Guided Learning Algorithms: An Application of Constrained Spectral Partitioning to Functional Magnetic Resonance Imaging (fMRI)

Henry L. Phillips[1], Peter B. Walker[1], Carrie H. Kennedy[2], Owen Carmichael[3], and Ian N. Davidson[3]

[1] United States Navy
[2] Marine Corps Embassy Security Group
[3] University of California – Davis
{henry.phillips,peter.b.walker}@navy.mil, carrie.kennedy@usmc.mil, ocarmichael@ucdavis.edu, davidson@cs.ucdavis.edu

Abstract. Innovations in neuro-technology have created a potential gap in our ability to measure human performance and decision making in dynamic environments. Therefore, a need exists to create more reliable testing methodologies and data analytic solutions. The primary aim of this paper is to describe work to integrate subject matter expertise with algorithms designed to measure human brain activity in real time. Specifically, Guided Learning using constrained spectral partitioning to increase the reliability and interpretability of fMRI data is explicated and applied as a test case to the Default Mode Network in the elderly population. How Guided Learning can be further applied to other neuro-imaging technologies that may be more conducing to furthering the field of augmented cognition is discussed.

Keywords: augmented cognition, functional connectivity, fMRI.

1 Introduction

Since its inception as a scientific field at the turn of this century, augmented cognition has been one of the fastest growing research areas influencing several different academic disciplines including engineering, psychology, and human factors [1]. The excitement surrounding this field of research has allowed for an explosion of innovation in the ability to capture human performance and decision making through innovations in neuro-technology. However, a gap may soon develop as researchers attempt to develop research methodologies to integrate these innovations into the laboratory environment. The primary aim of this paper is to describe progress integrating subject matter expertise (SME) with algorithms designed to measure human brain activity in real time. We term this work Guided Learning to reflect the pursuit to develop a general class of algorithms that incorporate SMEs to help identify meaningful and insightful patterns within dense datasets.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 709–716, 2013.

For the research described in this paper, we will focus on the analysis and use of functional Magnetic Resonance Imaging (fMRI). fMRI provides the unique opportunity to visualize neural activity in the brain in an on-line modality. Perhaps one of the most promising applications of fMRI data has been on the analysis of functional connectivity [2], the temporal correlation of neuronal activation for spatially discrete locations. The excitement over this analytic approach is due, in part, to the applicability of these findings in both clinical and diagnostic settings. For example, functional connectivity studies have explored issues such as the study of post-traumatic stress disorder (PTSD), chronic traumatic encephalopathy (CTE) and traumatic brain injury (TBI) [3].

However, a review of the literature on the use of fMRI in the analysis and interpretation of functional connectivity data illustrates several limitations in the current methodology used to interpret this data. For example, analysis of functional connectivity data is often limited by poor test/re-test reliability. As is seen in Figure 1, results from separate scans may yield different results. In this figure, we see the results for two resting state fMRI data sets of the same healthy young individual, acquired in short succession. Barring a major medical event between the two scans, the spatial and temporal patterns of resting state activity should be very similar. Yet, the data analytic approach used on this data set identified different patterns of activation across the two sets of scans.

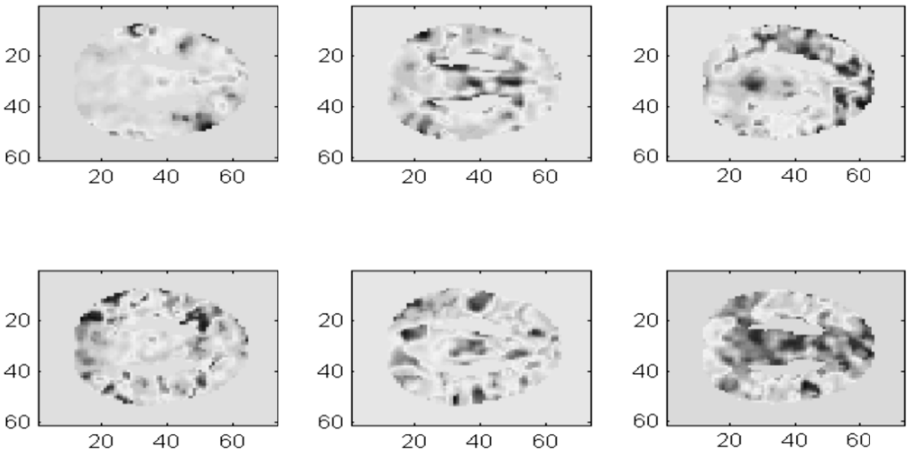

Fig. 1. Lack of test-retest reliability in fMRI. A healthy young individual received two fMRI scans in rapid succession. The top and bottom rows show the three top-ranked clusterings for the first and second scans.

