Genetic Algorithm Framework for Spike Sorting

Sajjad Farashi
Shahid Beheshti University of Medical Sciences, Faculty of Medicine, Tehran,Iran
E-mail: farashi@sbmu.ac.ir

Mohammad Mikaili
Department of Biomedical Engineering, Faculty of Engineering, Shahed University, Tehran, Iran
E-mail: mikaili@shahed.ac.ir

Abstract—Spike sorting involves clustering spikes according to the similarity of their shapes. Usually the sorting procedure is carried out by extracting appropriate features of neuronal spikes. In this study a new spike sorting procedure based on genetic algorithm is developed which contains two distinct phases. In the first phase a B-spline curve is fitted to each spike waveform and then the optimal features are selected from parameters of fitted B-spline curves. The genetic algorithm is used for searching the optimal parameters of B-spline curve in a way that the curve fitting error is minimized. In the second phase, clustering of spikes based on extracted features is performed by applying genetic algorithm. In this phase the fitness function is defined in a manner that both spatial distances between objects in the feature space and their similarity in the real world are considered. The proposed sorting method is tested on the real neural dataset which firstly are classified by an expert human. The results show that the proposed method based on genetic algorithm framework gives fewer errors of clustering in comparison with some other approaches currently used in the clustering purposes.

Index Terms—Neural data, spike sorting, feature extraction, clustering, B-spline, genetic algorithm.

I. INTRODUCTION

For exploring the mechanisms of nervous system functions, it is essential to record the electrical activity of neurons called spike. It is assumed that each neuron produces a distinct and reproducible electrical activity which can be considered as its finger print [1]. When more than one neuron is recorded by a single microelectrode, the recorded signal is the mixture of all neurons that surround the electrode. It is essential to separate the spikes events generated by each neuron. Such procedure is called spike sorting.

The most important step in sorting procedure consists of choosing suitable features with the optimal discrimination capability. These features transform spikes to a new space called feature space which sorting can be conducted. The feature extraction is carried out to diminish the redundancy of information in the real world and to concentrate on the differences between data originated from different sources. Projecting dataset to the feature space creates some clouds of points.

Clustering is a technique for grouping data points into disjoint groups named clusters. The data points in each cluster must have similar origin and be different from data points of other clusters. When an available dataset is projected to a feature space by optimal features, it is expected that the neighbor data in real world belong to the same cluster in the feature space. It means that optimal features conserve the distance between real world data in the feature space. This leads to feature space of compact and well -separated clusters.

During the past decade spike sorting techniques have been the subject of extensive development and so far several solutions have been advanced. Based on the assumption of optimal feature selection, some kind of clustering algorithms tries to allocate neighbor points in the feature space to distinct clusters. Such algorithms fail to achieve proper clustering results when clusters overlap or there is small separation between clusters[2]. The algorithms like *k*-means clustering try to minimize the within-cluster sum of squared distances. In this class of algorithms the number of clusters should be defined then in an iterative procedure the centers of clusters are adjusted. The main drawback of such algorithms is the necessity for determining the exact number of clusters which inappropriate choice may leads to unexpected results. This class of clustering methods usually is appropriate for spherical shape clusters. When the size of clusters are different and there are many outliers, the performance of *k*-means algorithm is not optimal [2]. Another approach in data clustering is multiobjective clustering approach which uses some criteria such as the compactness of clusters and the separation between clusters [3]. The clustering based on Gaussian mixture model (GMM) is considered widely in this area. The assumption underlying this approach is that the members in dataset are samples of a Gaussian distribution with specific mean and covariance matrix. In such assumption Gaussian mixture decomposition can be applied for solving clustering problem[1, 4]. The optimal labels for members are the final purpose of clustering. Such optimization aim can be dealt by Genetic Algorithm. Also there are several works in clustering problem which utilize different version of Genetic algorithm [2, 5-7].

There are several methods in the literature developed for optimal feature extraction. The majority of spike sorting methods transform the spikes to a feature space by extracting features that differentiate spikes. Such projection reveals clusters in the feature space. In such case the performance of the spike sorting is affected widely by the quality of extracted features. If suppose the neurons as the statistically independent sources, independent component analysis (ICA) can be used to extract the statistically independent waveforms from the measured signal. Such independent waveforms have the characteristics of each neuron [8]. Knowing the spike templates present in the recorded neurophysiological data, it is possible to classify spikes. The classification is performed by comparing each spike with all templates and selects the best matching template. Such methods suffer from the lack of information about templates and their correct numbers[9]. Principal component analysis (PCA) is a popular method to visualize high-dimensional data to the feature space. Usually some limited numbers of principal components contain most of the variance in the dataset, are used in the projection. In this manner PCA reduces the dimensionality of dataset [10]. Although PCA is widely used in feature extraction for spike sorting but for low signal-to- noise ratio conditions [11] and for the situation where there are small differences in spike templates[12], its performance is deteriorated.

