2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

# Robust Signal Identification for Dynamic Pattern Classification

Rui Zhao
ECSE Dept., RPI
Troy, NY, USA
zhaor@rpi.edu

Gerwin Schalk
Wadsworth Center, NYS Dept. of Health
Albany, NY, USA
schalk@neurotechcenter.org

Qiang Ji
ECSE Dept., RPI
Troy, NY, USA
jiq@rpi.edu

***Abstract*—This paper addresses the problem of identifying signals of interest from discrete-time sequences contaminated by erroneous segments, which we define as the part of time series whose dynamic patterns are inconsistent with that of the signals. Assuming the signals of interest consist of consecutive samples with arbitrary starting point, duration and following a stationary dynamic pattern, we propose a robust algorithm combining Random Sample Consensus (RANSAC) and Hidden Markov Model (HMM) to automatically identify the start and end of signals of interest from time series. To evaluate the identification quality, we perform a classification task, where the identified signals are used to train a classifier. A majority vote strategy is adopted to handle error contaminated testing sequences. Compared with manual selection approach and other unsupervised learning methods, the proposed method shows improvement in classification accuracy on both synthetic and real Electrocorticographic (ECoG) data.**

## I. Introduction

Time series data provide important information for the analysis of system dynamics. However, time series data are often contaminated with irrelevant part introduced during experiment process. Therefore the exact location and duration of the signals of interest in time series are often unknown. For instance, in brain computer interface application, brain signals in response to exterior stimulus are recorded. However, both onset and elapsed time of the response are difficult to determine exactly due to the interference of other neural activities which may be happening before and after the ones of interest. In order to analyze the underlying dynamic pattern, it is crucial to identify the signals of interest first, which we consider as signal identification problem.

The signal identification problem is closely related to the change-point detection or time series segmentation problem whose goal is to segment time series into disjoint statistically consistent parts. Time series segmentation has been studied in a variety of disciplines including speech signal segmentation [1], action recognition from videos [2], stock data mining [3] and climate change detection [4]. A typical approach to obtain segmentation is to minimize a specific cost function which gives optimal start and end points for each segment (See [5] for a review). More recently, Liu *et al.* [6] proposed to detect change-point based on a non-parametric divergence estimation between time series segments. Chen and Zhang [7] constructed a similarity graph among time series, based on which the change-point is determined.

Another approach is based on explicit modeling on the dynamics of time series. For instances, Takeuchi and Yamanishi [8] detected changing point and outliers in network security analysis using autoregressive model. Citi *et al.* [9] modeled electrocardiogram signals as a point process in order to detect abnormal heartbeats. Oh *et al.* [10] proposed a switching linear dynamic system for segmenting bee dance data. Lee and Kim [11] modeled time series using two HMMs to identify existence and uniqueness of the pattern among different classes.

The third approach treats signal identification as an unsupervised learning problem. Zhou *et al.* [2] proposed aligned cluster analysis based on kernel k-means for motion sequence segmentation. Tierney *et al.* [12] extended subspace clustering methods to segment sequential data with temporal smoothness constraints. However, the dynamic pattern is not modeled in these approaches.

Existing works often assume the availability of ground truth segmentation, based on which a statistical model or template can be developed to characterize the dynamic patterns. In contrast, we are faced with the case where the ground truth segmentation is not available. In addition, we do not assume any prior knowledge on the dynamic pattern of signals of interest except for consistency of signals belonging to the same class and the temporal continuity of the signals. Our goal is to identify signals of interest from discrete-time sequence containing irrelevant or erroneous segments. Our contribution lies in the following aspects:

- We propose a framework that can robustly identify the signals of interest from time series, which helps analysis of the corresponding dynamic pattern.
- A novel identification method called likelihood sum is proposed based on HMM learned from data. We analyze the identification quality with respect to the computational cost based on RANSAC framework.
- We support the proposed framework with empirical evidence obtained on both synthetic and human brain ECoG data by performing a classification task.

