion that the number of identified mediods in different parts of the distribution is proportional to the density of samples. The approach to identify the mediods is different for different feature spaces. After the identification of m -mediods, we identify the set of possible normality ranges for each modeled pattern (in step 2) to be used later for classification and anomaly detection. Instead of identifying a single normality range at a pattern level, we propose to automatically determine a normality range at a mediod level customized according to the sample distribution around a given mediod (step 3). Hence, a single pattern will be containing multiple mediods having different normality ranges enabling the proposed

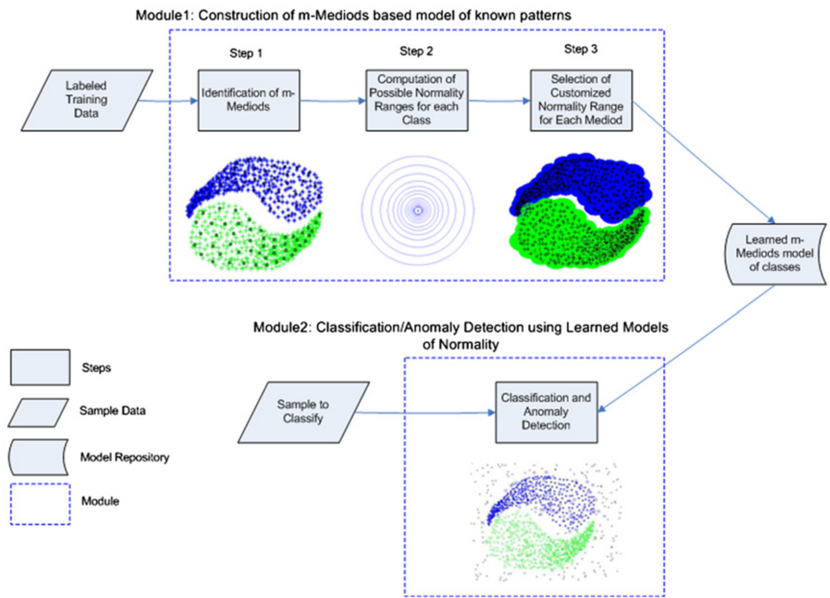

Fig. 1. Overview of our proposed m-mediods based modeling and classification framework.

approach to model variations in the distribution of samples within a pattern. The multivariate m-Mediods based models of patterns learned in first module is used by second module for classifying unseen samples and anomaly detection. The classification and anomaly detection in Euclidean feature spaces and generic feature spaces is based on the adaption of merged and unmerged approach respectively, as proposed in our previous work [23].

3. Multivariate modeling and classification for finite dimensional features in Euclidean space (MMC-ES)

In this section, we present a framework for multivariate modeling and classification using m-mediods approach that is applicable to features in Euclidean space with calculable mean.

3.1. Multivariate m-mediods based modeling

Given a feature space representation of training data for each pattern, we wish to model the underlying distribution of samples within a pattern using training data. Modeling of pattern using multivariate m-mediods approach in Euclidean space is a three step process, (i) identification of m-mediods, (ii) computation of set of possible normality ranges for the pattern and (iii) selection of customized normality range for each mediod.

3.1.1. Step 1: identification of m-mediods

The algorithm for identification of mediods using finite dimensional features in Euclidean space is based on the adaptation of neural gas based learning algorithm [24]. Let $DB^{(i)}$ be the labeled training instances associated to pattern i and W the weight vector associated to each output neuron. The modeling algorithm comprises the following steps:

1. Initialize the SOM network with a greater number of output neurons than the desired number of mediods m to avoid the modeling algorithm from stucking into the problems of local minima. A series of experiments have been conducted, using patterns with different statistical distributions, to determine a good value for the number of output neurons. We have observed that a good value for the number of output neurons can be obtained as:

$$\#_{output} = \begin{cases} \xi & \text{if } \xi < 150 \wedge \xi > (m \times 2) \\ m \times 2 & \text{if } \xi < (m \times 2) \\ 150 & \text{if } \xi > 150 \end{cases} \quad (1)$$

where $\xi = size(DB^{(i)})/2$.
2. Initialize weight vectors W_i (where $1 \le i \le \#_{output}$) from the $PDF\ N(\mu,\Sigma)$ estimated from training samples in $DB^{(i)}$.
3. Identify k-Nearest Weights (k-NW) to current training sample using:

$$k\text{-}NW(F,\mathbf{W},k) = \{C \in \mathbf{W} \,|\, \forall R \in C, S \in \mathbf{W}-C, \|F-R\| \le \|F-S\| \wedge |C| = k\} \quad (2)$$

where F is the training sample, $\mathbf{W}$ is the set of weight vectors, C is the set of k closest weight vectors, $\|.\|$ is the Euclidean distance function. The k closest output neurons to F is updated in the specific iteration of learning process. For a given training cycle t, $k=\delta(t)$ where $\delta(t)$ is a neighborhood size function whose value decreases gradually over time as specified in Eq. (5).
4. Train the network by updating a subset of the weights (C) using

$$W_c(t+1) = W_c(t) + \alpha(t)\zeta(j)\|F-W_c(t)\| \quad \forall W_c \in C \quad (3)$$

where W_c is the weight vector representation of output neuron c, j is the order of closeness of W_c to F ($1 \le j \le k$), $\zeta(j,k) = \exp(-(j-1)^2/2k^2)$ is a membership function that has value 1 when $j=1$ and falls off with the increase in the value of j, $\alpha(t)$ is the learning rate of SOM and t is the training cycle index.
5. Decrease the learning rate $\alpha(t)$ exponentially over time using:

$$\alpha(t) = 1 - e^{2(t-t_{max})/t_{max}} \quad (4)$$

where t_{max} is the total number of learning iterations.
6. Decrease the neighborhood size exponentially with training iterations as:

$$\delta(t) = \left\lceil \delta_{init}\left(1 - e^{2(t-t_{max})/t_{max}}\right) \right\rceil \quad (5)$$

where δ_{init} is the neighborhood size at the start of learning process. We set $\delta_{init} = 5$ after rigorous experimental evaluation.
7. Iterate through steps 3–6 for all the training iterations.

8. Ignore output neurons with zero membership.
9. Select the closest pair of weight vectors (i,j) (indexed by (a,b)) given by the condition

$$(a,b) = \arg\min_{(i,j)} [(W_i - W_j)^T (W_i - W_j) \times (|W_i| + |W_j|)]^{1/2} \quad \forall i,j \wedge i \neq j \tag{6}$$

where $|.|$ is the membership count function. Scaling the distance between two weight vectors by their membership counts discourages the merging of weight vectors that are modeling a dense distribution of samples. Merge the selected pair of weight vectors using

$$W_{ab} = \frac{|W_a| \times W_a + |W_b| \times W_b}{|W_a| + |W_b|} \tag{7}$$

10. Repeat step 9 till the number of weight vectors gets equivalent to $\#_{mediods}$. Append weight vector W_k to the list of mediods $\mathbf{M}^{(i)}$ modeling the pattern i.