For feature extraction, wavelet based methods have gained great attention [13-15] . In the multiresolution wavelet transform, time-frequency decomposition of the signal with optimal resolution in the time and the frequency domains is achieved which each coefficient contains some amount of the signal energy. Choosing some limited numbers of wavelet coefficients reduces the dimension of feature space. In the wavelet based methods mother wavelet selection is an important step as it alters the energy distribution between wavelet coefficients in different time and frequency scales. Generally the optimal choice for mother wavelet depends on dataset[12]. Approaches that utilize artificial neural network have been used in spike sorting[11]. The main weakness of methods based on neural networks is their necessity for the learning procedure and optimal structure that needs a-priori knowledge about the data which aren't usually accessible in the neural data processing [16].

In this paper an automatic method for spike sorting is presented which utilizes genetic algorithm (GA). The proposed method consists of two distinct phases. In the first phase a B-spline curve is fitted to each segmented spike. A method considering the morphological properties of spikes is proposed for spike segmentation. For such curve fitting the optimal parameters of B-spline curve are searched by genetic algorithm. The feature space is constructed based on the obtained parameters in a way that distinguishes between spikes. This constructed feature space is fed to clustering step. In the second phase another genetic algorithm is utilized for clustering which needs any predefined information about the correct number of clusters. The proposed approach has been tested using real neural data.

The rest of the paper has been structured as follows: in the first section B-spline curve fitting is defined briefly and its application for spike feature extraction is explained. The best parameters for fitting a B-spline curve to the spike waveforms are found using genetic algorithm. Section 2 presents the proposed genetic algorithm for clustering. Section 3 contains the implementation of the proposed methods for feature extraction and clustering where the performance of the proposed approach is evaluated.

II. MATERIAL AND METHODS

A. Bspline

A spline curve is a sequence of curve segments that are connected together to form a single continuous curve. In signal processing, B-splines are adopted to process noisy signals or for approximating complicated functions [17]. A B-spline curve of *d-th* degree is defined as the linear combination of a set of B-spline basis functions as (1).

$$f(t) = \sum_{i=0}^{m-d-1} p_i N_{i,d}(t, \boldsymbol{u})$$

(1)

Where p_i is *i-th* control point and $N_{i,d}(t,\boldsymbol{u})$, is the *i-th* B-spline basis function of degree d which is defined on a non-decreasing knot vector $\boldsymbol{u}=(u_0,u_1, \ldots, u_m)$ with length m as (2). Each knot span corresponds to one curve segment.

$$N_{j,0} = \begin{cases} 1 & u_i \leq t \leq u_{i+1} \\ 0 & otherwise \end{cases}$$

$$N_{i,d} = \frac{t - u_i}{u_{i+p} - u_i} N_{i,d-1}(t) + \frac{u_{i+d+1} - t}{u_{i+d+1} - u_{i+1}}$$

$$i = 0,1,\ldots,m-d-1$$

(2)

The set of linearly independent basis functions $N_{i,d}$, are piecewise polynomials with compact support and often defined in the interval [0 1]. The boundary values are clamped as $u_0=u_1=\ldots=u_d=0$, $u_{m-d}=u_{m-d+1}=\ldots=u_m=1$. The B-spline shape is affected by knot vector $\boldsymbol{u}$, control points $\boldsymbol{p}$ and degree d. The function $f(t)$ is d-r_j times differentiable at $\boldsymbol{u}_j$ and infinitely differentiable elsewhere. r_j is the number of multiplicities of $\boldsymbol{u}_j$. It is possible to approximate a curve passing through a set of observed data points based on B-spline interpolation. To fit such curve to the data points both control points and knot points must be determined based on observed data.

The number of B-spline basis function must satisfy the equality of *n=m-d-1*, where *m* is the knot vector size, *d* is the B-spline degree and *n* is the number of basis functions. Note that based on (1) each basis function needs one control point so *n* is also the number of control points.

An important property of B-spline curve is that it is contained in the convex hull of its control polyline. This means that there are only *d+1* basis functions engaged in

constructing a curve segment in each knot span and so the *p+1* corresponding control points have non-zero coefficients. The weighted average of such basis functions must lie in the convex hull specified by corresponding *p+1* control points in each knot span. This property makes polyline a useful tool for following the variation of a curve or waveform. It should be noted that lower degree of B-spline causes the B-spline curve to be closer to its control polyline. Changing the location of control points modifies the curve segments shape, so in curve fitting the precise location of control points is very important.

B. *Main idea of the application of B-spline in spike sorting*

The response of each neuron to an excitation is a sharp and stereotyped action potential or spike. The stereotyped spikes of a neuron are propagated along axon without change of shape over long distances. This means that spike shape can be used as an important tool for identifying different neurons. In this paper each spike is modeled by a B-spline curve which needs three parameters of B-spline include control points, knot vector and spline degree to be defined and then B-spline control points are used in the next step i.e. feature extraction. The parameters of B-spline are affected directly by spike shape. When spike waveform is modeled by a B-spline curve, the spike lies in polyline constructed by B-spline control points. Choosing the lower B-spline degree makes control points closer to data points. For determining knot vector some critical points in each spike waveform are specified as is described in the following section. Genetic algorithm is used for optimizing the best location of control points to reduce curve fitting error.