Most often, datasets such as those illustrated in the figure above are modeled as a graph: each node corresponds to an instance and (the weight of) each edge corresponds to the similarity between two instances. To partition the dataset, a commonly used algorithm is spectral clustering [4], which finds the (relaxed) normalized min-cut of the corresponding graph. Traditional spectral clustering only applies to a single graph. However, in a wide range of applications, in addition to the graph under consideration, there is auxiliary information available in the form of a second graph, which shares the same set of nodes with the first graph, but has a different set of edges. A number of alternatives exist under which a second graph might meet these assumptions, including 1) the edges of the second graph are constructed based on a different set of features; 2) the edge weights of the second graph are computed using different similarity functions; and/or 3) the two graphs represent the evolution of a graph over time. Intuitively, the extra knowledge from a second graph may help to identify a better partition than the best one that can be identified using the first graph only [5]. As will be explained later, for optimal utilization, this information must be partially theoretically driven, using qualitative input from a SME, rather than derived solely through algorithm application.

A direction already explored by the community is to consider any two such graphs as two independent views and combine them into one graph, to which the traditional spectral clustering algorithm is then applied [6]. However, this approach relies on the assumption that the two views are complementary and thus helpful to each other, which is not always the case in practice. The approach outlined in this paper attempts to transfer "knowledge" from more stable fMRI images to those scans where a particular activation function may not be as readily apparent. To accomplish this, a form of spectral clustering to two separate groups of fMRI scans was applied. Unlike traditional clustering algorithms, such as spectral clustering that attempt to segment a graph based on a single image, our approach incorporates knowledge from multiple graphs that might share the same set of nodes with the first graph, but have a different set of edges.

2 Limitations of Spectral Clustering

Clustering analytic approaches remain one of the most widely used techniques for exploratory data analysis and have been used extensively in areas ranging from image processing to functional connectivity analysis [7,8,9]. Furthermore, some forms of clustering may be more applicable to particular problem sets over others. For example, spectral clustering has been argued to be superior to traditional clustering algorithms like K-means because it yields a deterministic polynomial-time solution, provides researchers the ability to model arbitrary shaped clusters, and affords equivalence to certain graph cut problems.

However, as mentioned previously, traditional clustering like spectral clustering can only be applied to a single graph. In a wide range of applications, such as the analysis of several distinct fMRI scans, it would be more beneficial to combine properties from different graphs to form a single cut from the data comprising the set. This approach, which has only been recently introduced by the clustering community, has come to be known as constrained spectral clustering.

Constrained spectral clustering attempts to incorporate auxiliary information from separate graphs to help improve clustering on both graphs. In general, constrained clustering is a category of techniques that tries to incorporate Must-Link (ML) and Cannot-Link (CL) constraints into existing clustering algorithms. It has been well studied on algorithms such as K-means clustering, mixture modeling, hierarchical clustering and density-based clustering.

Multi-view Spectral Clustering

In contrast to constrained spectral clustering, traditional multi-view spectral clustering algorithms attempt to consider a set of two graphs as two independent views and combine information from both graphs into one graph. However, in its basic form, this relies on the assumption that the two graphs are complementary to one another. That is, it is assumed that both graphs are noise-free. In the work presented in this paper, we no longer assume the two graphs are complimentary. Rather, our approach, which we term *Constrained Spectral Partitioning* (hereafter, CSP) attempts to discover an alternative direction of finding a cut whose edge weight is minimized based on information about both graphs [10, 11, 12].

We contend that CSP fits into a general category of algorithms, i.e., Guided Learning. This term is appropriate because, in addition to algorithm application, we allow for SME input to maximize the identification of appropriate cut(s) for a series of scans. Assume we have two graphs, an exemplar graph and a target graph, that share the same set of nodes but have different sets of edges or edge weights. The goal of applying SME input, is the utilization of information from the exemplar graph to identify a more representative and replicable cut on the target graph.

Further, we believe this work represents a hitherto unattempted technique to "close the loop" with respect to Augmented Cognition. Previous work in algorithm development for Augmented Cognition was focused on utilizing machine learning to maximize human performance. However, our work here attempts to close the loop by allowing for SME input to further maximize the efficiency of the learning algorithm.

3 Background and Graph Theory Notation

Formally, the set of points in a network may be represented as a weighted undirected graph $G = (V, E)$, where the nodes are the set of points in a feature space and an edge is formed between each pair of nodes. The weight (similarity) on each edge $w(\boldsymbol{i},\boldsymbol{j})$ is a function of the similarity between nodes $\boldsymbol{i}$ and $\mathbf{j}$.