3.1.2. Step 2: computation of possible normality ranges

After the identification of mediods $\mathbf{M}^{(i)}$ for pattern i, we intend to identify and pre-compute a set of possible normality ranges for a given pattern. Values of normality ranges for a given pattern is determined by the inter-mediod distances within a given pattern. Hence, different patterns will have different set of possible normality ranges depending on the distribution of samples, and in turn mediods, within a pattern. In this step, a set of possible normality ranges $\mathbf{D}^{(i)}$ for the pattern i is computed as follows:

1. Identify the closest pair of mediods (i,j) (indexed by (p,q)) from $\mathbf{M}^{(c)}$ as follows:

$$(p,q) = \arg\min_{(i,j)} dist(M_i, M_j) \quad \forall i,j \wedge i \neq j \tag{8}$$

where dist(.,.) is the distance function which is Euclidean for MMC-ES framework.

2. Populate the distance array for the current number of mediods using

$$\mathbf{D}_l^{(c)} = (p, q, dist(M_p, M_q)) \tag{9}$$

where l is the current number of mediods.

3. Merge the closest pair of mediods using

$$M_{pq} = \frac{|M_p| \times M_p + |M_q| \times M_q}{|M_p| + |M_q|} \tag{10}$$

4. Iterate through steps 1–3 till the number of mediods gets equivalent to 1.

3.1.3. Step 3: selection of customized normality range for each mediod

After the identification of mediods and a set of possible normality ranges for a given pattern, we select different normality range for each mediod depending on the distribution of samples from the same and different patterns around a given mediod. The normality range is selected to minimize false positives (false identification of training samples from other pattern as a normal member of pattern that is being modeled) and false negatives (classification of normal samples of the pattern being modeled as anomalies). The algorithm for selection of customized normality range for each mediod, to enable multivariate m-mediod based modeling of pattern, comprises of following steps:

1. Initialize significance parameter τ with the number of mediods (m) used to model pattern c.
2. Sequentially input labeled training instances belonging to all classes and identify the closest mediod, indexed by r, using:

$$r = \arg\min_k dist(Q, M_k) \quad \forall k \tag{11}$$

where Q is the test sample.

3. Perform an anomaly test using the anomaly detection system, as proposed in Section 3.2, assuming a one class classifier containing only pattern c represented by mediods set $\mathbf{M}^{(c)}$.
4. Increment false positive count $FP(r)$, corresponding to closest mediod M_r, each time when the sample is a normal member of pattern c but is identified as anomalous.
5. Increment false negative count $FN(r)$, corresponding to closest mediod M_r, each time when the sample is misclassified to pattern c.
6. Iterate through steps 2–5 for all the samples in DB.
7. Calculate Significance Parameter Validity Index ($SPVI$) to check the effectiveness of current value of τ for a particular mediod using:

$$SPVI(k,\tau) = \beta \times FP(k) + (1-\beta) \times FN(k) \quad 0 \leq \beta \leq 1 \quad \forall k \tag{12}$$

where β is a scaling parameter to adjust the sensitivity of proposed classifier to false positives and false negatives according to specific requirements.

8. Set $\tau = \tau - 1$.
9. Iterate through steps 2–8 till $\tau = 1$.
10. Identify the value of significance parameter for a given mediod as:

$$\widehat{\tau}_{(c,k)} = \arg\min_{\tau} \quad SPVI(\tau, k) \quad \forall M_k \in \mathbf{M}^{(c)} \tag{13}$$

where $\widehat{\tau}_{(c,k)}$ is the dynamic significance parameter that have a different normality range for each mediod depending on the local density.

The space complexity of the proposed modeling algorithm is $O(|DB| +_{output})$. For large datasets, $|DB| \gg \#_{output}$ and the space complexity reduces to $O(|DB|)$. The time complexity of our algorithm is the sum of time complexities of the three steps and is equivalent to $O((t_{max} * \#_{output} * \log(\#_{output}))) + O((\#_{mediods} * \log(\#_{mediods}))) + O(|DB|^2 * \#_{mediods} * \log(\#_{mediods}))$ where

- $\#_{output} * \log(\#_{output})$ is the time complexity of ranking of nodes w.r.t. the closeness to the training sample in each iteration
- $\#_{mediods} * \log(\#_{mediods})$ is the time complexity of possible normality range
- $|DB|^2 * \#_{mediods} * \log(\#_{mediods})$ is the time complexity for selecting customized normality range for each mediod.

3.2. Classification and anomaly detection

Once the multivariate m-mediods based model for all the classes have been learnt, the classification of unseen samples to known classes and anomaly detection is performed by checking the closeness of unseen sample to the models of different classes. We use a k-NM classifier in which the unseen sample is posed as a query to the entire set of mediods ($\mathbf{M}$) belonging to different classes and k nearest mediods from $\mathbf{M}$ are identified. Instead of applying a voting mechanism as in case of conventional k-NN classifier, we try to classify the sample to the class of the closest mediod given it lies in the customized normality range of the mediod as identified during the modeling phase. However, if the sample lies outside the normality range of the closest mediod, we check its closeness w.r.t. the next closest mediod. This process continues till the sample lies in the normality range of one of the k mediods in which case it is classified as a normal member of the corresponding mediod. The sample is marked as anomalous if the sample does not fall in the normality range of any of the k closest mediods. The value of k in the proposed k-NM classifier is chosen using the experimental analysis as suggested in Section 6.5. The classification and anomaly detection algorithm, in the presence of multivariate settings, comprises the following steps:

1. Identify k nearest mediods to unseen sample Q as:
$$k{-}NM(Q,\mathbf{M},k) = \{\mathbf{C} \in \mathbf{M} \,|\, \forall R \in \mathbf{C}, S \in \mathbf{M}-\mathbf{C},\ Dist(Q,R) \le Dist(Q,S) \wedge |\mathbf{C}| = k\} \quad (14)$$
where $\mathbf{M}$ is the set of all mediods from different classes and $\mathbf{C}$ is the ordered set of k closest mediods starting from the nearest mediod.
2. Initialize nearest mediod index $\imath$ to 1.
3. Set r to the index of $\imath^{th}$ nearest mediod and c to the index of its corresponding class.
4. Initialize index l with m.
5. Select the most similar pair of mediods in $\mathbf{M}^{(c)}$ along with the distance between them, for the current number of mediods l as:
$$(p,q,d_{pq}) = \mathbf{D}_l^{(c)} \quad (15)$$
where d_{pq} contains the distance between mediods indexed by p and q.
6. Q is classified as normal member of class c if:
$$Dist(Q,M_r) \le d_{pq} \quad (16)$$
7. If the condition specified in Eq. (16) is not satisfied, decrement the index l by 1.
8. Merge the most similar pair of mediods using
$$M_{pq} = \frac{|M_p| \times M_p + |M_q| \times M_q}{|M_p| + |M_q|} \quad (17)$$
9. Iterate steps 3–8 till l gets equivalent to the significance parameter $\widehat{\tau}_{(c,k)}$. If the test trajectory Q has yet not been identified as a valid member of class c, it is considered to be an outlier w.r.t. to the mediod r belonging to class c.
10. Increment the index $\imath^{th}$ index by 1.
11. Iterate steps 3–10 till $\imath$ gets equivalent to k. If Q has not been classified as a normal member of any class, it is marked as anomalous.