The rest of this paper is organized as follows. In section 2, we define the problem and the used assumptions. We describe our approach in details in section 3. The experiments on synthetic data and ECoG data are discussed in section 4 and

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

5 respectively, followed by conclusion and future work.

## II. Problem Statement

Consider discrete-time sequence $\mathbf{X} = \{X_t\},\ t = 0, ..., L-1$, where length $L > 0$ containing signals of interest $\{X_a, ..., X_{a+N-1}\}$ with $a$ and $N$ be the onset time and duration of signals respectively. The remaining samples are called erroneous segments. We assume within each sequence, the signals of interest, which possess some stationary dynamic pattern, consist of consecutive samples in discrete-time with a duration $N \geq 0.5L$.

Define subsequence be a subset of $\mathbf{X}$ which contains consecutive samples in discrete-time. A subsequence is called signal subsequence if it only contains samples belonging to signals of interest and otherwise called erroneous subsequence. Fig. 1 is an illustration of different parts in a sequence. Given a set of sequences $\{\mathbf{X}\}_K$, we are trying to identify signals of interest for each sequence, i.e. find starting point and length of the signals of interest.

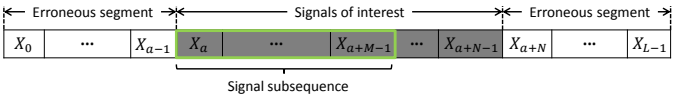


Fig. 1. Illustration of different parts of discrete-time sequence of length $L$. Each small rectangle represents a discrete-time sample. The length $N$ signals of interest $\{X_a, ..., X_{a+N-1}\}$ is shaded. The two unshaded segments $\{X_0, ..., X_{a-1}\}$ and $\{X_{a+N}, ..., X_{L-1}\}$ are erroneous segments. The subsequence in green bounding box is an example of length $M$ signal subsequence.

## III. Methods

We provide a brief overview of RANSAC method and HMM, where our method is build on. Then we describe in details about the proposed signal identification method.

### A. Random Sample Consensus (RANSAC)

RANSAC is a robust algorithm for parameters estimation of a mathematical model from a set of data points which contains a minority portion of outliers [13]. In an iterative process, RANSAC randomly selects a sufficiently large subset of data points to estimate the model parameters. The quality of estimation is evaluated using all the remaining data points by some quantitative metric, where the ones that agree with the learned model form a consensus set. The consensus set is expected to be of larger size if the learning subset contains more inliers. By repeating this process multiple times, we may select a subset containing only inliers, which can produce the largest consensus set.

Given the total number of data points and proportion of outliers, we can compute the number of random selections required such that at least one selection does not contain any outliers up to a pre-specified probability value. Finally, the parameters are re-estimated by the identified inliers, which include all the data points in the largest consensus set as well as the corresponding data points used to train the model. In this work, we tailor RANSAC method for time series data and use HMM as the underlying model for the data. The process of generating consensus set naturally provides us a way of identifying signals of interest, which are considered as inliers in our case.

### B. Hidden Markov Model (HMM)

HMM is a popular generative model that characterizes the dynamic pattern of time series [14]. The sequential observations are modeled by a series of random variables, each of which is associated with a latent discrete state variable. The dynamic behavior of the sequence is characterized by the status change between neighboring state variables. The joint distribution of hidden state $\mathbf{X} = \{X_1, ...X_T\}$ and observation $\mathbf{Y} = \{Y_1, ...Y_T\}$ can be factorized as follows.

$$P(\mathbf{X}, \mathbf{Y}) = P(X_1) \prod_{t=2}^{T} P(X_t|X_{t-1}) \prod_{t=1}^{T} P(Y_t|X_t) \quad (1)$$

Therefore, the model can be parametrized by three conditional probability distributions, namely prior distribution $P(X_1)$ for initial state variable, transition distribution $P(X_t|X_{t-1})$ for neighboring pair of state variables and emission distribution $P(Y_t|X_t)$ for each observation and its associated state variable. Given time series, we can estimate the parameters of the three distributions using EM algorithm. For inference, we compute likelihood $P(\mathbf{Y})$ using forward-backward algorithm [15]. In this work, we use multivariate Gaussian distribution for emission distribution. The number of hidden state is selected through a cross-validation process on training data.

