Spatial Proximity based Subspace Decomposition for Movement Direction Decoding of Local Field Potentials

Vijay A. Tadipatri[1], B. Vikrham Gowreesunker[1], Ahmed H. Tewfik[1], Nuri F. Ince[1, 2, 3], James Ashe[2, 3], Giuseppe Pellizzer[2, 3]
1. Department of Electrical Engineering, University of Minnesota
2. Department of Neuroscience, University of Minnesota
3. Brain Sciences Center, VA Medical Center, Minneapolis, MN, USA

Abstract - **Local Field Potentials (LFP) provides higher spatial resolution and SNR than EEG data and can be used to construct a Brain Computer Interface. In [7], we have shown that movement direction decoding can be done with about 90 % classification accuracy using spatial patterns (CSP) and Error Correction Output Codes (ECOC). However, a major challenge in this study is to make this method more robust to inter-session variability of the LFP data, where state-of-the-art results are in the high 70 percent. In [8], we have demonstrated that LFP features that are recurrent across sessions can be extracted using a subspace learning method and used to improve the CSP +ECOC classifier.**

In this work, we propose an extension of the subspace learning method that exploits the spatial topology of the channels. This allows us to learn spatially diverse features, while previously the subspaces were being learned independently of the channel layout. We proposed a method where a block of samples from neighboring channels is used to find the subspaces and decode the directions. This approach is analogous to analyzing an 8x8 pixel map in image processing. Furthermore, this method allows a spatio-temporal classification, and it is indeed observed that different directions were providing higher accuracies at different time blocks. The proposed method can boosts the accuracy by at least 6% to bring classification to the mid 80 percent. Furthermore, we show early results where adding a pilot trial from the test session can be used as a calibration to further improve the spatio-temporal classification.

I – INTRODUCTION

Current Brain Computer Interfaces [1] suffer from the lack of robustness due to signal variability in time. The inconsistency of brain signals over time is documented [2]-[5]. It is unclear if the brain signals are correlated with similar tasks. Specifically, the decoding of motor control based on bran activity collected over time is not addressed by the current Brain Machine Interface systems. In general, a bypass solution is suggested by using large training data sets and repeated learning techniques.

The main contribution of this paper is a neighborhood based subspace decomposition algorithm to overcome the long term decoding problems. The algorithm seeks to train subspaces that are local in time and spatial configuration. We show that this method is more successful in identifying motion related behaviors specific to a particular movement direction. We use multiple subspaces to represent the recurrent signal while ignoring the non-recurrent part of the signal.

The selection of Local Field Potentials is due to their higher spatial resolution and SNR than the non-invasive modalities. Also, LFP signals are more robust than the Single Unit Activity (SUA), which are unstable even within a day. We apply our method to data that is collected over several weeks. In specific, we demonstrate that model is robust in time.

We also explore the use of trials from a test session as possible calibration trials. Also, an insight into the use of direction specific time blocks is provided.

The paper is organized in the following way. Section II discusses the data collection paradigm and the challenges in the problem at hand. In Section III, an overview of the feature selection algorithm is presented. Then we provide our novel Subspace Projection algorithm in Section IV. Results are discussed in Section V.

II – DATA

The Local Field Potentials were acquired from rhesus monkey subject - H564. The monkey was left-handed and was trained to move a robotic manipulandum in one of the eight directions, 0, 45, 90, 135,180, 225, 270, and 315 when provided a visual stimulus for the direction. The experimental paradigm proposed by Georgopoulos et.al. [6] for movement is used in the experiment as shown in Figure 1.

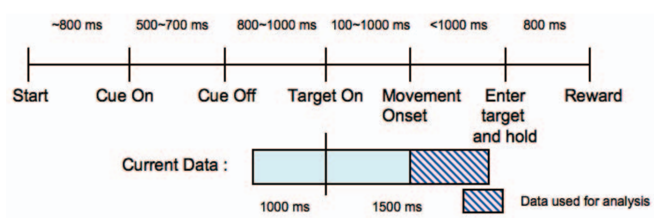

Figure 1: Experimental time line set up for a single trial

The data was collected using two 64 channel - 10x10 Utah microelectrode grid arrays. They were placed in the primary and the dorsal pre-motor cortex areas of the brain. This data is low-pass filtered in 0.4 – 200 Hz band at a sampling rate of 1 KHz. Visual inspection of the channels revealed artifacts, power line noise or large DC wanders.