C. *Genetic algorithm for control point adjustment*

The Genetic Algorithm methods usually are used in optimization problems. They are less susceptible to getting stuck at local optima compared with gradient search methods.

To use a genetic algorithm, each solution of the problem should be represented as a *genome* or *chromosome*. The genetic algorithm creates a population of solutions and then applies genetic operators such as mutation and crossover to the selected chromosomes to generate new chromosomes called offspring. The quality of each solution or chromosome is evaluated by a fitness function. The algorithm evolves the solutions in order to find the best one.

Each control point is a pair of numbers as *(x,y)* where x and y indicate the location of control point in time and the amplitude of control point, respectively. The genetic algorithm is used to search the optimal locations of control points in fitting the B-spline curve to spike waveform.

The proposed genetic algorithm is implemented as follows:

1. Specify the number of control points: the number of control points(*n*) is restricted by the number of knot points(*m*) and B-spline degree(*d*) as *n=m-d-1*.

2. Create initial population for x and y values of control points individually. The x values are selected randomly in spike length interval and the corresponding y values are chosen in the vicinity of spike amplitude in the related x locations. This will accelerate the rate of convergence to the optimal control points. The population size is set by a predefined value.

3. For each *x* and *y* chromosome in the population, Compute the fitness value: knowing the control points, knot vector and B-spline degree, it is possible to construct a B-spline curve. How the curve is fitted to given data points is computed by coefficient of determination criterion. In statistics, the coefficient of determination indicates how well data points fit a line or curve. This criterion is denoted by R^2 and is defined as (3):

$$R^2 = 1 - {S_{res}}/{S_{tot}}$$
$$S_{res} = \sum_i (y_i - f_i)^2$$
$$S_{tot} = \sum_i (y_i - \overline{y})^2 \tag{3}$$

Where in (3) S_{res} is the sum of squares of residuals in curve fitting and S_{tot} is the total sum of squares which is proportional to the sample variance. Also y_i is the given data point of spike waveform and f_i is the point of fitted B-spline curve. Closer R^2 to one, exhibits greater fitness of chromosome.

4. Generate new offspring by genetic operator:

4-1. Crossover: from the population, two parents with closer fitness value to one are selected (p_1,p_2). This selection is carried out based on roulette wheel selection.

From these parents three offspring are generated by crossover as follows:

$$offspring1 = kp_1 + (1-k)p_2$$
$$offspring2 = p_1 + t(rand - 0.5)p_1$$
$$offspring3 = p_2 + h(rand - 0.5)p_2 \tag{4}$$

In (4), *t* and *h* are constant coefficients which decrease exponentially with time or iteration of parent generation. *k* is another constant coefficient which decreases linearly with time. The rand function in (4) generates a random number in the interval [0 1] and it causes the new generated control points move in all directions.

4-2 Mutation: for mutation one possible solution of control points is selected randomly. Small values are added or reduced to selected chromosome to generate a new offspring. The probability of mutation is decreased with time exponentially.

The new generated offspring give new control points that can be used for generating new B-spline curves. The

fitness of such curves are computed and finally two most fitted are selected. If the fitness of selected offspring is further than their parents, the parents are removed from the population list and the offspring is substituted.

The offspring generation can be continued for a predefined number of iterations or fetch up if the fitness values of new generated offspring are not better for some successive iteration.

D. Knot vector definition

In fitting a curve to the physiological data, increasing the complexity and variation levels in data will increase the degree of curve. In this paper action potential signals are detached to shorter segments in some critical points and in each segment with lower complexity, the curve fitting task is carried out. For finding critical points in action potential time series, their morphological shape is considered.

The action potential generation can be divided into five phases including the rising phase, the peak phase, the falling phase, the undershoot phase, and the refractory period which are managed by ionic gates. In rising phase the membrane potential becomes more positive and depolarization occurs. This phase is concluded in the action potential peak where membrane potential reaches the maximum. After peak location the falling phase starts which the membrane potential will be more negative. In this phase an undershoot event which referred as hyperpolarization is possible. In such phase the membrane potential is more negative than resting potential. Finally, the time during which a subsequent action potential is impossible or difficult to fire is called the refractory period, which may overlap with the other phases.

In action potential detection, extracted spikes are usually aligned with their dominant peak location so the peak is selected as one critical point. The location before dominant peak with maximum slope value and the location after dominant peak with maximum absolute slope value are selected as other critical points. Such points are highly depends on the action potential waveform which is important in the sorting step.

The probable valley location produced by hyperpolarization is another critical point. Note that single or array microelectrode placed in the extracellular space record the activity of several neighboring neurons. Closer the neuron to the recording site causes higher activity of such neuron to be recorded. Due to the activity of other neurons the low amplitude part of an intended action potential (the first and last segments of spike) are contaminated more severely by interferences so to reduce the effect of such interferences the first and last segments of each spike are discarded in curve fitting procedure. Discarded segments have duration equal to 0.1 times of action potential duration. Selected critical points and produced segments for a sample action potential are shown in Fig 1.

