Robust Signal Identification for Dynamic Pattern Classification

Rui Zhao[1] zhaor@rpi.edu

Gerwin Schalk[2] schalk@wadsworth.org

Qiang Ji[1] jiq@rpi.edu

1. Rensselaer Polytechnic Institute, Troy, NY, USA. 2. Neural Injury and Repair Lab Wadsworth Ctr, NYS Dept of Health, Albany, NY, USA

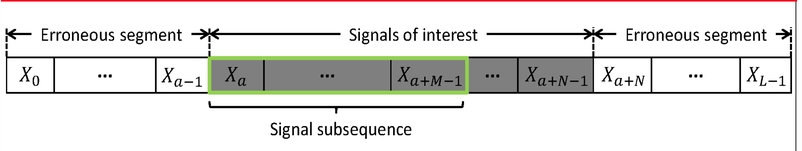
Overview

- **Goal**: Identify signals of interest in time-series that are contaminated by erroneous segments.
- **Challenge:** The start and duration of the signals of interest are difficult to determine.

Our Contribution

- Proposed a framework that can robustly identify the signals of interest from time series, which combines RANSAC method and HMM for joint identification and classification.
- Provided empirical evidence obtained on both synthetic and human brain ECoG data that proposed framework improves identification and classification quality.

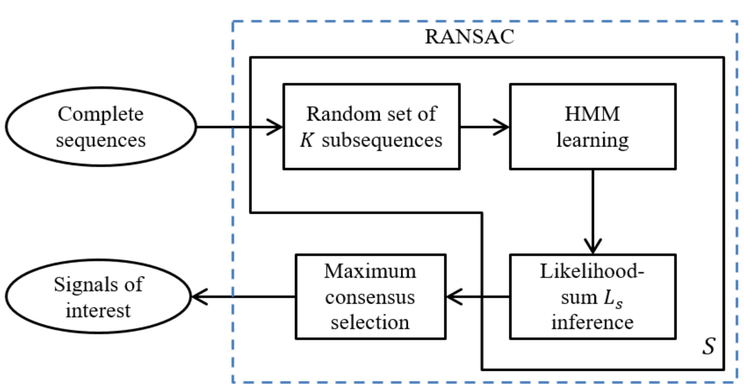
Problem Statement



- Notation: Sequence $\{X_t\}, t = 0, \dots L-1$ contains signals of interest $\{X_a, \dots, X_{a+N-1}\}$ with start time a and duration N.
- Assumption: signals of interest consists of consecutive samples with duration $N \geq 0.5L$. Signals from same class share similar dynamic pattern.
- Goal: estimate a and N for given sequences.

Methods

Phase I. Identification



Probability of successful identification: p

Number of selections: $S = \frac{\ln(1-p)}{\ln 1-(1-\epsilon)^K}$ (Eq. 2)

Likelihood sum: $L_s = \sum_i l_i^s, s = 1, \dots S$ (Eq. 3)

HMM is learned using EM algorithm.

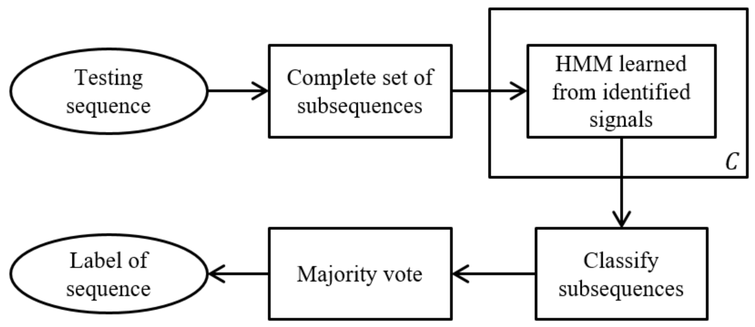
Algorithm 1 Robust signal identification

Input: K: number of sequences, M: length of subsequence, S: number of iterations

Output: Signals of interest

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

Phase II. Classification



C classes of signals. $\hat{k} = \arg\max_k P(\mathbf{Y}|\theta_k)$

Algorithm 2 Time series classification using subsequences

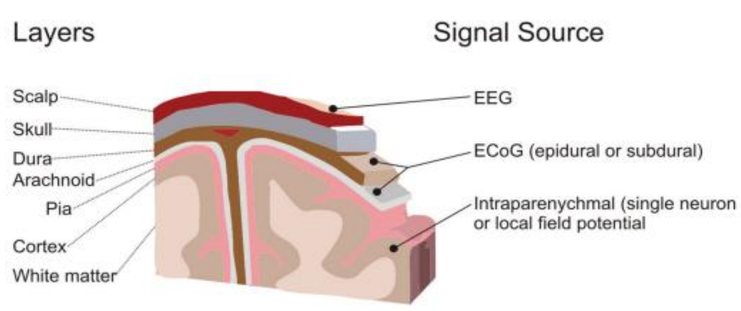
Input: $\mathbf{X}$: testing sequence, M: length of subsequence