The time complexity of MMC-ES based classification and anomaly detection algorithm is $O(|\mathbf{M}| + O(k*(m*\log(m) - \tau*\log(\tau)))$ for anomalous samples where $|\mathbf{M}|$ is the total number of mediods in $\mathbf{M}$. However, for most of the normal samples the time complexity is $O(|\mathbf{M}| + O(m*log(m) - \tau*log(\tau))$. The time complexity can be further reduced by using efficient indexing structure like kd-trees to index $|\mathbf{M}|$ mediods for efficient k-NM search.

4. Multivariate modeling and classification for general feature spaces with a computable pairwise similarity measure (MMC-GFS)

The framework of multivariate m-mediods based modeling and classification, as presented in Section 3, works only with feature spaces with calculable mean. However, for complex feature spaces, it is not always possible to calculate a mean. This section provides a modified multivariate m-mediod based framework for any feature space, given that there is a computable pairwise similarity measure.

4.1. Multivariate m-mediods based modeling

The proposed algorithm for modeling of pattern in general feature spaces is a three step process.

4.1.1. Step 1: Identification of m-mediods

The algorithm for identification of mediods using finite dimensional features in general feature space with a computable similarity matrix is based on the affinity propagation based clustering algorithm [25]. Let $DB^{(i)}$ be the classified training samples associated to pattern i and W the weight vector associated to each output neuron. The modeling algorithm comprises the following steps:

1. Form the affinity matrix $A \in R^{n\times n}$ defined by
$$A(a,b) = \begin{cases} \exp\left(\dfrac{-dist(s_a,s_b)}{2\sigma^2}\right) & \text{if} a \ne b \\ P(a) & \text{otherwise} \end{cases} \quad (18)$$
Here s_i and s_j are the feature vector representation of training samples, σ is the scaling parameter and $P(a)$ is the preference parameter indicating the suitability of sample a to be selected as a mediod. We set $P(a)$ to the median of affinities of sample a with n samples.
2. Initialize availability matrix $\Im(a,b) = 0 \quad \forall a,b$.
3. Update responsibility matrix $\Re$ as
$$\Re(a,b) = A(a,b) - \max_{\forall c \text{ s.t. } b \ne c} \{\Im(a,c), A(a,c)\} \quad (19)$$
4. Update availability matrix $\Im$ as
$$\Im(a,b) = \begin{cases} \min\{0, \Re(b,b) + \sum_{\forall c \text{ s.t. } c \ne a \wedge c \ne b} \{0, \Re(c,b)\}\} & \text{if } a \ne b \\ \sum_{\forall c \text{ s.t. } a \ne c} \max\{0, \Re(c,a)\} & \text{otherwise} \end{cases} \quad (20)$$
5. Identify the exemplar for each sample as
$$\xi_a = \arg\max_b [\Im(a,b) + \Re(a,b)] \quad (21)$$
6. Iterate through steps 3–5 till the algorithm is converged or maximum number of learning iterations (t_{max}) is exceeded. The algorithm is considered to have converged if there is no change in exemplar identification for certain number of iterations ($t_{convergence}$).
7. If the number of exemplars identified are smaller than the desired number of mediods, set higher values of preference and vice versa. The algorithm is repeated till the desired number of exemplars are identified. An appropriate value for identification of desired number of mediods is searched using a bisection method.
8. Append exemplars ξ_a to the list of mediods $\mathbf{M}^{(i)}$ modeling the pattern i.

4.1.2. Step 2: computation of possible normality ranges

After the identification of mediods $\mathbf{M}^{(i)}$ for pattern i, a set of possible normality ranges $\mathbf{D}^{(i)}$ for the pattern i is computed using the Step 2 of modeling algorithm as specified in Section 3.1.2. However, instead of using Euclidean distance as distance function in Eqs. (8) and (9), appropriate distance function for a particular feature space should be incorporated.

4.1.3. Step 3: selection of customized normality range for each mediod

Customized normality ranges for each mediod, to enable multivariate m-mediod based modeling of pattern, is selected using the algorithm similar to the one presented in Section 3.1.3. A distance function appropriate to a given feature space has to be utilized in Eq. (11). Instead of performing the anomaly detection in step 3 using MMC-ES based anomaly detection system, we apply the anomaly detection algorithm of MMC-GFS framework as proposed in Section 4.2.

The space complexity of the proposed modeling algorithm in general feature space is $O(3{*}n^2)$. The time complexity of our algorithm is the sum of time complexities of the three steps. The time complexity of step 1 is $O(\omega{*}(n^2+n^2{*}\log(n)))$ where

- $O(n^2)$ is the time complexity of affinity matrix computation
- $O(n^2{*}\log(n))$ is the time complexity of message passing to compute availability and responsibility matrix
- ω is the number of times the modeling algorithm is repeated to identify m mediods. It has been observed that the value of ω normally lies in the range 3–10.

The time complexity of steps 2 and 3 is similar to the time complexity of modeling algorithm as specified in Sections 3.1.2 and 3.1.3, respectively.

4.2. Classification and anomaly detection

Once the m-mediods based model for all the classes have been learnt, the classification of unseen samples to known classes and anomaly detection is performed using following steps:

1. Identify k nearest mediods, from the entire set of mediods ($\mathbf{M}$) belonging to different classes, to unseen sample using Eq. (14).
2. Initialize nearest mediod index $\imath$ to 1.
3. Set r to the index of $\imath^{th}$ nearest mediod and c to the index of its corresponding class.
4. Initialize index l with the value of significance parameter $\widehat{\tau}_{(c,k)}$.
5. Identify the closest pair of mediods in set $\mathbf{M}^{(c)}$ and their corresponding distance, for the current number of mediods l, using $\mathbf{D}^{(c)}$ as:
$$(p,q,d_{pq}) = \mathbf{D}_l^{(c)} \tag{22}$$
where d_{pq} contains the distance between mediods indexed by p and q.
6. Test sample Q is considered to be a valid member of class c if:
$$Dist(Q,M_r) \le d_{pq} \tag{23}$$
7. If the condition specified in Eq. (23) is not satisfied, increment the index $\imath$ by 1.
8. Iterate steps 3–7 till $\imath$ gets equivalent to k. If the test trajectory Q has not been identified as a valid member of any class, it is considered to be an outlier and deemed anomalous.

The time complexity of MMC-ES based classification and anomaly detection algorithm is $O(|\mathbf{M}|)+O(k)$ for anomalous samples However, for most of the normal samples the time complexity is $O(|\mathbf{M}|)$. The time complexity can be further reduced by using efficient indexing structure like kd-trees to index $|\mathbf{M}|$ mediods for efficient k-NM search.

5. Relative merits of proposed modeling and classification algorithms

In this section, we provide a comparative evaluation of the proposed multivariate m-mediods and localized m-mediods [23] based frameworks (LMC-ES) for modeling, classification and anomaly detection. These frameworks can be characterized in terms of the following attributes:

- Ability to deal with multimodal distribution within a pattern
- Time complexity of generating m-mediods based model of known patterns
- Time complexity of classification and anomaly detection using learned models of normality
- Scalability of modeling mechanism to cope with increasing number of training data

For the ease in understanding of the comparative analysis, simulation of the working of proposed modeling and classification algorithms for arbitrary shaped patterns having multimodal distributions is presented in Fig. 2. In the left image of Fig. 2, each point represents the training sample and instances belonging to the same class are represented with same color and marker. Squares superimposed on each group of instances represent the mediods used for modeling the pattern. Normality region generated using different frameworks for classification and anomaly detection is depicted in the right image of Fig. 2. Test sample is considered to be a normal member of the class if it lies within the normality region, else it is marked as anomalous. Visualization of LMC-ES, MMC-ES and MMC-GFS based modeling is provided in Fig. 2(a)–(c) respectively.