By segmentation the variability in each segment is reduced so the lower degree curve can be utilized for fitting each segment. The critical points are considered as knot vector for B-spline curve fitting.

It is possible that interferences caused by the activity of neighboring neurons destroy the hyperpolarization valley. In such case associated critical point will be chosen in the mid distance between critical point with maximum absolute slope and starting point of the last segment. The critical points, the first and last point of each spike form the knot vector for B-spline curve fitting.
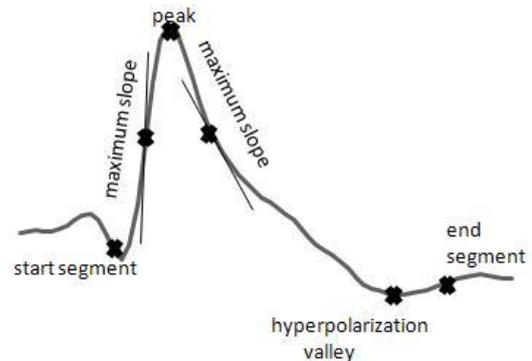

Fig. 1. Sample action potential and selected critical points. Critical points divide each spike to seven segments. These critical points are considered as knot vector for B-spline curve fitting

E. Feature selection

In the previous section each spike was fitted with a B-spline curve. The control points of the most fitted B-spline curve were considered as the extracted features for that spike. In this section, selection of features is addressed. Selection is done in a manner that discriminates between different waveforms. Suppose all spikes have been generated by one distinct neuron so all waveforms in dataset should have similar shapes, therefore *k-th* control point among all fitted curves should have approximately similar location and then distribution of such values follows a monomodal distribution. If there are more neurons therefore different waveforms will be appeared in the data set. In this case it is possible to find the control points that their distribution among all fitted curves is multimodal. These control points can distinguish between different spikes and are demanded in our task. For multimodality test, the Lilliefors modification of a Kolmogorov-Smirnov (KS) test for normality is applied[15]. Given *i-th* control point location among all fitted curves *(x)*, the KS compares the cumulative distribution function of the data *(F(x))* with that of a Gaussian distribution with the same mean and variance *(G(x))*. Deviation from normality is measured by(5):

$$max|F(x) - G(x)| \qquad (5)$$

The control point which its distribution among all fitted curves is more deviated from normal distribution is selected as the first optimal feature. As the standard

deviation exhibits the variability of a waveform so for each fitted curve the standard deviation of control points is computed as the second optimal feature. It should be noted that each control point defines by a pair of numbers which the first one specifies the time location of control point and second one specifies the amplitude of control point. We have chosen the amplitude of control point for above discussion.

F. Genetic Algorithm for clustering

In this paper a simple genetic algorithm method is introduced for clustering. This algorithm is explained as follows:

1.   Initial population Generation: as clustering aims to assign labels to data points in the feature space, in generating initial population a random label from existing labels is allocated to each data point. For this purpose k data points in feature space are selected randomly as the centers of clusters and finally each data point is assigned to nearest center. Note that the assumption is that neighbor points in feature space project the similar objects in real world. k or the number of clusters is chosen large enough because the similar clusters will be merged during genetic operations. Each member of population is called chromosome.

2.   Fitness computation: for each chromosome clustering must be evaluated in term of an objective measure. Such measure gives the fitness value of clustering. In this paper objective function is computed based on both spatial and similarity distances. For each given cluster in feature space the intra-cluster distances (ICD) which is the average Euclidean distance between each pairs of cluster members are computed. Smaller value for such measure indicates more compact cluster. Simultaneously the between-cluster distance (BCD) as the average Euclidean distance between specified cluster's center and other cluster centers is computed. Much more the value of this measure indicates more separated clusters. At the same time the intra-cluster correlation (ICC) between corresponding waveforms of data points in the feature space which belong to a distinct cluster is considered. This is similarity measure for each cluster. A small value of similarity shows the cluster is far from optimality. In brief the fitness value (F)for each class is computed as(6):

$$BCD(i) = \sum_{j=1}^{k} d^2(\mu_i - \mu_j)$$

$$ICD(i) = (^2/_{n(n-1)}) \sum_{t \in C_i} \sum_{h \in C_i} d^2(x_t, x_h)$$

$$ICC(i) = (^2/_{n(n-1)}) \sum_{t \in C_i} \sum_{h \in C_i} corr(x_t, x_h)$$

$$F(i) = ICD(i) / (BCD(i) \times \sqrt{ICC(i)})$$

$$(6)$$

In (6) n is the number of data points in feature space which belong to i-th cluster C_i, k is the number of clusters, the centroid of the i-th cluster represented by μ_i, and *corr* is the correlation function. Note that $ICC(i) \epsilon [0\ 1]$, so for better performance, the measured values of $BCD(i)$ and $ICD(i)$ are normalized in $[0\ 1]$. For the cases which the templates in dataset are very similar and only differ in small-scale structures, the value of ICC for different clusters may be very close. For increasing the sensitivity of F to small differences in templates, $\sqrt{ICC}$ is considered in (6).