### C. Our method

Our method essentially consists of three components. First, iteratively select random subsequences of time series data following the framework of RANSAC method. At each iteration, HMM is learned using selected data. Second, evaluate the learned HMMs using all the subsequences and choose the best one, which is then used to identify the signals of interest. Third, evaluate the quality of identified signals from different classes by performing a classification task.

#### *Data selection and model generation*

Inspired by RANSAC method, we randomly select one subsequence from each complete sequence. If all the selected subsequences are signal subsequences, we have a better representation than using the complete sequences. By repeating this process multiple times, we increase the probability of selecting only signal subsequences. To evaluate the selected subsequences, we fit a HMM to represent the overall dynamic pattern at each selection. In the following analysis, we show that given $p$, a pre-defined probability value, we can determine $S$, the number of selections needs to be performed such that at least one selection contains only signal subsequences.

Given a sequence of length $L$ with signals of interest length $N$ $(0.5L \leq N \leq L)$, we have $L-M+1$ subsequences, where $M$ $(M \leq N)$ is the length of subsequence. Similarly, there are $N-M+1$ signal subsequences. Let $\epsilon$ be the probability of a subsequence being an erroneous subsequence. We can

compute $\epsilon = \frac{\eta L}{L-M+1}$, where $\eta = 1 - N/L$ is the proportion of outliers. Suppose we have $K$ training sequences each with length $L$. From each one of the $K$ sequences, we randomly select a length $M$ subsequence, yielding $K$ subsequences. The probability of all $K$ subsequences being signal subsequences is $(1-\epsilon)^K$. Within $S$ sets of selections, the probability that at least one out of $S$ sets has no erroneous subsequences is $p = 1 - (1-(1-\epsilon)^K)^S$. Therefore

$$S = \frac{\ln(1-p)}{\ln(1-(1-\epsilon)^K)} \tag{2}$$

Eq. (2) determines the number of selections $S$ needed to achieve identification accuracy $p$ given $\epsilon$. For example, Fig. 4 shows value change of $S$ with different $p$ under different $\epsilon$, which depends on $\eta$ and $M$. The value of $M$ is often application dependent. While the analysis holds true for $M \leq N$, we will show in experiment part how we choose $M$. Note that the training sequences do not need to have the same length in general. We consider the same length case here and the varied length case can be extended by similar arguments. After selecting subsequences at each iteration, we learn a HMM using EM algorithm, which has complexity $O(MQ^2)$ and $Q$ is number of hidden states. For the details of learning HMM, readers are referred to [15]. The same process is applied to each class of time series.

*Model selection and signal identification*

Model selection is to identify the HMM that is most likely to be generated by a set $s^*$ consists of only signal subsequences. We propose a method using $\log P(\mathbf{Y})$, namely log-likelihood computed by HMM for model selection. The signals of interest are then identified by the selected best model. We work with log-likelihood values to avoid numerical underflow.

The intuition of this method is that the log-likelihood of a signal subsequence computed by HMM should be larger if the model is learned from mostly signal subsequences due to the consistency assumption. Since the signals of interest are the majority portion of a sequence, the summation of log-likelihood of all the subsequences should increase as the quality of learned model improves. To be specific, at each iteration we keep record of the log-likelihood of each subsequence computed using learned HMM, yielding $K(L-M+1)$ log-likelihood values. The computational cost is $O(LMQ^2)$. Let $l_i^{(s)}$ be the log-likelihood computed by $s^{\text{th}}$ HMM for $i^{\text{th}}$ subsequence. We calculate $\mathcal{L}_s$, the sum of log-likelihood values of all subsequences as