Multivariate modeling using MMC-ES and MMC-GFS frameworks caters for the multimodal distribution within a pattern. On the other hand, LMC-ES framework always assumes a unimodal distribution within a pattern and hence can not cater for the dynamic distribution of samples within a pattern. It is apparent from Fig. 2 that MMC-ES and MMC-GFS frameworks have generated more accurate models that have accommodated the variation in sample density within a given pattern. Particularly, normality region generated using MMC-ES framework appears to be least affected by the multivariate distribution of training samples.

LMC-ES framework performs a hard classification of unseen sample. A sample is classified to a pattern represented by the majority of mediods from a set of k nearest mediods. The sample may not lie in the normality region of a pattern to which it is classified and hence deemed anomalous. However, it is likely that it may still fall in the normality region of the second closest but less dense pattern having larger normality range. The hardness of LMC-ES based classification algorithm will result in the misclassification of such samples. However, the classification and anomaly detection algorithms proposed in MMC-ES and MMC-GFS do not give a hard decision and checks for the membership of test w.r.t. different patterns until it is identified as a valid member of some pattern or it has been identified as anomalous w.r.t. k nearest mediods. This relatively softer approach enables the MMC-ES and MMC-GFS based classification algorithm to adapt to the multimodal distribution of samples within different patterns. This phenomena is highlighted in Fig. 3. The samples, represented by '$\times$' marker, will be classified to blue pattern but is marked as anomalous using LMC-ES classifier as it falls outside the normality range of dense mediods belonging to the closest pattern. On the other hand, soft classification technique as proposed in MMC-ES and MMC-GFS frameworks will correctly classify the sample as normal members of green pattern.

Algorithms to generate m-mediods model, as proposed in MMC-ES framework and LMC-ES framework, are efficient and scalable to large datasets. On the other hand, the modeling algorithm of MMC-GFS is relatively inefficient and is not scalable to large datasets due to the requirement of affinity matrix computation. The space and time complexity is quadratic which is problematic for patterns with large number of training sample. The complexity problem can be catered by splitting the training sample into subsets and selecting candidate mediods in each subset using algorithm specified in Section 4.1.1. The final selection of mediods can be done by applying the same algorithm again but now using the candidate mediods instead of all the training sample belonging to a given pattern. The classification algorithm of MMC-GFS framework is relatively efficient as compared to MMC-ES framework. This efficiency gain is due to the non-iterative unmerged anomaly detection with respect to a given mediod. The anomaly detection is done by applying a single

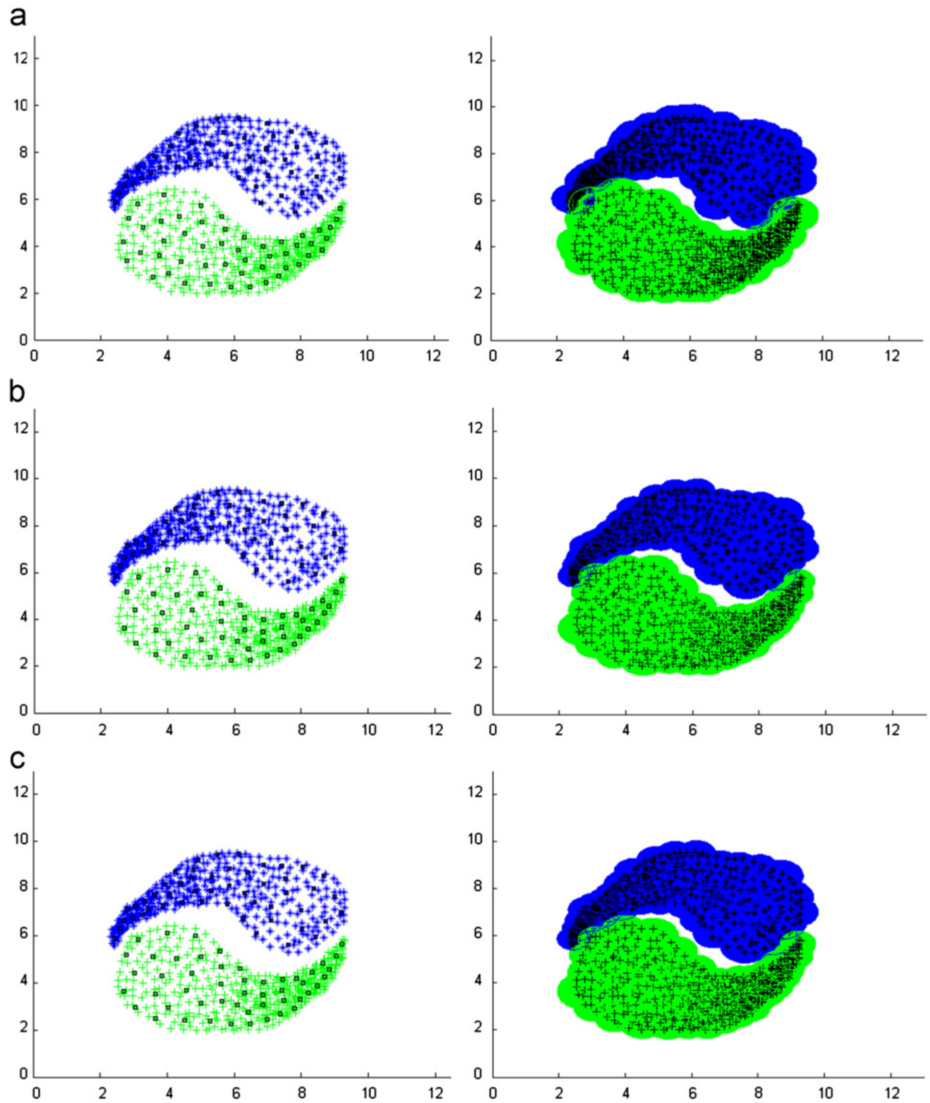

Fig. 2. m-mediods based modeling of patterns using (a) LMC-ES framework, (b) MMC-ES framework, (c) MMC-GFS framework. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

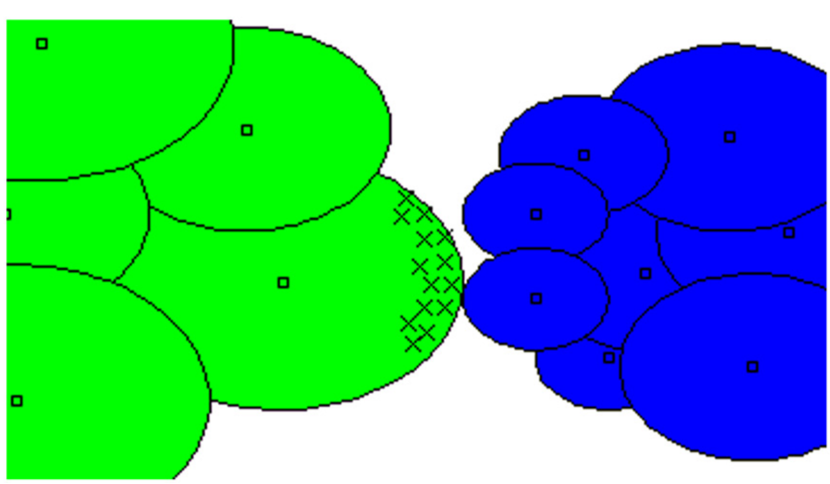

Fig. 3. Scenario for evaluating the adaptation of classification algorithms as proposed in different m-mediods based frameworks.

threshold to the distance of the test sample from its ι^{th} closest mediod as specified in Eq. (23). On the other hand, MMC-ES implements iterative merged anomaly detection, which is more accurate but time consuming. as compared to the modeling algorithm proposed in MMC-GFS framework. The time complexity of merged anomaly detection is $O(m{*}\log(m)-\tau{*}\log(\tau))$.

