Abstract—In this paper, we consider the problem of localizing the subsequence in time series which contains the dynamic pattern of interest. This is motivated by brain computer interface application where we need to analyze the dynamic pattern of brain signals in response to external stimulus. We treat the localization as a binary label assignment problem and formalize a mixed integer linear programming (MILP) problem. The optimal solution is obtained by minimizing a cost function associated with label assignment subject to empirical constraints induced by data acquisition process. We first experiment with synthetic data to evaluate the effectiveness of the proposed MILP formulation and achieve 10.8% improvement on $F_1$-score. We then experiment with electrocorticographic (ECoG) data for a classification task and achieve 8.8% improvement on accuracy using subsequences localized by our method compared to other methods.

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. The major cause of this issue is that the data collection process itself is often insensitive to the dynamic pattern change over time. To be inclusive, data collection is often conducted continuously in time regardless of the presence or absence of interested event. For example, surveillance camera record videos all the time without knowing the presence or absence of abnormal event. The uncertainty in localization of dynamic pattern can also be caused by the technical limitation of data collection process. For instance, in the application of biophysical signal analysis, the data collection process follows a pre-defined experiment protocol. However, the starting moment of data collection may not be coincide with the actual response of the subject due to the variation of response time. Similarly, the stop moment of data collection is not aligned well with the end of subject response. In order to analyze the dynamic pattern possessed by signals of interest, it is crucial to identify the signals segment first. In this work, we address the problem of localizing signals of interest from discrete-time series containing irrelevant segments for the purpose of dynamic pattern classification. The rest of this paper is organized as follows. In section II, we review techniques to identify signals from un-segmented data. In section III, we define the problem and list the assumptions we use. We formalize the optimization problem in section IV. The experiments on synthetic data and electrocorticographic (ECoG) data are discussed in section V. Finally we draw conclusion and point out future work.

II. RELATED WORK

Existing approaches for temporal pattern localization in time series data can be divided into several major categories. The first major category treats time series data as a random process whose data points are samples from probabilistic distribution. The signals of interest are found by detecting change-point in time series. To declare a change-point, methods evaluating the difference between two distributions are used. The two major modeling issues need to be addressed in this category of approach is 1) how to characterize the probabilistic distribution given samples and 2) how to determine whether two distributions are different. To characterize the distribution, various statistics have been proposed to describe the random process in time series. To declare a change-point, methods evaluating the difference between two distributions are used. The two major category is to formalize the localization as a segmentation problem where an optimal segmentation can produce the minimum cost in terms of the pre-defined optimization objective. The segmented data can then be used for pattern recognition. Representative works include piece-wise linear model [11], max-margin segmentation [12], [13], [14], [15], and clustering [16], [17], [18]. Keogh et al. [19] provided a comprehensive review on existing algorithms on segmentation of time series represented by piece-wise linear function. The third major category directly models time series using probabilistic dynamic models and convert the localization problem into a model fitting and probabilistic inference problem such as Gaussian AR [20], GGM [21], HMM [22], HSMM [23], [24], SLDS [25], [26], [27], HCRF [28], Coupled HMM [29], Sequence Memorizer [30], etc. An evaluation metric is derived from the fitted model to determine whether a change point occurred. Existing approaches often require the availability of ground truth label of whether a data sample belongs to the signals of interest in order to train the identification model. We propose an approach that can be trained unsupervisely by exploiting the relationship among
labels. This strategy is useful when ground truth labels are difficult to obtain such as in brain signals analysis application.

III. PROBLEM STATEMENT

A. Definition and Goal

We now formally define the temporal pattern localization problem as follows. Suppose we have a multi-dimensional time series \( x^L = \{x_0, ..., x_{L-1}\} \), \( x_i \in \mathbb{R}^d \), \( \forall i = 0, ..., L-1 \), where \( d \) is the dimension of each sample and \( L \) is the length of the sequence. Let \( x'_n = \{x_n, ..., x_{n+l-1}\} \) \((0 \leq n \leq L - l, 1 \leq l \leq L)\) denote a subsequence consisting of \( l \) consecutive samples with the first sample be \( x_n \). Let \( y^I_n \) be a binary variable indicating the label of subsequence \( x'_n \). A subsequence \( x'_n \) is called signal subsequence with \( y^I_n = 1 \) when it only contains samples from signals of interests. A subsequence that contains samples of irrelevant data is called irrelevant subsequence. The corresponding \( y^I_n = -1 \). The union of all signal subsequences is called signal segment. The remaining part is called irrelevant segment. For segmentation problem, we want to separate the signal segment from irrelevant segment. The overall goal is to learn a mapping function \( f : \mathbb{R}^d \rightarrow \mathbb{R} \). Then the label is decided as follows

\[
y^I_n = \text{sign}(f(x'_n))
\]  

Finally, we can identify the signal segment by taking the union of all the subsequences \( x'_n \) with \( y^I_n = 1 \). An illustration of different components in a time series is shown in Fig. 1.