$$\mathcal{L}_s = \sum_{i=1}^{K(L-M+1)} l_i^{(s)}, \ s = 1, ..., S \tag{3}$$

The best model among $S$ selections is determined as

$$s^* = \arg\max_s \mathcal{L}_s \tag{4}$$

To identify the signals of interest, we set a threshold $h = \bar{\mathcal{L}}_{s^*}$ as the average log-likelihood over $l_i^{(s^*)}$. The $i^{\text{th}}$ subsequence is considered as a signal subsequence if its log-likelihood is higher than $h$. The signals of interest within each sequence is the union of its signal subsequences. The overall process of signal identification is summarized by Algorithm 1. The overall computational cost is $O(SLMQ^2)$.

**Algorithm 1** Robust signal identification

**Input:** $K$: number of sequences, $M$: length of subsequence, $S$: number of iterations
**Output:** Signals of interest

**for** $s = 1$ to $S$ **do**
  $\mathcal{D} \leftarrow$ randomly select $K$ length $M$ subsequences, one from each of the $K$ sequences
  $\theta_s \leftarrow$ learn HMM parameters with $\mathcal{D}$
  $l_i^{(s)} \leftarrow$ compute subsequence $i$ log-likelihood using $\theta_s$
  $\mathcal{L}_s \leftarrow$ compute log-likelihood sum using Eq.(3)
**end for**
$s^* \leftarrow \arg\max_s \mathcal{L}_s$
$h \leftarrow \mathcal{L}_{s^*}$/ total number of subsequences
**return** union of subsequences $i$ with $l_i^{(s^*)} \geq h$

*Signal evaluation*

We evaluate the quality of identified signals of interest by performing a classification task. After robust identification process, we relearn a HMM using identified signals of interest for each class. Let $\theta_k, \ k = 1, ..., C$ be the parameters of $k^{\text{th}}$ HMM, where $C$ is the number of classes. Since the testing sequence may also be corrupted by erroneous segments. We propose to perform classification based on subsequences. We divide the complete testing sequence into subsequences in the same way as training, i.e. a length $L$ sequence is divided into $L-M+1$ length $M$ subsequences. Then we decide the label of each subsequence using some classifier. In our experiment, we tried two different classification methods. The first one directly uses log-likelihood computed by learned HMM from different classes. The second one uses multi-class linear SVM. Nevertheless, the framework is completely general and any off-the-shelf classifier can be applied. After we classify all the subsequences, we decide the label of the complete sequence as the majority class of subsequences. The algorithm of complete testing process is described in Algorithm 2, where $classifier()$ represents some classification function.

**Algorithm 2** Time series classification using subsequences

**Input:** $\mathbf{X}$: testing sequence, $M$: length of subsequence
**Output:** Class label

$n \leftarrow 0$, $a[1...C] \leftarrow 0$
**while** $n + M - 1 <$ length of $\mathbf{X}$ **do**
  $l \leftarrow classifier(\mathbf{X}(n : n + M - 1))$
  $a[l] \leftarrow a[l] + 1$
  $n \leftarrow n + 1$
**end while**
**return** *label* $\leftarrow \text{argmax}_i \, a[i]$

## IV. Experiments on Synthetic Data

### A. Data generation

In order to evaluate the effectiveness of proposed robust segmentation algorithm, we perform two experiments on synthetic

data, which consist of three classes of univariate time series generated by pre-defined functions. Gaussian random noise is added to each data point in each sequence. 1) Saw-tooth waveform with negative slope: $c_1(t) = 1 - (0.5t - \lfloor 0.5t \rfloor)$; 2) Saw-tooth waveform with positive slope: $c_2(t) = 0.5t - \lfloor 0.5t \rfloor$; 3) Sine waveform: $c_3(t) = 0.5\sin(\pi t) + 0.5$, where $\lfloor x \rfloor$ is the largest integer that is smaller than or equal to $x$. The discrete-time series are obtained by sampling the continuous-time waveforms with period $1/20$, i.e. $d_i(n) = c_i(n/20),\ n = 0, ..., 99,\ i = 1, 2, 3$. Then each sequence has length 100 and a complete pattern lasts $20\%$ of total length. Finally, we add Gaussian noise with mean 0 and $\sigma = 0.25$ to the generated sequence.