The overall fitness (OF) for each chromosome (possible data point labels) is computed as the summation of fitness for all classes as (7):

$$OF = \sum_{i=1}^{k} F(i)$$

$$(7)$$

Smaller value for OF indicates more optimized clustering.

3.   Genetic operators:

Some chromosome with smaller value of OF are selected in initial population. Among the selected chromosomes the averaged fitness of each class is computed. The worst class is the cluster with largest averaged fitness value. Before generating new offspring, M members of population which have been labeled more by worst cluster labels are deleted. In genetic operation the deleted member are replaced by generated offspring as follows.

Crossover: The population members are sorted in a list based on their fitness values. The most qualified chromosome which has minimum OF value is selected as first parent. The second parent is selected with a probability from the sorted population. For this aim the span $[0\ 1]$ is divided to many sub-spans and each sub-span is assigned to one population member. Corresponding sub-span for n-th individual in the population list is $[1/2^{(n-2)}, 1/2^{(n-1)}]$. Note that before this the first chromosome in population is selected as first parent and the selection is continued from second location in the population list. Finally a random number is produced and the member related to the span of random number will be selected as the second parent. This allows population members with lower fitness to be engaged in crossover.

From the first parent, genes labeled as worst group are searched and replaced by the corresponding genes in the second parent. If any replaced gene of second parent is labeled by worst group label, the label of neighbor cluster of the worst class is substituted (see Fig 2.). The center of the neighbor cluster for worst cluster has nearest distance to the center of worst cluster. In this scheme the worst cluster is discarded and its members are allocated to other clusters so the generated offspring has lower number of clusters.

Parent 1	1	1	2	③	1	1	③	2	2	③	③	1
Offspring	1	1	2	1	1	1	2	2	2	①	2	1
Parent 2	1	2	2	1	1	2	2	2	2	3	2	1

Fig. 2. Proposed offspring generation. Suppose a three clusters case, the worst class is cluster number 3 and its neighbor is cluster number 1. The worst class labels in parent 1(indicated by circle) are replaced by corresponding labels in parent 2. If parent 2 contains worst class label, the neighbor label of that is substituted (indicates by arrow).The offspring contains no genes labeled by worst class label.

Mutation:

For mutation one member of population (expect two selected parents in crossover step) is selected randomly and a limited number of their genes labeled by worst group label are replaced by the label of worst cluster's neighbor. The mutation probability is decreased during offspring generation.

Before offspring generation the limited numbers of population members denoted by S which contain maximum number of genes specified by worst class label will be deleted. The deleted members will be replaced by S-1 offspring generated by crossover and one offspring generated by mutation.

III. RESULT AND DISCUSSION

In this section the performance of both proposed algorithms for feature selection and clustering are compared with traditional methods.

A. Performance of feature extraction algorithm

For evaluating the performance of proposed method for feature extraction, two data sets are used. The dataset 1 consists of 3 different template waveforms of 402 spikes. The dataset 2 consists of 2 template waveforms which are different in small-scale structures including 274 spikes. The template waveforms of these two data sets are depicted in Fig 3. The neural signal was recorded from the restricted cockroach by inserting a single microelectrode on cockroach body [18] where the sampling frequency was 30Kb/s and to reduce the induction of interferences faraday cage was utilized. The spikes are extracted from recorded neural data by simple amplitude thresholding method. Each spike in these two data sets is labeled by an expert.
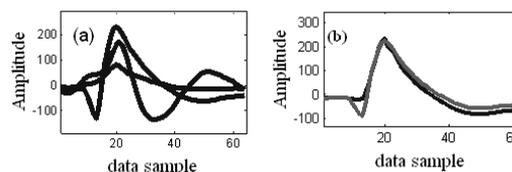

Fig. 3. (a) dataset 1 template waveforms. (b) dataset 2 template waveforms. data set1 consists of three different templates and data set 2 consists of two templates which are different in small-scale structures.