To more effectively interpret the graph, grouping or clustering techniques may be applied that attempt to segment the graph into more similar sub graphs containing similar features. This may be accomplished by partitioning the graph into multiple disparate sets $V_1, V_2, \ldots, V_m$, where some measure of similarity is high among vertices within set V_i but very low across different sets of vertices between sets V_i and V_j.

However, as was discussed previously, the traditional approaches do not take into account those cases where we may only wish to extract certain features from some of the graphs. In this section, we describe how we adapted the classical spectral clustering to increase reliability when segmenting across one or several graphs.

To accomplish this, CSP was applied such that one or several source graphs were identified and used to segment several target graphs. The knowledge to transfer was derived from the source graph in the form of what we termed, degree-of-belief constraints. Specifically, let $G_S(V, E_S)$ be the source graph and $G_T(V, E_T)$ the target graph. A_S and A_T are their respective affinity matrices. Then, A_S can be considered a constraint matrix with only *ML* constraints. It carries the complete knowledge from the source graph, and we can transfer it to the target graph using our constrained spectral clustering formulation:

$$\underset{\mathbf{v}\in\mathbb{R}}{argmin}\, \mathrm{v}^T \bar{L}_T\mathbf{v}, s.t.\, \mathbf{v}^{\mathrm{T}}\mathrm{A}_{\mathrm{S}}\mathbf{v} \geq \alpha, \mathbf{v}^{\mathrm{T}}\mathbf{v} = \mathrm{vol}(\mathrm{G}), \mathbf{v} \neq \mathrm{D}_{\mathrm{T}}^{1/2}\mathbf{1}$$

α is now the lower bound of how well the knowledge from the source graph must be enforced on the target graph. The solution to this is similar:

$$\bar{L}_T\mathbf{v} = \lambda(\bar{\mathrm{A}}_{\mathrm{S}} - \frac{\beta}{\mathrm{vol}(\mathrm{G_T})}\mathrm{I})\mathbf{v}$$

Note that since the largest eigenvalue of $\bar{\mathrm{A}}_{\mathrm{S}}$ corresponds to a trivial cut, in practice we should set the threshold such that $\beta < \lambda_1\mathrm{vol}(G)$, $\lambda 1$ is the second largest eigenvalue of $\bar{\mathrm{A}}_{\mathrm{S}}$. This will guarantee a feasible eigenvector that is not the trivial cut.

4 Application of CSP to fMRI Analyses

In this paper, we apply CSP to the analysis of the Default Mode Network [13]. The DMN is an interconnected brain system that activates simultaneously and periodically while in rest state. It has been hypothesized that the DMN is only active when individuals are focused on internal tasks such as daydreaming, memory retrieval, or introspection. The DMN is composed of several subsystems including part of the medial temporal lobe, medial prefrontal cortex, the posterior cingulated cortex, and the lateral and inferior parietal cortex.

The DMN provides a relevant test-bed for measuring the reliability/stability of the technique presented herein, as it has been reliably shown that the DMN is more pronounced and observable in young healthy patients than individuals suffering from various mental pathologies [14]. In this case, we compared DMN signatures across fMRI scans for young healthy patients with those suffering from early and late Alzheimer's disease. If CSP is able to reduce the impact of noise within fMRI scans across different abnormal groups, then it is hypothesized that CSP might increase the sensitivity to allow for the detection of specific phenomena – in this case, degree of similarity to an exemplar DMN among a set of Alzheimer's patients.

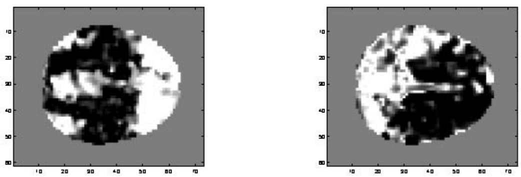

Fig. 2a and 2b. Figure 2a (left) displays segmentation results from a normal healthy participant. Figure 2b (right) displays segmentation results from a participant diagnosed with Alzheimer's disease. DMN activation appears as lighter colored pixels.

As shown in Figure 2a above, segmentation of a graph from a normal participant (P1) captures the DMN (the light pixels). However, if we apply spectral clustering to another graph constructed from an Alzheimer's patient's (P2) fMRI scan, the normalized mincut shows an entirely different pattern (Figure 2b).