### *B. Processes and results*

The synthetic dataset is partitioned into training and testing set. Training set contains 10 sequences for each class. Testing set contains 30 sequences in total.

The first experiment involves training set only and the hidden state number of HMM is fixed to be 3. For each sequence, we introduce erroneous segments by substituting the beginning and ending portion of each sequence by Gaussian noise with $\mu = 0$ and $\sigma = 0.5$. The proportions of substitution at the beginning and end are chosen randomly with total proportion equals to $30\%$ of the sequence length. This generation mimics the situation in real data where the onset of signals of interest can be arbitrary. As a comparison, we train HMM using either the entire sequence or only signal portion for each class. Finally, we train HMM using signals of interest identified by the proposed robust identification method. For evaluation, we first compute log-likelihood of all the overlapping subsequences from training set using HMM as shown in Table I.

TABLE I
AVERAGE LOG-LIKELIHOOD PER SUBSEQUENCE COMPUTED BY HMM LEARNED FROM DIFFERENT TYPES OF SUBSEQUENCES

| Class | Entire | Identified by Alg. 1 | Signals |
|---|---|---|---|
| 1 | -8.02 | **-6.23** | -6.27 |
| 2 | -5.07 | -3.44 | **-2.73** |
| 3 | -1.87 | **-1.08** | -1.77 |

We see that using erroneous corrupted sequences yield the lowest log-likelihood. The proposed method produces comparable log-likelihood as the ideal case where only identified signals of interest is used to train HMM. This provides an empirical evidence that a higher log-likelihood sum is effective in identifying signals of interest. We point out that a threshold on log-likelihood value can be applied to filter out extremely low values before computing the sum. However, the choice of threshold is non-trivial. Here we relies on the majority assumption of signal part to prevent extremely low log-likelihood produced by poor model. For the third class, the robust method produces higher log-likelihood than the ideal case. This is due to the erroneous segment may mimic the signal pattern since the log-likelihood computed by the entire sequence is comparable with the one computed by signals only.

We then directly evaluate the quality of identified inliers by comparing against the ground truth signal subsequences. This is considered as a binary classification on individual sample of a sequence. As shown in Table II, our method improves precision and F1-score by 11.6% and 17.2% compared to the case where treating all the samples as inliers.

TABLE II
IDENTIFICATION QUALITY USING ALGORITHM (1)

| Class | 1 | 2 | 3 | Average | Entire |
|---|---|---|---|---|---|
| Precision (%) | 81.8 | 82.5 | 80.5 | 81.6 | 70.0 |
| F1-score (%) | 89.9 | 89.6 | 89.2 | 89.6 | 82.4 |

The second experiment is classification which involves both training and testing sets and the hidden state number of HMM is chosen through five-fold cross validation on training set. We introduce erroneous segments in a similar way as the first experiment with three different variations as follows. Case 1: Only training sequences contain erroneous segments. Case 2: Only testing sequences contain erroneous segments. Case 3: Both training and testing sequences contain erroneous segments. We decide class label as the one whose corresponding HMM produces the largest log-likelihood. For training part, we compare the cases with and without using Algorithm 1. For testing part, we compare the cases with and without using Algorithm 2. The baseline approach is using the entire sequence in both training and testing. The classification accuracy of all different cases are summarized in Table III.