Those particular channels were removed from the analysis. Consequently, 116 useful channels were obtained for the subject H564. We refer the reader to [7] for more information on the data. In this study we our models were trained on one second of data after the movement cue.

Data was collected on three sessions spread over 8 days. There is a gap of 7 days between the first and the second session and a gap of 1 day between the second and third.

III – FEATURE GENERATION

Common Spatial Patterns

During our visual inspection of data, it was observed that the mean LFP activity had different patterns for different directions. To exploit this we used Common Spatial Patterns [8]. The motivation and the performance of this classifier is discussed in [7]. Common Spatial Patterns (CSP) are generalized eigenvectors of the average covariance matrices of two conditions. These provide spatial weights to the two conditions such that the variance of the weighted linear combination of the channel data is discriminative across the conditions. A Linear Discriminant Analysis (LDA) classifier is then enough to classify the two classes.

Error Correction Output Codes

Common Spatial Patterns is a supervised binary classifier. However, to employ it efficiently we use error correction output codes. The idea is to have a redundant assortment of binary CSP+LDA classifiers. Here we use two types of classfiers

1. Pairwise Classifiers: These classifiers are one single direction vs. another single direction. We implement such classifiers in an exhaustive manner encompassing all 28 possible combinations of directions. Ex: {0} vs. {45}, {0} vs. {90}… {270} vs. {315}.
2. Hierarchical Classifiers: In these classifiers, we group the neighboring directions into one class and pit them against a group in the diametrically opposite direction. We use a group with 2 directions, 3 directions and 4 directions. Ex: {0,45} vs. {180,225}, {0,45,90} vs. {180,225,270}…

Orthogonal Subspace Pursuit

The impetus for subspace approach is that there exists a subspace such that the recurrent features pertaining to a particular direction "live" on it [9]. It seems that the neural data can be segmented as two components - A recurrent component that approximates to the condition related component (here direction related); and a non-recurrent component that explains to the state and conditions at the time of the experiment. By estimating these recurrent components we will be able to separate the trials into different classes.

Orthogonal Subspace Pursuit (OSP) is used to extract the subspaces. This dictionary learning method is a data dependant technique. In this method dictionary is iteratively updated using clustering and orthogonalizing techniques.

To employ the OSP, data from all the trials is collected in a huge array. The dimensions of the array reflect the time length of each trial and the number of channels for each channel. The algorithm is initiated by setting this data set as the initial dictionary. The first signal from this array is selected and a sparse representation with the remainder of the elements is pursued. This provides us with an initial subspace. All the signals that can be represented by this are clustered and removed from the array. This process is repeated till the array is empty.

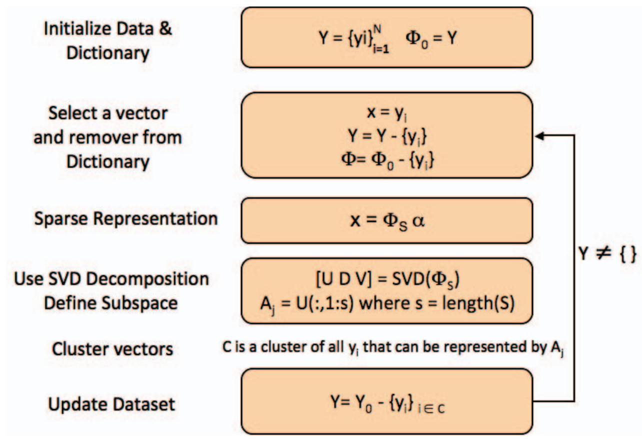

Figure 2: A Flowchart for the Orthogonal Subspace Pursuit

IV – SUBSPACE PROJECTION USING SPATIAL PROXIMITY

In [9] we have shown that a subspace-based approach to classification could be used to extract direction dependent recurrent patterns and provide high decoding power (DP). In this work, we use the spatial neighborhood of the channels to learn subspaces that exploit the spatial arrangement of the sensors.