For comparison, two feature extraction methods based on multiresolution wavelet transform and sammon's mapping are used. In wavelet based method, multilevel one-dimensional wavelet analysis using a specific wavelet is performed. To select a few coefficients that best separate the different spike classes, a multimodal distribution test using modification of Kolmogorov-Smirnov (KS) test is carried out [15]. Another method is Sammon's mapping which is a non-linear mapping that maps a set of input waveforms on a feature space (usually 2D or 3D) trying to preserve the relative distance within the input space approximately. Sammon's projection tries to minimize a cost function referred as sammon's stress function in an iterative procedure[19]. In our case each data set is transformed to a two dimensional feature space then k-means clustering is performed. It should be noted that our focus in feature extraction is on finding optimal features which get more compact and well-separated clusters by preserving the distance between data set members in feature space as real world. The result of feature extraction for dataset 1 is depicted in Fig 4.
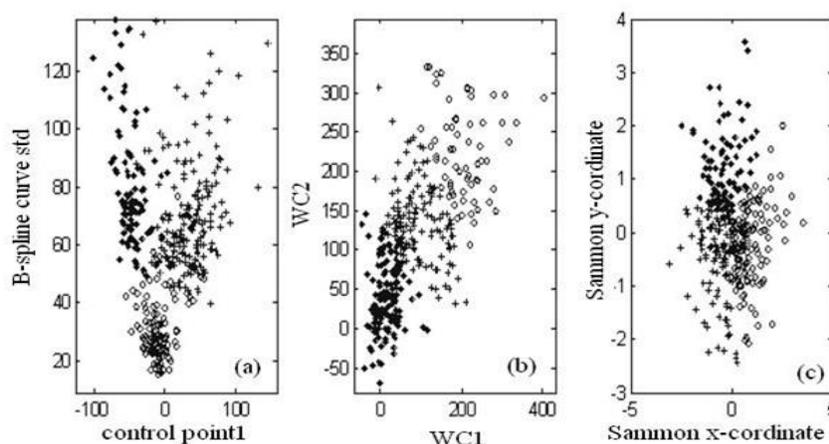

Fig. 4. Feature space generation by (left) proposed method, (middle) wavelet based method and (right) sammon's mapping. K-means clustering is performed in each case to test the ability of features in obtaining well-separated clusters.

Fig 4 shows the result of feature space construction and corresponding *k*-means clustering for dataset 1. From Fig 4 it is visually obvious that the proposed method has higher ability to obtain well-separated clusters than two other methods. To ensure that close objects in data set are projected in neighboring points in feature space, the classification error as the percentage of misclassified spikes is computed and reported in Table 1. We observed that wavelet based method is very susceptible to mother wavelet selection. For comparison purposes symlet7 mother wavelet is selected due to its higher similarity with our template waveforms. The multiresolution wavelet decomposition is performed in 5 levels. Table1 shows that the proposed method for feature extraction gives the lower classification error. As proposed feature extraction includes genetic algorithm and the operations in genetic algorithm occur with a probability so the proposed feature extraction is carried out 5 times and the reported classification error is an average value. Also the error standard deviation is reported. It should be noted that in Sammon's mapping the convergent to the optimal solution during its iteration is not always guaranteed and

hence in comparison with GA based method it suffers from stacking in local minima of sammon's stress function. This increases the classification error.

For dataset 2 feature selection is carried out by three mentioned methods. This data set has two templates which are different in small-scale structures. For such dataset simple feature extraction methods like amplitude and spike width and the methods which rely on global features of data are not such suitable because such features have lower discrimination capability in such case. The feature space of this dataset, applying three mentioned methods, are displayed in Fig 5. Again the constructed feature space by each method (spline curve feature extraction, wavelet and sommon's mapping) is fed to a *k*-means clustering and the percentage number of misclassified spikes is considered for performance evaluation. The result is demonstrated in Table1. Visually all three methods have obtained well-separated clusters but the optimality of these features by computing the percentage of misclassified spikes in Table1 shows that the selected features based on proposed method give lower percentage of error for data set 2.
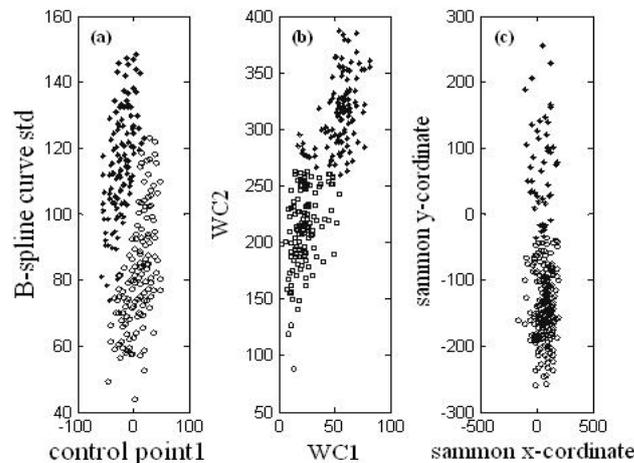

Fig. 5. Feature space generation by (a) proposed method, (b) wavelet based method and (c) sommon's mapping. K-means clustering is performed in each case to test the ability of features in obtaining well-separated clusters.