TABLE III
CLASSIFICATION ACCURACY UNDER DIFFERENT CONDITIONS WITH DIFFERENT METHODS.

| Case | Training / Testing | | | |
|---|---|---|---|---|
| | All/All | Alg.1/All | All/Alg.2 | Alg.1/Alg.2 |
| 1 | 80.0 | **96.7** | 80.0 | **96.7** |
| 2 | 76.7 | 60.0 | **90.0** | 80.0 |
| 3 | 66.7 | 63.3 | 83.3 | **90.0** |

From the results in case 1 and 2, we see that erroneous segments dampen the performance of classification. In case 1, the deployment of Algorithm 1 produces higher classification accuracy, which indicates the model is more likely to be learned from signal subsequences. In case 2, the deployment of Algorithm 2 improves the classification accuracy under the same HMM learned from training data. This means a majority vote strategy on classification is effective to alleviate the contamination of testing sequences. In case 3, using a combination of Algorithm 1 for training or Algorithm 2 for testing yields significant improvement on classification accuracy. Consider case 3 is the most realistic situation in practice, therefore it is critical to use both algorithms together.

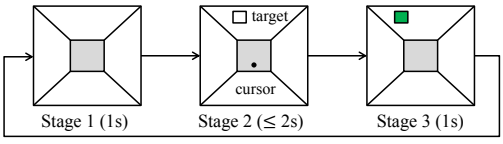


Fig. 2. Timeline of a trial, where the goal is to hit the target using cursor controlled by a joystick.

## V. Experiments on Electrocorticographic Data

### A. Data collection and pre-processing

We evaluate the performance of proposed robust signal identification algorithm on ECoG data, which has higher spatial and temporal resolution comparing to other brain signal modalities such as EEG and fMRI [16]. The signals were recorded from patients with electrodes placed subdurally on the surface of brain for solely clinical purpose of identifying epilepsy seizure foci prior to surgical resection. The subjects had normal cognitive capability and was given informed consent. During the motion control experiment, the subject held a joystick to move a cursor appearing on the screen to hit a virtual target. Each trial consists of three stages with total lasting time 3-4 seconds. Multiple trials were recorded in a consecutive manner. Fig. 2 shows the time line of a trial.

The target can appear in eight different locations, which leads to eight different directions of hand movement. We try to differentiate the dynamic pattern of signals in response to different hand movements, which is cast as a classification problem. Based on the design of experiment, the stimulus onset time, namely the beginning of stage 2 can be located. However, the accurate onset and duration of motor cortex activities, which are of primary interest, are difficult to obtained. The lack of such ground truth information makes the pattern recognition problem highly challenging.

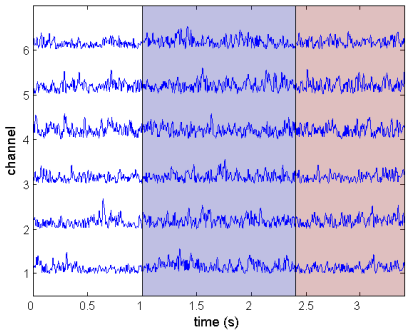


Fig. 3. Features from multiple channels of a trial of subject A. The unshaded, blue and red regions correspond to stage 1, 2 and 3 respectively. Target appears at the start of stage 2, whose duration varies across different trials and subjects. (Best view in color)

The ECoG signal was originally sampled at 1200Hz for each channel. For signal processing, we adopt a similar procedure used in [17]. We first exclude channels with significant line noise and then apply common average filter to all channels. We then apply notch filter to further reduce line noise and their harmonics. For feature extraction, we apply spectrum filter followed by Hilbert transform to extract amplitude envelope of representative frequency band. In this experiment, we use $\gamma$ band of range 70-170 Hz, which has been demonstrated with significant correlation to motor activity [18]. Finally, we down-sample the signal to 400Hz. Fig. 3 shows processed signal from multiple channels over one trial from one subject. Data from four subjects are used for further analysis. For each subject, we identify the electrodes covering motor cortex area resulting 4-6 channels of signals per subject.

### B. Processes and results

For classification experiment, for each subject, we select $K = 10$ trials from each direction of movement as training set and additional 10 sequences for testing, in total of 80 testing sequences. Each sequence has $L = 800$ sample, i.e. 2 seconds, which is the maximum allowed time for completing the motion portion of a trial. The chance level of classification is $12.5\%$. Due to the sophisticated inter-subject variation, we restrict the classification to be performed within each subject and the results are reported for each individual subject.