Output: Class label

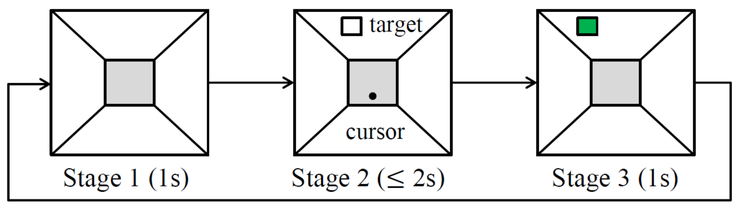
1: $n \leftarrow 0$, $a[1 \dots C] \leftarrow 0$
2: **while** $n + M - 1 <$ length of $\mathbf{X}$ **do**
3: $l \leftarrow classifier(\mathbf{X}(n : n + M - 1))$
4: $a[l] \leftarrow a[l] + 1$
5: $n \leftarrow n + 1$
6: **end while**
7: **return** $label \leftarrow \text{argmax}_i\, a[i]$

Data

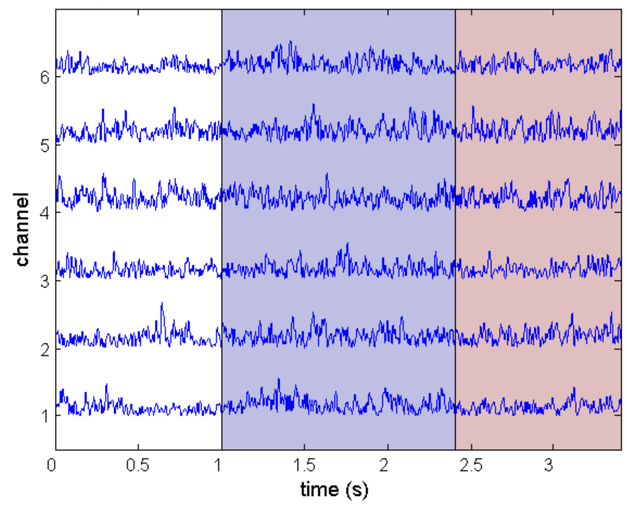
- **Synthetic data:** three classes of univariate time series with different waveforms.
- **ECoG signal**: high resolution and high signal noise ratio in both space and time. Important for brain computer interface (BCI) application.



- **Human motor control task:** Hold joystick to move cursor on the screen to hit a virtual target located in one of eight possible locations.



- **Feature extraction:** From two channels that cover motor cortex, perform spectrum filtering followed by Hilbert transform to obtain envelope features from γ band (70-170 Hz).



Experiments

A. Results on synthetic data

- Average log-likelihood per subsequence

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

- Identification quality

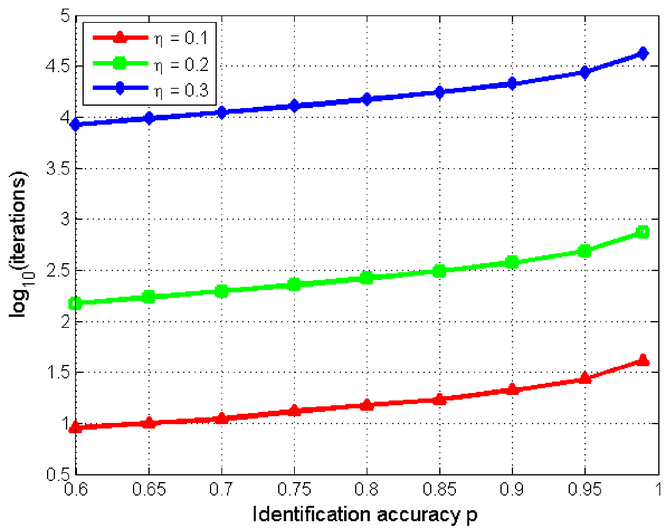
Class	1	2	3	Average	Entire
Precision (%)	81.8	82.5	80.5	81.6	70.0
F1-score (%)	89.9	89.6	89.2	89.6	82.4

- Classification accuracy 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**

B. Results on ECoG data

- Determine S based on p and ϵ, given $K = 10$.



- No groundtruth of segmentation available. Evaluate classification accuracy.

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)

Conclusion

- By evaluating on synthetic data for both identification and classification, the proposed method can robustly identify the signals of interest from erroneous corrupted sequences.
- Using proposed robust identification, we improved the classification accuracy of motor signal pattern in ECoG data despite the uncertainty of temporal location of the signals.
- For future work, we plan to extend this method 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.