6. Experimental results

In this section, we present some results to analyze the performance of the proposed multivariate *m*-mediods based modeling, classification and anomaly detection as compared to competitive techniques.

6.1. Experimental datasets

Experiments are conducted on synthetic SIM_1 and SIM_2 and real life ASL [9,26–30] datasets. Details of these datasets can be found in Table 1.

6.2. Experiment 1: evaluation of m-mediods based frameworks for classification and anomaly detection

The purpose of this experiment is to evaluate the performance of proposed MMC-ES, MMC-GFS and LMC-ES based frameworks for classification of unseen data samples to one of the known patterns. The effectiveness of the proposed frameworks to perform anomaly detection is also demonstrated here. The experiment has been conducted on simulated SIM_1 and SIM_2 dataset. Training data from simulated datasets is shown in Fig. 4. Test data for SIM_1 dataset is obtained by generating 500 samples from a uniform distribution such that $(x,y) \in (U(1,12),U(1,12))$. Test data for SIM_2 dataset is obtained by generating 1000 samples from a uniform distribution such that $(x,y) \in (U(0,20),U(0,20))$.

We have used 50 mediods to model a class using its member samples. The classification and anomaly detection results for SIM_1 and SIM_2 datasets, using LMC-ES, MMC-ES and MMC-GFS frameworks are presented in Fig. 5(a)–(c) respectively. Training samples are represented using '+' marker whereas classified normal samples are represented by small circles. Data points belonging to same class are represented with same color and marker. Samples from test data which have been identified as anomalous are represented with a black '×' marker. It is apparent from Fig. 5 that multivariate *m*-mediods based classification system as proposed in MMC-ES and MMC-GFS framework performs better classification and anomaly detection while catering for multimodal distribution within the modeled pattern. On the other hand, LMC-ES based framework performs univariate modeling of patterns and therefore the classification system does not adjust well to the variation of density within a pattern, specifically for multivariate class distributions in SIM_1 dataset, as highlighted in Fig. 5(a).

6.3. Experiment 2: comparison of proposed classifiers with competitive techniques

The purpose of this experiment is to compare the performance of classifiers as proposed in LMC-ES, MMC-ES and MMC-GFS frameworks. For comparison of our results with competitive techniques, we establish a base case by implementing three different systems for comparison including Mahalanobis, GMM and LFDA-GMM classifier. Real life ASL dataset is used for the experiment. Signs from different numbers of word classes are selected. Classified training data is obtained by randomly selecting half of the samples from each word class leaving the other half to be used as test data. Trajectories from ASL dataset are represented using DFT-MOD based coefficient feature vectors [28]. Patterns are modeled using 20 mediods per pattern. We have computed single multivariate Gaussian for modeling of patterns for Mahalanobis classifier. Modeling of patterns and classification of unseen samples using GMM is based on the approach as described in [30]. Each class is modeled using a separate GMM. The number of modes to be used for GMM-based modeling is automatically estimated using a string of pruning, merging and mode-splitting processes as specified in [30]. We have implemented LFDA-GMM classifier as proposed in [22]. After modeling of normal patterns,

Table 1
Overview of datasets used for experimental evaluation.

Dataset	Description	# of trajectories	Extraction method	Labeled (Y/N)
SIM_1	Simulated dataset comprising of two dimensional coordinates	Arbitrary #	Simulation	Y
SIM_2	Simulated dataset comprising of two dimensional coordinates generated randomly along two concentric circles	Arbitrary #	Simulation	Y
ASL	Trajectories of right hand of signers as different words are signed. Dataset consists of signs for 95 different word classes with 70 samples per word.	6650	Extracting (*x*,*y*) coordinates of the mass of right hand from files containing complete sign information	Y

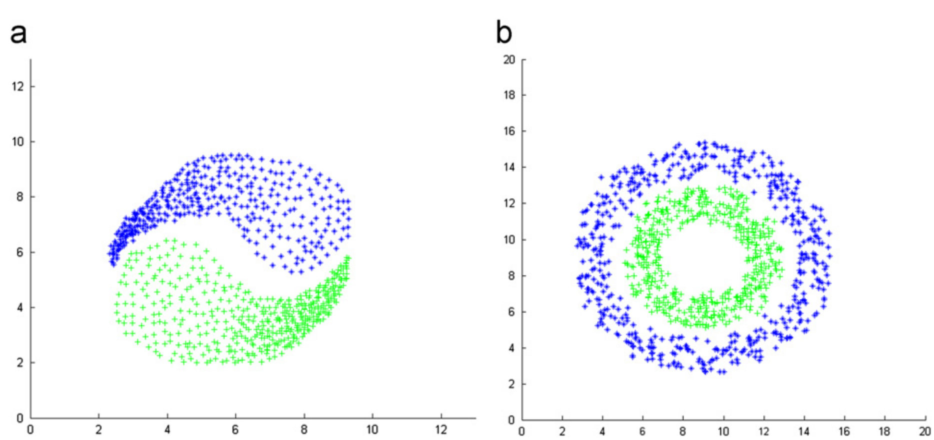

Fig. 4. Training data from (a) SIM_1 dataset and (b) SIM_2 dataset.