B. *Performance of genetic clustering algorithm*

In this section the performance of proposed genetic clustering algorithm is evaluated by dataset 1 and dataset 2. For this aim these two dataset are projected to a two dimensional feature space by our proposed method in feature extraction section. Finally the proposed genetic algorithm clustering method is applied to the feature space. The clustering error is defined by false positive (FP) and false negative (FN) terms. Allocation of a point to a wrong cluster is denoted by FP and not including a point to its cluster is denoted by FN. Error index for each cluster is computed by the summation of FP and FN values as (8). The overall clustering error is the sum of errors for all clusters.

$$error\ index = \sum_i \sqrt{FN^2 + FP^2}$$

(8)

Where *i* is the number of clusters. For comparison *k*-means and fuzzy C mean clustering are performed, as these methods are used extensively in clustering tasks. The result of clustering for dataset1 and dataset 2 are reported in Table 2. The results show that proposed genetic algorithm outperforms *k*-means and fuzzy c-mean clustering. Note that *k*-means and c-mean clustering performance deteriorates in the case of overlapped clusters because these methods operate only based on spatial distances between objects whereas in the proposed genetic algorithm both spatial distances in feature space and similarity distances in real world are considered. Large error reported in Table 2 may results due to the small number of spikes in our dataset. Note that for dataset 2 where spikes are more similar, the error is larger in comparison with dataset 1.

The *k*-means and fuzzy C-mean clustering algorithm are affected by initial cluster center selection. The

Genetic Algorithm clustering is also affected by initial population generation and some random operation in parent selection. As all methods are affected by random operations so each algorithm implemented 10 times and the average results are selected for comparison (see Table 2). An example of implementing of the proposed method and its comparison with *k*-means clustering is depicted in Fig 6. Note that for the cluster indicated by circle, the error of *k*-means clustering is dramatic. It should be mentioned that the number of clusters in *k*-means clustering and fuzzy C-mean is a predefined parameter but in the proposed genetic algorithm this is found automatically. This is done by merging clusters during genetic operations.

IV. CONCLUSION

In this study genetic algorithm is used for spike sorting. The proposed algorithm consists of two distinct phases. In the first phase a B-spline curve is fitted to each spike. In this case knot vector, B-spline degree and control points are important parameters. For finding the optimal B-spline control points that minimize the curve fitting error, genetic algorithm is used. In this algorithm a population of possible control points are generated and then by genetic operation the new offspring(control points) are created. Finally the optimized solution is found which minimizes curve fitting error. For projecting spikes to feature space the control point which its distribution among all fitted curves is more deviated from normal distribution and the standard deviation of control points as a measure of waveform variation are selected as features. In the second phase, for clustering the feature space, genetic algorithm is utilized to search the optimal labeling. In the proposed method spatial and similarity distances are used to evaluate the performance of labeling simultaneously. The results show that proposed method for feature extraction outperform wavelet based method and sammon's mapping method in the point of lower misclassified spikes which is achieved by obtaining well-separated clusters. Also it is shown that the proposed genetic algorithm for clustering outperforms *k*-means clustering and fuzzy C-mean clustering due to lower false negative and false positive errors. Although the number of clusters in *k*-means or fuzzy C-mean algorithms must be specified for the classifier as a predefined parameter, the proposed method needs no predefined parameters.
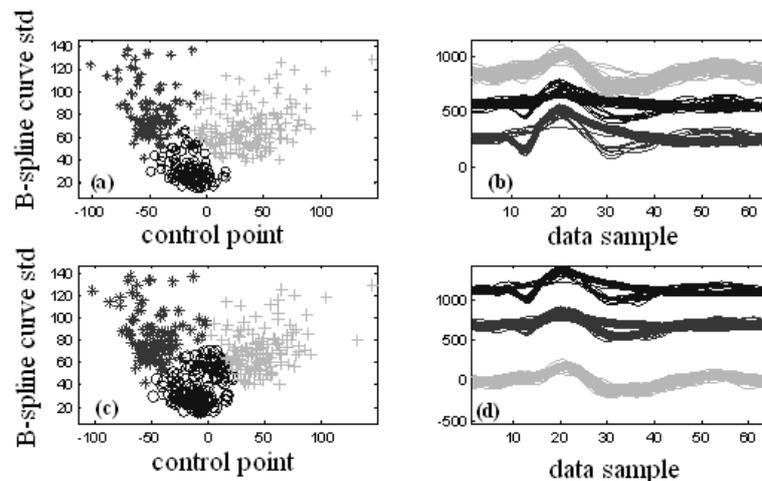

Fig. 6. Implementation of (a) proposed genetic algorithm and (c) k-means algorithm for clustering. The clustered spikes are depicted in (b) and (d) respectively.

Table 1. Percentage of misclassification error in k-means clustering

Method/data set	dataset 1	dataset 2
Proposed method	2.5 (±0.4)	6(±0.8)
Wavelet based	4.2	8.1
Sommon method	3.1(±0.3)	6.2(±0.7)

Table 2. Comparison of performance for different clustering methods

Algorithm/dataset	dataset 1				dataset 2		
	FN/FP(%)			Error index(%)	FN/FP(%)		Error index(%)
Cluster #	C1	C2	C3		C1	C2	
Kmeans	6/3	5.1/1.45	0.2/3.6	4.26	16.2/10.1	12.2/5.3	16.19
Genetic Algorithm	3.5/3	2.4/1.5	1.24/3.46	3.34	9/5.5	9.5/4.05	10.43
Fuzzy C-mean	5.8/3	5.2/1.5	0.4/4.75	4.067	15.8/10.2	11.7/5	15.76

REFERENCES

[1] S. Shoham, M. R. Fellows, and R. A. Normann, "Robust, automatic spike sorting using mixtures of multivariate t-distributions," *J Neurosci Methods* vol. 127, pp. 111-122, 2003.