The average time to hit the target correctly starting from stimulus onset over 80 selected trials is 1.41s. Based on this average operation time, a rough estimate of erroneous portion can be as much as $1 - (1.41/2) \approx 0.3$. We choose our subsequence length as $M = 400$, which corresponds to duration 1 second. Based on Eq.(2), we compute the number of iterations required given different identification accuracies in log scale as shown in Fig. 4. In the following experiments, we assume the outlier proportion be $\eta = 20\%$ (green curve in Fig. 4). Choose the identification accuracy be $95\%$, we need $S = 463$ iterations.

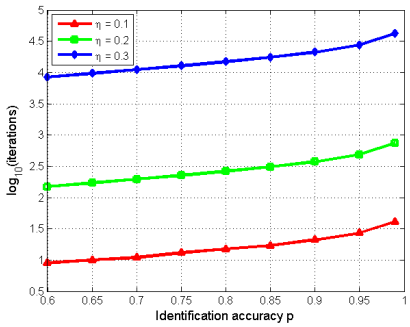


Fig. 4. Number of iterations needed to attain designated identification accuracy under different values of outlier proportion $\eta$. (Best view in color)

We compare three different methods against proposed robust identification method including aligned cluster analysis (ACA) [2], spectral clustering (SC) and manual selection. For ACA, we use the implementation provided by the original authors, where an entire sequence is segmented into multiple subsequences. We choose the longest subsequence as signals of interest. For SC, we implemented the algorithm described in [19], where we cluster each length $M$ subsequence into two clusters. The scaling parameter $\sigma$ is set to the average L2-norm of subsequences for computing affinity matrix. The signals of interest are identified as the union of the subsequences of the larger cluster. For manual selection, we simply use the stimulus onset time as the start of signals of interest.

The signals of interest identified by each method is then used to train a multi-class linear SVM, where we use libSVM [20]. For fair comparison, we use the same type of classifier for different methods so that the classification results only depend on the quality of identified subsequences. For each method, the first $M$ time stamps of identified signals of interest are used to train SVM, yielding the feature dimension be $lM$, where

number of channels $l = 6$ for subject A,B,C and $l = 4$ for subject D. To decide the hidden state number of HMM used in our method, we perform five-fold cross validation on training set and choose the number with highest average accuracy. The classification accuracy on testing set is summarized in Table IV. We also compute the 95% confidence interval for the average accuracy following the method of [21].

TABLE IV
CLASSIFICATION ACCURACY AND 95% CONFIDENCE INTERVAL (CI95) USING LINEAR SVM TRAINED ON SUBSEQUENCES IDENTIFIED BY DIFFERENT ALGORITHMS

| Subject | A | B | C | D | Average (CI95) |
|---|---|---|---|---|---|
| Manual | 32.5 | 65.4 | 74.5 | 42.5 | 53.7 (42.9,64.2) |
| ACA | 63.0 | 58.1 | 58.8 | 41.8 | 55.4 (44.5,65.8) |
| SC | **63.8** | 60.6 | 80.4 | 38.8 | 60.9 (49.9,70.9) |
| Ours | **63.8** | **67.3** | **100** | **50.0** | **70.3** (59.5,79.2) |

From the above results, we have the following observations. First of all, comparing results across different subjects, we observe significant variation, which suggests the variation of brain signals across different subjects in response to the similar motor task. Temporal alignment purely based on data collection protocol may not be sufficient. This variation is also partially due to inaccurate mapping between the placement of electrodes and brain function area. Nevertheless, the average accuracy of different subjects are all better than random guess which only has $12.5\%$ expected accuracy. These results show promising in identifying motor signal pattern from ECoG data despite the uncertainty of spatio-temporal location of the signals. Second, comparing results across different methods, we observe that the proposed robust identification method yields superior performance in improving classification accuracy. It is consistently better than other competing methods across different subjects with average relative improvement to the second best method be $15.4\%$. This results indicate that by exploiting the consistency assumption of dynamic pattern, we are able to identify subsequences that are more distinctive to the underlying neural activity patterns. Such subsequences are therefore more likely to be the signals of interest.