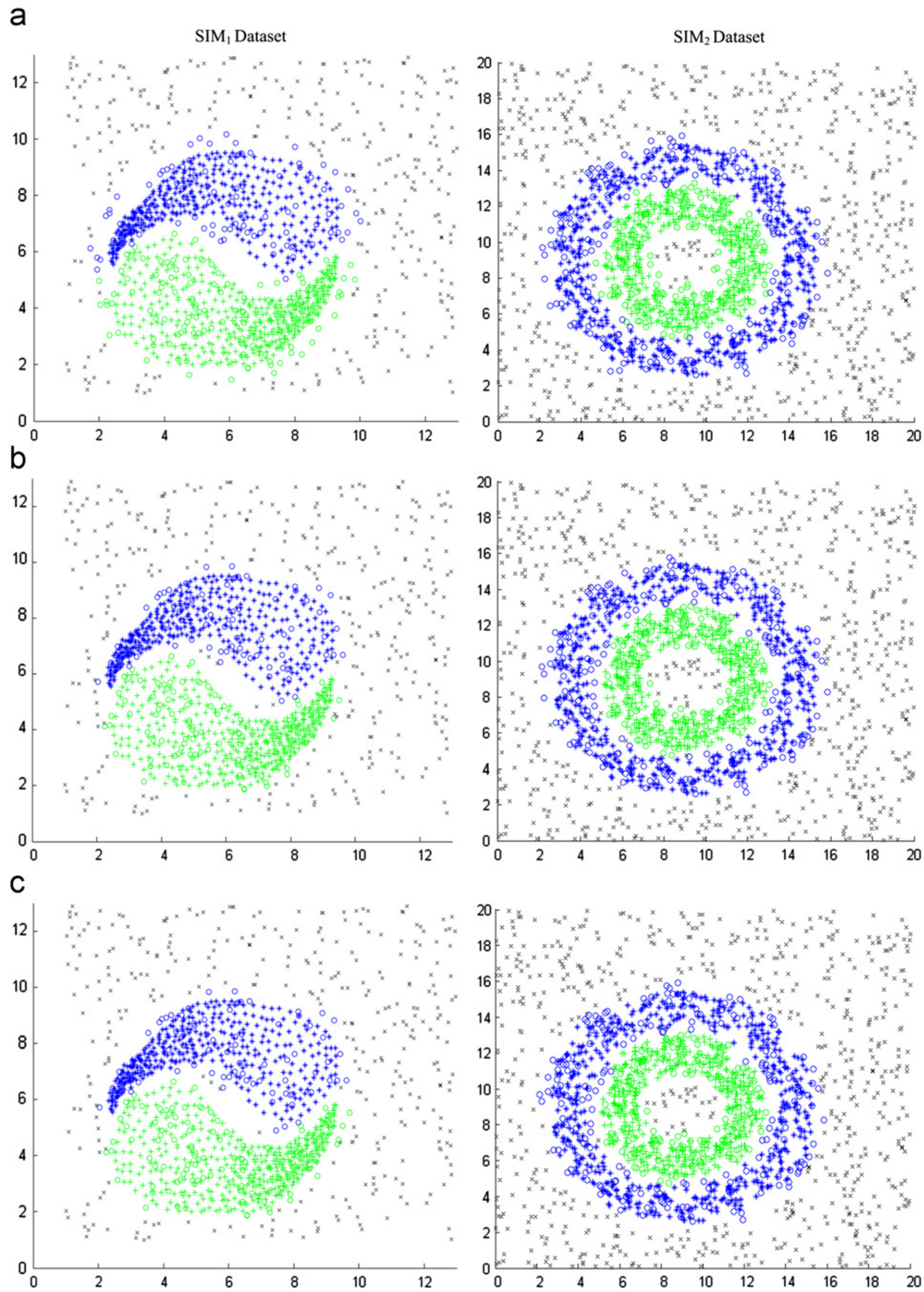

Fig. 5. Classification of test data, based on SIM_1 and SIM_2 classes, using (a) LMC-ES framework, (b) MMC-ES framework, (c) MMC-GFS framework.

the test data is classified using different approaches. Classification accuracies are computed by comparing the classification results with the ground truth. The experiment is repeated with different numbers and combinations of word classes. Each classification experiment is averaged over 50 runs to reduce any bias resulting from favorable word selection.

The accuracy of different classifiers for wide range of word classes from ASL dataset is presented in Table 2. Based on these results, we can see that the multivariate m-mediods based classifier as proposed in MMC-ES and MMC-GFS frameworks yield a superior classification accuracy as compared to other classifiers closely followed by univariate LMC-ES framework and LFDA-GMM. GMM yields good results for lower number of classes but its performance deteriorates for higher number of word classes. However, applying LFDA based supervised transformation before generating GMM model enhances the effectiveness of GMM classifier as obvious from the classification accuracies of LFDA-GMM classifier. It can also be observed from Table 2 that the relative accuracy of proposed m-mediods based MMC-ES, MMC-GFS and LMC-ES classifiers increases with an increase in the number of classes as compared with competitive techniques; thus making them more scalable for larger number of classes. The superior performance of MMC-ES and MMC-GFS, as compared to competitive techniques, can be explained by the fact that the proposed multivariate m-mediods based frameworks do not impose any restriction on the probability distribution function of modeled patterns. The proposed frameworks can effectively model arbitrary shaped patterns and can effectively handle variation in sample distribution within a pattern as demonstrated in Figs. 2 and 5. On the other hand, the competitive approaches impose assumptions on the PDF of patterns (normally Gaussian). These approaches do not have the capacity to handle multivariate distribution within a pattern. As a result, the

Table 2
Classification accuracies for different number of classes from ASL dataset.

Classifiers	ASL (#classes:#samples)				
	2:70	4:140	8:280	16:560	24:840
MMC-ES	0.99	0.95	0.92	0.88	0.85
MMC-GFS	0.99	0.94	0.91	0.86	0.83
LMC-ES	0.98	0.92	0.88	0.83	0.78
Mahalanobis	0.95	0.88	0.82	0.75	0.71
GMM	0.97	0.92	0.83	0.74	0.69
LFDA-GMM	0.98	0.91	0.88	0.82	0.79

model generated by these approaches will not give an accurate representation of complex patterns and hence result in poor classification performance as compared to the proposed multivariate m-mediods based approaches.

Similar experiment with ASL dataset (using similar experimental settings) has been conducted by Bashir et al. [29] using their proposed GMM and HMM-based classification system. They reported classification accuracies of 0.96, 0.92, 0.86 and 0.78 for 2, 4, 8 and 16 word classes respectively. A comparison of these classification accuracies with the results obtained using our approach reveals that classifiers from m-mediods classifier family performs better than GMM and HMM-based recognition system [29] despite the fact that our proposed classification approach is conceptually simpler and computationally less expensive.

6.4. Experiment 3: quantitative evaluation of anomaly detection algorithms

Here we provide a quantitative evaluation and comparison of m-mediods based anomaly detection algorithms, as proposed in MMC-ES, MMC-GFS and LMC-ES frameworks, with competitors. We implemented three different anomaly detection techniques based on statistical test as proposed in [9], Grown When Required (GWR) novelty filter as proposed in [11] and one-class classifier based anomaly detection as proposed in [31]. Naftel et al. [9] performs anomaly detection by using Mahalanobis classifier and conducting Hotelling's T^2 test. Tax et al. [31] perform anomaly detection by generating model of one class (referred to as target class) and distinguishing it from samples belonging to all other classes. There generation of model of the target class is done using SVM and GMM. For SVM-based one class classifier (OCC-SVM), we have used RBF kernel for the modeling of target class. For GMM-based one class classifier (OCC-GMM), we have used the approach as specified in Experiment 2 to generate the GMM-based model.

The experiment has been conducted using different numbers of word classes from ASL dataset. We have extracted half of the samples belonging to each word class for training purposes leaving the other half of the samples to be used as test data. DFT-MOD based coefficient feature vector representation of sign trajectories from training data is generated and used to generate models as required by the different classification approaches. MMC-ES, MMC-GFS and LMC-ES framework based model of each class is generated using the algorithm as presented in Sections 3.1, 4.1 and [23] respectively. Patterns are modeled using 20 mediods per pattern.

Once the model learning phase is over, anomaly detection using different techniques is carried out using test dataset. We would expect that few instances drawn from class X would be recorded as anomalous when tested against the same class, whereas nearly all instances would be detected as anomalous when tested against a different class Y. The experiment is repeated with different numbers and combinations of word classes. Each anomaly detection experiment is averaged over 50 runs to reduce any bias resulting from favorable word selection.