B. Assumptions

We assume each time series \( x^L \) contains one and only one signal segment of length \( l' \geq L/2 \). This majority assumption is reasonable in our application where signals of interest usually dominate during data collection process. We also assume the mapping function is a linear function of \( x'_n \) which can be parametrized as follows

\[
f(x'_n) = w^T x'_n + b
\]  

Despite the use of linear model here, our formulation is general enough to allow the use of nonlinear model through kernel trick [31].

IV. OPTIMIZATION PROBLEM

In the sequel, we omit \( l \) in the superscript for compact notation. The value of \( l \) will be determined empirically during experiment satisfying \( 1 \leq l < L/2 \). From Eq. (1), we have

\[
y_n f(x_n) > 0
\]

In addition, we can scale the parameters \( (w, b) \) so that

\[
\min(y_n f(x_n)) = 1,
\]

which yields

\[
y_n f(x_n) \geq 1
\]

When the label \( y_n \) is available during training, one popular way of solving this problem is support vector machine (SVM) [32]. The parameters \( (w, b) \) are learned by solving the following optimization problem, which essentially maximizes the distance from decision boundary to the closest data point.

\[
\min_{w, b} \frac{1}{2} ||w||^2 \\
\text{s.t. } y_n (w^T x_n + b) \geq 1, \; n = 0, ..., L - l
\]

We can allow mis-classification by introducing positive slack variables.

\[
\min_{w, b, \xi_n} \frac{1}{2} ||w||^2 + \gamma \sum_{n=0}^{L-1} \xi_n \\
\text{s.t. } y_n (w^T x_n + b) \geq 1 - \xi_n, \; \xi_n \geq 0, \; n = 0, ..., L - l
\]

where \( \gamma \) is a constant weighting the penalty on mis-classification. However, the challenge is that the ground truth label \( y_n \) is not available during training. Instead, we have some other information on \( y_n \) induced from the assumption that exactly one signal segment exists in a sequence. We reformulate the optimization problem as follows.

\[
\min_{y_n, w, b, \xi_n} \frac{1}{2} ||w||^2 + \gamma \sum_{n=0}^{L-1} \xi_n \\
\text{s.t. } y_n (w^T x_n + b) \geq 1 - \xi_n, \; \forall n = 0, ..., L - l
\]

\[
\sum_{n=0}^{L-1} y_{n+1} - y_n \leq 2, \; \forall i = 0, ..., \left\lfloor \frac{L}{2} \right\rfloor - 1
\]

\[
y_n \in \{-1, 1\}, \; \xi_n \geq 0, \; \forall n = 0, ..., L - l
\]

where \( \left\lfloor x \right\rfloor \) is the largest integer smaller than or equal to \( x \). The second inequality constraint indicates the presence of at least one signal segment. The third set of inequality constraints requires the signal segment to have at least length \( L/2 \) since \( |y_{n+1} - y_n| = 1 \) indicates \( x_n \) is either the beginning or the end of signal segment.

We adopt the framework proposed by [33] to decompose Eq. (5) into two sub-problems. The overall optimization problem is solved iteratively by alternating between two sub-problems. For the first sub-problem, we assume \( y_n \) is known, then Eq. (5) reduces to Eq. (4), where we solve for \((w, b)\). For the second sub-problem, we use \((w, b)\) estimated in the first sub-problem and solve for \(y_n\). We introduce new variables \( u_n = \frac{y_n + 1}{2}, \; z_n = |y_{n+1} - y_n| \) and \( a_n = w^T x_n + b \). The second
Fig. 1. Illustration of different parts of discrete-time sequence of length $L$. Each small rectangle represents a discrete-time sample. The length $N$ signal segment $x_0, \ldots, x_{a+N-1}$ is shaded, where $a$ is the starting index. The two white segments $x_0, \ldots, x_{a-1}$ and $x_{a+N}, \ldots, x_{L-1}$ are irrelevant segments. The subsequence in green bounding box is an example of length $M$ signal subsequence and the subsequence in red bounding box is an example of irrelevant subsequence.

The second sub-problem Eq. (6) is a mixed integer linear programming problem. We solve the first sub-problem using \texttt{liblinear} [34]. We solve the second sub-problem by combining branch-and-bound and cutting plane methods [35]. The overall algorithm is summarized in Algorithm 1. Notice that the algorithm is described for one sequence. The generalization to multiple sequences is straightforward by adding the constraints associated with each sequence together.

\begin{algorithm}[H]
\caption{MILP based temporal localization}
\label{alg:MILP}
\begin{algorithmic}[1]
\Require $L$: length of complete sequence, $l$: length of subsequence, $S$: number of iterations.
\Ensure Signal segment $x^*$
\State 1: Divide complete sequence into subsequence $x_n$ and set initial value of $y_n$, $\forall n = 1, \ldots, L - l + 1$
\For {$s = 1$ to $S$}
\State 2: \{w, b\} $\leftarrow$ solve Eq. (5) given $y_n, x_n$
\State 3: \{u_n\} $\leftarrow$ solve Eq. (6) given \{w, b\}, $x_n$
\State 4: $y_n = 2u_n - 1$
\EndFor
\State 6: return $x^* \leftarrow \text{Union}(x_n), \forall y_n > 0$
\end{algorithmic}
\end{algorithm}