[2] L. E. Agustın-Blas, S. Salcedo-Sanz, S. Jimenez-Fernandez, L. Carro-Calvo, J. D. Ser, and J. A. Portilla-Figuera, "A new grouping genetic algorithm for clustering problems," *Expert Systems with Applications,* vol. 39, pp. 9695–9703, 2012.

[3] J. Handl and J. Knowles, "An evolutionary approach to multiobjective clustering," *IEEE Trans Evol Comput,* vol. 11, pp. 56-57, 2007.

[4] A. Jain, R. Duin, and J. Mao, " Statistical pattern recognition : a review," *IEEE Trans Pattern Anal Mach Intell* vol. 22, pp. 4-37, 2000.

[5] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay, "Multiobjective Genetic Algorithm-Based Fuzzy Clustering of Categorical Attributes," *IEEE Trans Evol Comput,* vol. 13, pp. 991-1005, 2009.

[6] E. R. Hruschka and N. F. Ebecken, "A genetic algorithm for cluster analysis," *Intelligent Data Analysis,* vol. 7, pp. 15-25, 2003.

[7] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique.," *Pattern Recogn,* vol. 33, pp. 1455-1465, 2000.

[8] A. M. Mamlouk, H. Sharp, K. M. L. Menne, U. G. Hofmann, and T. Martinetz, " Unsupervised spike sorting with ICA and its evaluation using GENESIS simulations," *Neurocomputing* vol. 65–66, pp. 275–282, 2005.

[9] T. I. Aksenova, O. K. Chibirova, O. A. Dryga, I. V. Tetko, A. L. Benabid, and A. E. Villa, "An unsupervised automatic method for sorting neuronal spike waveforms in awake and freely moving animals," *Methods* vol. 30, pp. 178–187, 2003.

[10] D. A. Adamos, E. K. Kosmidis, and G. Theophilidis, " Performance evaluation of PCA-based spike sorting algorithms," *Comput Methods Programs Biomed,* vol. 91, pp. 232-244, 2008.

[11] K. H. Kim and S. J. Kim, "Neural spike sorting under nearly 0 dB signalto- noise ratio using nonlinear energy operator and artificial neural network classifier," *IEEE Trans. Biomed Eng* vol. 47, pp. 1406– 1411, 2000.

[12] A. Pavlov, V. A. Makarov, I. Makarova, and F. Panetsos, " Sorting of neural spikes: When wavelet based methods outperform principal component analysis," *Natural Computing,* vol. 6, 2007.

[13] E. Hulata, R. Segev, and E. Ben-Jacob, " A method for spike sorting and detection based on wavelet packets and Shannon's mutual information," *J Neurosci Methods,* vol. 117, pp. 1-12, 2002.

[14] J. C. Letelier and P. P. Weber, "Spike sorting based on discrete wavelet transform coefficients," *J Neurosci Methods,* vol. 101, pp. 93–106, 2000.

[15] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering," *Neural Computation,* vol. 16, pp. 1661–1687, 2004.

[16] P. M. Horton, A. U. Nicol, K. M. Kendrick, and J. F. Feng, "Spike sorting based upon machine learning algorithms (SOMA)," *J Neurosci Methods,* vol. 160, pp. 52-68, 2007.

[17] M. Unser, A. Aldroubi, and M. Eden, " B-Spline Signal Processing. Ii. Efficiency Design and Applications," *IEEE Transactions on Signal Processing,* vol. 41, pp. 834-848, 1993.

[18] S. Farashi, M. D. Abolhassani, Y. Salimpour, and J. Alirezaie, "Combination of PCA and Undecimated wavelet transform for neural data processing," presented at the Proc of the 32 th Annual. Intern. Conf. of the IEEE Eng. in Med. & Biol. Soc., 2010.

[19] J. W. Sammon, "A Nonlinear Mapping for Data Structure Analysis," *IEEE Transactions on Computers,* vol. 18, pp. 401-409, 1969.

Authors' Profiles

Sajjad Farashi was born in 1983 in Nahavand, Iran. He received his B.S. degree in Electronic.Eng. and the M.S.in biomedical.Eng. from the Tehran University of Medical Sciences, Tehran, Iran, in 2007 and 2010, respectively, and is now preparing his Ph.D. thesis at the Shahid Beheshti University of Medical Sciences, Faculty of Medicine, Tehran, Iran. His main research interests lie in the fields of bioelectromagnetism and biomedical signal processing, especially neural signal processing.

Mohammad Mikaili is currently an Assistant Professor at the Faculty of Electrical Engineering of the Shahed University, Tehran, Iran. His research areas of interest are biomedical engineering and biomedical image and signal processing.