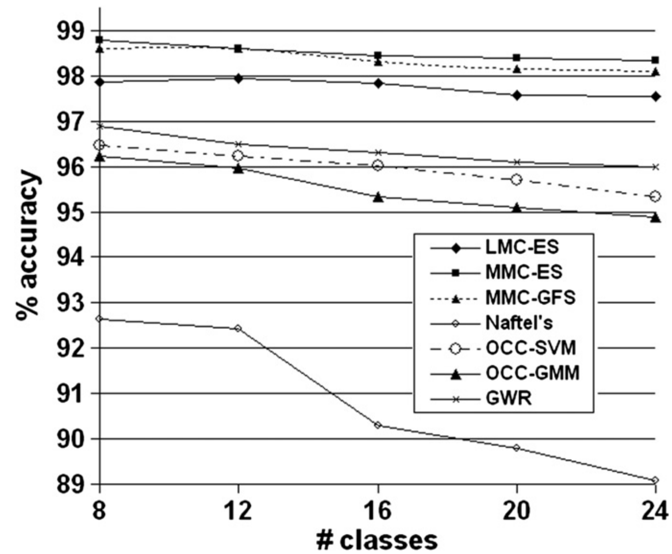

Fig. 6. Percentage anomaly detection accuracies for different number of classes from ASL dataset.

Fig. 6 reports the result in terms of percentage of correct anomaly detection using various number of word classes from ASL dataset. The results demonstrate the superiority of anomaly detection using m-mediods based MMC-ES, MMC-GFS and LMC-ES frameworks. The anomaly detection accuracies obtained using multivariate anomaly detection algorithms as presented in MMC-ES and MMC-GFS frameworks are higher than univariate LMC-ES based anomaly detection algorithm. MMC-ES, MMC-GFS and LMC-ES perform better than OCC-SVM, OCC-GMM, GWR and Mahalanobis framework-based Naftel's method. The superior performance of proposed approach as compared to state-of-the-art techniques is due to the fact that our approach gives importance to correct classification of normal sample and to the filtration of abnormal samples during the model generation phase. On the other hand, OCC-SVM generates good model of normal classes but classifies many of the abnormal samples as member of normal classes whereas GWR gives extra importance to filtering abnormal samples and in the process, identifies many normal samples as abnormal.

6.5. Experiment 4: evaluation of parameters of proposed algorithms

The purpose of this experiment is to evaluate the effect of important parameters on the performance of proposed classification and anomaly detection algorithm. The accurate modeling of known classes is dependant on the number of mediods m used to model a given class. Similarly, classification using k nearest mediod is somewhat dependant on the value of k. We will analyze the effect of selecting different values of these parameters on classification and anomaly detection accuracy of proposed algorithm. We have selected MMC-ES framework for the evaluation of these parameters.

The experimental setup is similar to the one as presented in Experiment 3. We have used fixed number (*#classes* = 24) of ASL classes in the experiment. To analyze the effect of m on the modeling process, we have repeated the experiment using different number of mediods. Classification and anomaly detection accuracies using proposed k-NM (k=9) approach for different value of m are presented in Fig. 7. The accuracy is low for small values of m and increases with the increase in the number of m used to model the pattern. However, increasing the value of m

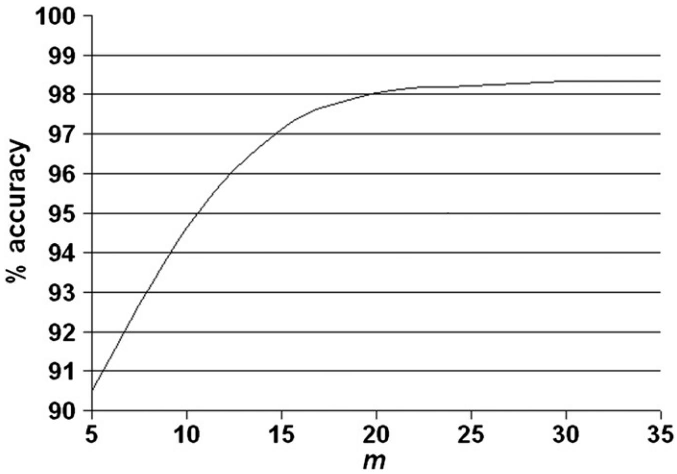

Fig. 7. Percentage classification and anomaly detection accuracies for different values of m.

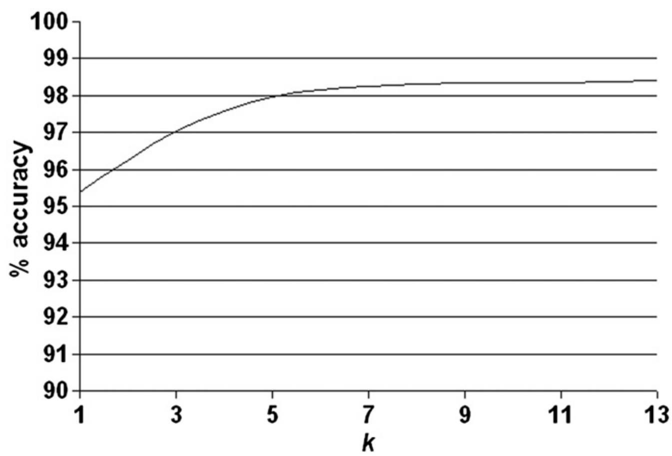

Fig. 8. Percentage classification and anomaly detection accuracies for different values of k.

above certain level does not have significant impact on accuracy whilst increasing the computational complexity of proposed algorithm. This is explained by the fact that using smaller number of mediods to model a pattern will result in generation of coarse representation of a give pattern and hence poor classification and anomaly detection accuracies. Increasing the number of mediods will result in modeling of finer details of multivariate distribution of samples within pattern and hence more accurate results. Based on these results, we recommend to set $m \geq 20$ for good classification results. Although, higher values of m results in higher computational complexity, the k-NM based classification can always be speeded up with indexing. To analyze the effect of different value of k on the accuracy of proposed k-NM approach, we have repeated the above experiment by keeping the value of m constat ($m=5$) but varying the values of k and the results are presented in Fig. 8. The classification and anomaly detection accuracies increase with the increase in the value of k and flattens out for higher values of k.

7. Discussion and conclusions

In this paper, we have presented an extension of localized m-mediods based modeling technique to cater for multimodal distribution of samples within a pattern. The strength of the proposed approach is its ability to model complex patterns without imposing any restriction on the distribution of samples within a given pattern. Once the multivariate m-mediods model for all the classes have been learnt, the classification of new trajectories and anomaly detection is then performed using a proposed soft classification and anomaly detection algorithm which is adaptive to multimodal distributions of samples within a pattern. Two variations of multivariate m-mediods based framework, namely MMC-ES and MMC-GFS, are proposed which enables the proposed approach to be used for modeling, classification and anomaly detection in any feature space with a computable similarity function. MMC-ES is a specialized framework tuned for feature vector spaces with a computable mean whereas MMC-GFS is a general framework for any feature space with a computable affinity matrix.