The purpose of using synthetic data is to evaluate the quality of solution to Eq. (6) as we can compare the integer programming solution to the ground truth subsequence labels, which are known beforehand. We choose $L = 100, l = 60, l = 20$ and randomly select a number $\alpha$ between 0 and $L - l'$. The subsequence labels $y_n$, $n = 0, \ldots, L - l$ are generated as

$$y_n = \begin{cases} 1, & \text{if } \alpha \leq n < \alpha + l' \\ -1, & \text{otherwise} \end{cases}$$

Then the corresponding sequence satisfies the assumptions we described in section III-B. To solve Eq. (6), we need to know the function value $a_n$ for each subsequence $x_n$. We generate $a_n$ by drawing samples from normal distribution with mean depending on $y_n$.

$$a_n \sim \mathcal{N}(y_n, 1), \quad y_n \in \{-1, 1\}$$

A. Synthetic Data Experiment

We perform experiments using both synthetic data and human brain signals.

V. EXPERIMENT

We perform experiments using both synthetic data and human brain signals.
B. ECoG Data Experiment

The ECoG data was recorded from a patient with electrodes placed subdurally on the surface of brain for solely clinical purpose of identifying epilepsy seizure foci prior to surgical resection. The subject 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. The subject can move joystick to eight different directions. Each trial lasts 3-5 seconds with average hit time 1.4 second. The trial protocol is illustrated in Fig. 2. Multiple trials were recorded during experiment session. We select 20 sequences for each direction of movement as our dataset.

The ECoG signal is originally sampled at 1200Hz for each electrode. We follow the same process as in [29] to extract envelope features from $\gamma$ band 70-170 Hz and downsample the signal to 400Hz. Fig. 3 shows two instances of filtered time series with color indicating signals from different electrodes. All sequences last 2 seconds, yielding $L = 800$ points for each sequence. The assumptions we described in section III-B is motivated by this application. Since we are only interested in the signal segment corresponding to the brain’s response to external stimulus. In this case, the external stimulus is the visual cue of joystick movement. During the experiment, though we have accurate control of the visual cue onset, we do not know the accurate onset of brain signals due to different reaction time in different trials.

For real data we do not have ground truth label for subsequence. In order to evaluate the quality of identified signal segment. We perform a classification experiment where the classifier is learned using identified signal segments and original complete sequences. The overall experiment is divided into two phases namely the identification phase and the classification phase. In the identification phase, we perform segmentation for all the sequences simultaneously using Algorithm 1 by aggregating constraints from each sequence together. As an initialization, we choose $l = 200$ and the labels of subsequences are given by

$$y_n = \begin{cases} 1, & \text{if } n \geq 400 \\ -1, & \text{otherwise} \end{cases}$$

Therefore the presumed signal segment has length $l' = 600$. After the identification phase is finished. We obtain a set of signal segments from the original set of sequences. We found in experiment the localization results become stable after a few iterations. For comparison, we use other methods that can perform temporal clustering and localization including SC [36], ACA [16] and HMM+R [22]. For baseline, we use the initial assignment of signal segment. The second phase is classification where we split the dataset into two subsets. We train a HMM classifier using signal segment identified by each method of comparison from training subset and test the classification accuracy on testing subset. Since the same classifier is used, the final results only depend on the quality of localization.

The overall accuracy of each method is shown in Table II. From the accuracy results we first see that all the methods exceed the chance-level. This indicates the presence of different dynamic patterns corresponding to different stimulus responses. Second, compared to baseline method, MILP improves the accuracy significantly by 30.0%. Compared to dynamic model based approach [22], MILP achieves 8.8% improvement. MILP also outperforms unsupervised clustering methods [16] and [36] by 17.0% and 20% respectively. These results demonstrate the effectiveness of the proposed approach.

VI. CONCLUSION

In this paper, we addressed the time series pattern localization problem motivated by brain signal analysis application. We formalize the localization as binary classification problem which decides whether a subsequence belongs to signal segment or not. We leverage on the empirical knowledge from data collection process to constrain that each sequence should
contain exactly one signal segment. To handle the unknown labels of subsequence during learning, we decompose the learning problem into two sub-problems by alternating the optimization between model parameters and label assignments. We solve the proposed mixed integer linear programming problem using branch-and-bound and cutting plane methods. We experiment this formulation on both synthetic data and real ECoG data. On synthetic data, we demonstrate the improvement of localization quality. On real data, we use classification as a surrogate metric to evaluate the quality of localization since the ground truth labels are not available. The improvement of classification accuracy shows the effectiveness of the localization method. For future work, we are interested in replacing the heuristics of using fixed subsequence length \( l \) and model it as a variable.

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 Fon-dazione Neurone. We thank Dr. John E. Mitchell for the discussion on problem formulation.

REFERENCES