The strategy is to take time samples of the data from a block of channels in close spatial proximity. An example with a 2x2 block is shown in Figure3. This is analogous to using an 8x8 pixel group in image processing. We vectorize the resulting block of data to create a spatio-temporal element. Due to the close neighborhood, the signal behavior across these channels is usually very similar. We try to capture this similarity using the subspaces.

In the training phase, we used non-overlapping blocks of data from the same direction to learn a good subspace for that element. Orthogonal Subspace Pursuit is employed to learn subspaces for each particular element. This process is repeated for all the individual elements. The same process is repeated for all the directions.

Once the subspaces have bee calculated, each trial is again divided into spatio-temporal blocks. The subspace that provides the minimum projection error is selected for each of the blocks. This process is applied for all the spatio-temporal blocks. After this is done, we re-assemble the data into its original form. The paired and hierarchical CSP classifiers are calculated from this projected data. LDA classifiers corresponding to each of these classifiers is calculated.

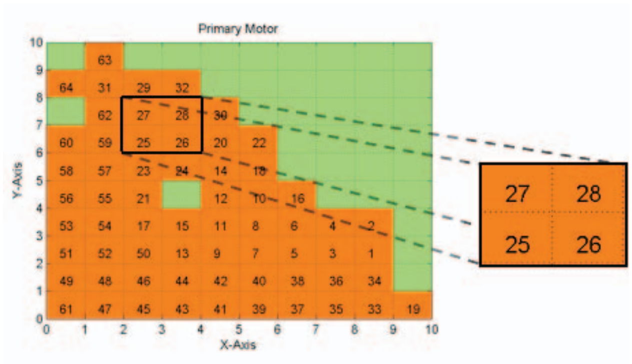

Figure 3: Grouping of channels based on their spatial neighborhood

Each test trial is decomposed into the spatio-temporal element, similar to the training data. Then the subspace that is the closest to each spatio-temporal element is found, and then the element projected onto that subspace. This projected data is then sent to the classifiers to determine the direction. The same CSP and ECOC algorithms used in [7] are employed for classification purposes.

Post-Processing Algorithm

The classification of the trials is done continuously over time. A block of 1s is considered at a time. The trial is classified as one of the eight directions based on this data. Then with a shift of 100ms, the next block of data is considered and the process repeated. The experiment proceeded in the way shown in Figure 4. For each time block the decision of the algorithm is recorded.

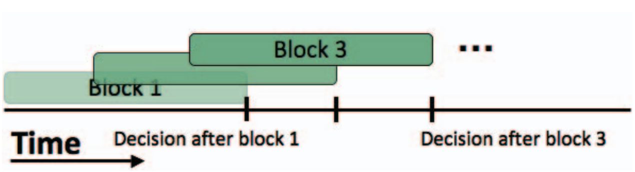

Figure 4: Time Blocks and Decoding Power Calculation

During this analysis it was observed that some directions were identified earlier than others. For example, directions 0 and 135 had high decoding power at the end of the movement onset while 45, 225, 270 and 315 had high Decoding power at the starting of the movement onset. Figure 5 shows the decoding power of the different directions at different time points. To exploit this effect, we used time windows to extract those directions to emphasize such directions.

Addition of a calibration trial from test-set

In this part of the study, we use a single trial per direction from the test trials and add it to the training set. Thus we add 8 trials corresponding to the eight directions to the training data. These test trials act as a calibration data for the model. Please note that we only re-model the CSP classifiers with the calibration data. The subspaces are calculated from the original training set only. The test trials added into the training set are selected at random. To have a fair estimate of this procedure, the process is repeated 5 different times with different trials added to the training set. The average of the decoding power after this is reported.