Experimental results are presented to show the effectiveness of proposed multivariate MMC-ES and MMC-GFS frameworks based classifiers. Modeling of pattern and classification using proposed frameworks is unaffected by variation of sample distribution within a pattern as demonstrated in Fig. 5. Quantitative comparison of MMC-ES and MMC-GFS based classifiers with competitive techniques demonstrates the superiority of our multivariate approach as it performs consistently better than commonly used Mahalanobis, GMM and HMM-based classifiers.

Experiments are also conducted to show the effectiveness of anomaly detection capabilities of proposed frameworks. Anomaly detection results for different classes of ASL datasets, using different variants of proposed anomaly detection algorithm, are presented. It has been shown that anomaly detection using multivariate MMC-ES and MMC-GFS frameworks gives better anomaly detection accuracies as compared to the univariate LMC-ES approach. Although LMC-ES enables the anomaly detection system to adapt to the normality distribution of individual classes, it is insensitive to the variation of distributions within a pattern which results in degradation of its performance as compared to MMC-ES and MMC-GFS frameworks. Comparison of proposed anomaly detection algorithms with an existing approach demonstrates the superiority of our approach as they consistently perform better for different number of classes.

References

[1] T. Brotherton, T. Johnson, G. Chadderdon, Classification and novelty detection using linear models and a class dependent elliptical basis function neural network, in: Proceedings of IJCNN Conference, vol. 2, 4–9 May 1998, pp. 876–879.
[2] S. Roberts, L. Tarassenko, A probabilistic resource allocating network for novelty detection, Neural Computation 6 (1994) 270–284.
[3] D.Y. Yeung, C. Chow, Parzen window network intrusion detectors, in: IEEE International Conference on Pattern Recognition, Canada 2002, pp. 385–388.
[4] T. Xiang, S. Gong. Video behaviour profiling and abnormality detection without manual labelling, in: IEEE International Conference on Computer Vision, London, UK, vol. 2, October 2005, pp. 1238–1245.
[5] T. Xiang, S. Gong, Incremental visual behaviour modelling, in: Sixth IEEE Workshop on Visual Surveillance, Graz, May 2006, pp. 65–72.
[6] D. Zhang, Gatica-Perez, S. Bengio, I. McCowan, Semi-supervised adapted HMMs for unusual event detection, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2005, pp. 611–618.
[7] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S. Maybank, A system for learning statistical motion patterns, IEEE Transactions on Pattern Analysis and Machine Learning 28 (9) (2006) 1450–1464.
[8] W. Hu, D. Xie, Z. Fu, W. Zeng, S. Maybank, Semantic based surveillance video retrieval, IEEE Transactions on Image Processing (2007) 1168–1181.
[9] S. Khalid, A. Naftel, Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space, Multimedia Systems 12 (3) (2006) 227–238.
[10] J. Owens, A. Hunter, Application of the self-organising map for trajectory classification, in: Proceedings of Third IEEE International Workshop on Visual Surveillance, Dublin, Ireland, July 2000, pp. 77.
[11] S. Marsland, J. Shapiro, U. Nehmzow, A self-organising network that grows when required, Neural Networks 15 (8–9) (2002) 1041–1058.
[12] H. Zhang, A.C. Berg, M. Maire, J. Malik, SVM-KNN: Discriminative nearest neighbor classification for visual category recognition, in: CVPR, 2006.

[13] G. Ratsch, S. Mika, B. Scholkopf, K. Muller, Constructing boosting algorithms for SVMs: an application for one-class classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (9) (2002) 1184–1199.
[14] C.P. Diehl, J.B. Hampshire II, Real-time object classification and novelty detection for collaborative video surveillance, in: Proceedings of IEEE IJCNN Conference, 2002.
[15] E. Xing, A. Ng, M. Jordan, S. Russell, Distance metric learning with application to clustering with side-information, NIPS, vol. 15, 2003, pp. 505–512.
[16] K. Weinberger, J. Blitzer, L. Saul, Distance metric learning for large margin nearest neighbor classification, NIPS, vol. 18, 2006, pp. 1473–1480.
[17] S. Xiang, F. Nie, C. Zhang, Learning a Mahalanobis distance metric for data clustering and classification, Pattern Recognition 14 (2008) 3600–3612.
[18] A. Bar-Hillel, T. Hertz, N. Shental, D. Weinshall, Learning distance functions using equivalence relations, in: International Conference on Machine Learning, 2003, pp. 11–18.
[19] L. Yang, R. Jin, R. Sukthankar, Y. Liu, An efficient algorithm for local distance metric learning, in: AAAI, 2006.
[20] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed., Academic Press, Inc., Boston, 1990.
[21] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis, Journal of Machine Learning Research 8 (2007) 1027–1061.
[22] J. Yu, Localized Fisher discriminant analysis based complex chemical process monitoring, 2010, doi: 10.1002/aic.12392.
[23] S. Khalid, Activity classification and anomaly detection using m-mediods based modeling of motion patterns, Pattern Recognition 43 (10) (2010) 3636–3647.
[24] T.M. Martinetz, S.G. Berkovich, K.J. Schulten, Neural-gas network for vector quantization and its application to time-series prediction, IEEE Transactions on Neural Networks 4 (4) (1993) 558–569.
[25] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science 3154 (581) (2007) 972–976.
[26] F.I. Bashir, A.A. Khokhar, D. Schonfeld, View-invariant motion trajectory based activity classification and recognition, ACM Multimedia Systems 1 (2006) 45–54 (Special Issue on Machine Learning Approaches to Multimedia Information retrieval).
[27] F.I. Bashir, A.A. Khokhar, D.Schonfeld, HMM based Motion Recognition System using Segmented PCA, in: IEEE International Conference on Image Processing, Genova, Italy, September 11–14, 2005, pp. 1288–1291.
[28] S. Khalid, Motion based behaviour learning profiling and classification in the presence on anomalies, Pattern Recognition 43 (1) (2010) 173–186.
[29] F.I. Bashir, A.A. Khokhar, D. Schonfeld, Object trajectory-based activity classification and recognition using hidden Markov models, IEEE Transactions on Image Processing 16 (7) (2007) 1912–1919.
[30] F.I. Bashir, A.A. Khokhar, D. Schonfield, Automatic object trajectory based motion recognition using Gaussian mixture models, in: IEEE International Conference on Multimedia and Expo, Netherland, July 6–8, 2005, pp. 1532–1535.
[31] D.M.J. Tax, One-class Classification, Ph.D. Thesis, Delft University of Technology, ⟨http://ict.ewi.tudelft.nl/davidt/thesis.pdf⟩, June 2001.

Shehzad Khalid graduated from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2000. He received the M. Sc. degree from National University of Science and Technology, Pakistan in 2003 and the Ph.D. degree from the University of Manchester, UK, in 2009. He is currently an Associate professor at the Bahria University of Management and Computer Sciences, Pakistan. His research interests includes: dimensionality reduction, indexing and retrieval, profiling and classification, trajectory-based motion learning profiling and classification, computer vision, machine learning.

Shahid Razzaq received the BS degree in Computer System Engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan in 2000 and an MS degree in Computer Science from University of Washington in 2008. He worked at Microsoft Corporation from 2001 to 2008 at the Redmond campus near Seattle, USA. He is currently a lecturer in the Department of Computing at the School of Electrical Engineering and Computer Science (SEECS), a constituent of the National University of Sciences and Technology, Islamabad, Pakistan. His research interests include machine learning and pattern recognition.