## VI. CONCLUSION

In this paper, we proposed a robust signal identification algorithm to automatically identify signals of interest from time series. The algorithm selects random subsequences, from which a dynamic model is learned as the representation of the entire time series. By repeating the process iteratively, a relatively best model and the corresponding signals of interest are identified. As the number of iterations increases, we can guarantee that the selected data do not contain erroneous segments up to a pre-specified probability value. As an evaluation, we use model learned from identified signals of interest for a classification task. Experiments on both synthetic and real ECoG data demonstrate the effectiveness of proposed method comparing to other automatic or manual approaches. We are planning to extend this strategy to analyze ECoG data from multiple regions of recordings for the same task to gain more insights on how brain signals propagate between different brain regions over time.

## ACKNOWLEDGMENT

This work was supported by the NIH (EB00856, EB006356 and EB018783), the US Army Research Office (W911NF-08-1-0216, W911NF-12-1-0109, W911NF-14-1-0440) and Fondazione Neurone.

## REFERENCES

[1] T. Svendsen and F. K. Soong, "On the automatic segmentation of speech signals," in *ICASSP*, vol. 12. IEEE, 1987, pp. 77–80.
[2] F. Zhou, F. Torre, and J. K. Hodgins, "Aligned cluster analysis for temporal segmentation of human motion," in *FG*. IEEE, 2008, pp. 1–7.
[3] F.-L. Chung, T.-C. Fu, V. Ng, and R. W. Luk, "An evolutionary approach to pattern-based time series segmentation," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 5, pp. 471–489, 2004.
[4] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Lu, "A review and comparison of changepoint detection techniques for climate data." *Journal of Applied Meteorology & Climatology*, vol. 46, no. 6, 2007.
[5] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *ICDM*. IEEE, 2001, pp. 289–296.
[6] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, 2013.
[7] H. Chen, N. Zhang *et al.*, "Graph-based change-point detection," *The Annals of Statistics*, vol. 43, no. 1, pp. 139–176, 2015.
[8] J.-i. Takeuchi and K. Yamanishi, "A unifying framework for detecting outliers and change points from time series," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 4, pp. 482–492, 2006.
[9] L. Citi, E. N. Brown, and R. Barbieri, "A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 10, pp. 2828–2837, 2012.
[10] S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert, "Learning and inferring motion patterns using parametric segmental switching linear dynamic systems," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 103–124, 2008.
[11] H.-K. Lee and J.-H. Kim, "An hmm-based threshold model approach for gesture recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 10, pp. 961–973, 1999.
[12] S. Tierney, J. Gao, and Y. Guo, "Subspace clustering for sequential data," in *CVPR*, 2014, pp. 1019–1026.
[13] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[14] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[15] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
[16] G. Schalk, J. Kubanek, K. Miller, N. Anderson, E. Leuthardt, J. Ojemann, D. Limbrick, D. Moran, L. Gerhardt, and J. Wolpaw, "Decoding two-dimensional movement trajectories using electrocorticographic signals in humans," *Journal of neural engineering*, vol. 4, no. 3, p. 264, 2007.
[17] R. Zhao, G. Schalk, and Q. Ji, "Coupled hidden markov model for electrocorticographic signal classification," in *ICPR*. IEEE, 2014, pp. 1858–1862.
[18] Z. Wang, A. Gunduz, P. Brunner, A. L. Ritaccio, Q. Ji, and G. Schalk, "Decoding onset and direction of movements using electrocorticographic (ecog) signals in humans," *Frontiers in Neuroengineering*, vol. 5, 2012.
[19] A. Y. Ng, M. I. Jordan, Y. Weiss *et al.*, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.
[20] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
[21] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, vol. 14, no. 2, 1995, pp. 1137–1145.