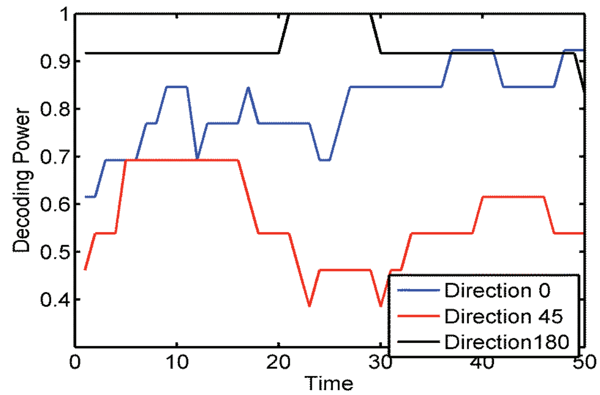

Figure 5: Decoding Power obtained at the end of time blocks for different directions

V – RESULTS

We consider and analyze data corresponding to three different sessions for a monkey H564. We use the trials from session 1 and 2 as training data set. The trials from the third session are used for the evaluation of the decoding power. This measure, Decoding Power, is the ratio of all the correctly decoded trials in the test data to the total number of trials in the test data. Since there can be 8 different directions that can be selected, the probability of random classification is 12.5%.

The results from the algorithms described above are summarized in Table1 below. To have a fair assessment of the algorithms the same CSP and LDA classifiers are used for all the algorithms. These also act as a baseline for comparison of our algorithms. Table 1 provides the decoding power of all the algorithms in different conditions.

It can be seen that, using subspace-projected data increases the decoding power considerably. In all the cases it can be seen that the subspace method consistently over performs the baseline method.

Methods	Original Data	Subspace Projected Data
CSP+ECOC	77.8%	82%
CSP + ECOC + Time Windows	80.7%	83.65%
Adding a Trial	81.65%	83.91%

TABLE 1: DECODING POWER OF DIFFERENT ALGORITHMS. ALL THE ALGORITHMS USE SESSION 1 AND 2 FOR TRAINING THE CLASSIFIER AND SESSION 3 FOR EVALUATION OF DECODING POWER

VI – CONCLUSIONS

In this paper we have described a novel method that uses the spatio-temporal features to build subspaces. We have shown that we can extract consistent features over time by using the subspace method. An approach to extract spatially close features is described. This approach provides up to 84% decoding accuracy, which is about 7% better than the baseline approach available. This paper also highlights that different directions can be decoded at different points in time. Time windows are used to tap the maximum decoding power from all the directions. These results are encouraging, and using better temporal feature selection techniques the results can be even furthered. Also we have shown that adding a calibration trial to the data increases the decoding power as expected. Throughout all these post-processing steps, the subspace projection method over performs the baseline CSP method.

REFERENCES

[1] J. Wolpaw et.al., "Braincomputer interfaces for communication and control," Clin. Neurophysiology, p. 767791, Sep 2002.

[2] W. Jensen and P. J. Rousche, "On variability and use of rat primary motor cortex responses in behavioral task discrimination" Journal of Neural Engineering, vol. 3, no. 1, pp. L7–L13, 2006.

[3] B. Blankertz, M. K. R. Tomioka, F. U. Hohlefeld, V. Nikulin, andK. Robert Muller, "Invariant common spatial patterns: Alleviating non-stationarities in brain-computer interfacing," in In Ad. in NIPS 20. MIT Press, 2008, p. 2008.

[4] N. Murata, M. Kawanabe, A. Ziehe, K.-R. Meuller, and S.-i. Amari, "Online learning in changing environments with applications in supervised and unsupervised learning", Neural Netw., vol. 15, no. 4, pp. 743–760, 2002.

[5] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D.Lawrence, "Dataset Shift in Machine Learning". The MIT Press, 2009.

[6] AP Georgopoulos et. al, "Neuronal Population Coding of Movement Direction", Science 226, 1986

[7] Ince, N.F.; Gupta, R.; Arica, S.; Tewfik, A.H.; Ashe, J.; Pellizzer, G.; "Movement direction decoding with spatial patterns of local field potentials", Neural Engineering, 2009. 4th International IEEE/EMBS Conference on April 29 2009-May 2 2009 Page(s):291 – 294

[8] H. Ramoser, J. Mueller-Gerking, and G. Pfurtscheller, "Optimal Spatial filtering of single trial EEG during imagined hand movement", IEEE Trans. Rehab. Eng, 8: 441-446, 2000.

[9] V. B. Gowreesunker, A. H. Tewfik, V. A. Tadipatri, N. F. Ince, J. Ashe, G. Pellizer; "Overcoming Measurement Time Variability in Brain Machine Interface", in Proceedings of 31st Annual International Conference of the IEEE EMBS on September 2-6, 2009, pp 3134 -3137