Here, CSP was applied as a new approach for assessing inter-individual clustering commonalities at a population level. The principal benefit yielded by the reliance of an exemplar scan as the basis for partitioning decisions among target group members is an improvement in the reliability of intra-individual fMRI clustering in the target group. CSP incorporates user-provided guidance about which voxels should and should not cluster together. Our approach is to use the clustering of an exemplar scan to generate guidance (constraints), and use them in CSP to cluster a target scan. The exemplar scan is explicitly assumed to exhibit desirable or representative clustering behavior. If multiple diverse clusterings of the target scan all yield similar cut costs, CSP identifies the one most similar to that of the exemplar at each timepoint, yielding improved intra-individual clustering reliability across different scans.

We used real resting-state fMRI scans of young and elderly (Normal, Mild Cognitive Impairment, and Demented) individuals to demonstrate the advantages of CSP over spectral partitioning. In comparing the two groups, we applied segmentation algorithms so as to identify the DMN for each group of scans. As has been previously shown, we expected that the tightness of this clustering would be decreased in elderly individuals, especially those with Alzheimer's disease. Therefore, we identified an exemplar scan of a young individual whose spectral partitioning clearly indicated the DMN as one of its clusters. We then applied CSP to partition target scans including young and elderly individuals based on constraints derived from this exemplar.

In order to assess whether CSP increased reliability over and above the use of spectral partitioning alone, we first compared the test-retest reliability of the spectral partitioning with that of CSP on a group of individuals who received a pair of fMRI scans at two different time intervals. For each pair of fMRI scans, we calculated the percent difference in spectral partitioning and CSP costs between scans (i.e., the absolute difference in partition costs divided by average partition cost). The data from this study supported the claim that CSP increases reliability over and above what is found from spectral partitioning alone.

Our next analysis focused on assessing the biological validity of CSP. To assess the biological validity of CSP, we compared partition costs for each of three groups of participants: Elderly, Mild Cognitive Impairment, and Demented. Figure 3 below plots the measure of reliability within different groups of participants. As can be seen from the figure below, the average CSP cut cost was greater in MCI compared to healthy elders, and greater in dementia compared to MCI. The MCI partition costs were more variable, spanning most of the range of normal and demented values. The difference between normal elderly and demented cut costs was statistically significant ($p = .046$). This finding is consistent with previous findings that suggest that the DMN is less pronounced (therefore exhibiting higher cut costs) in individuals with Alzheimer's disease.

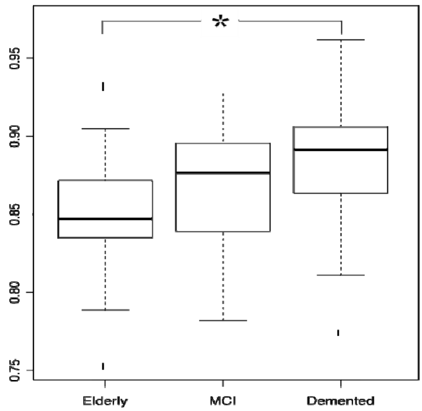

Fig. 3. Partition costs within healthy elderly, MCI, and demented groups. * Significant group difference at the $p<.05$ level.

Several key findings were presented above. First, we showed that deriving clustering constraints from an exemplar fMRI scan, and using them to cluster a target fMRI scan via CSP leads to resting state fMRI clustering results that are more reliable than traditional clustering approaches. In addition, we showed that the application of CSP in this dataset resulted in better identification of known biological differences between elderly individuals that are associated with neurodegenerative disease.

5 Guided Learning and Augmented Cognition

As newer technologies allow for the visualization of the brain during performance and decision making tasks, it is imperative that researchers incorporate paradigms that allow for more accurate assessments of network activation during these tasks. In this paper, we argued that the way forward for Augmented Cognition in this endeavor is a "closed loop" whereby SME input is incorporated with machine learning algorithms to measure and assess human performance. Algorithms such as CSP will provide augmented cognition researchers the opportunity to incorporate real and practical theory with basic science in the hopes of closing a potential gap between technology and the paradigms augmented cognition practitioners incorporate in their research.

Acknowledgements. The research reported in this paper was supported by Office of Naval Research Grants NAVY 00014-09-1-0712 and NAVY 00014-11-1-0108. The opinions of the authors do not necessarily reflect those of the United States Navy. We would also like to acknowledge Owen Carmichael from the Alzheimer's Disease Institute at the University of California – Davis for his assistance in data collection.

References

1. Schmorrow, D.D., Reeves, L.M. (eds.): HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565. Springer, Heidelberg (2007)
2. Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S.J.: "Functional Connectivity: The Principle Components Analysis of Large (PET) Data Sets. Journal of Cerebral Blood Flow and Metabolism 13, 5–14 (1993)
3. Hoge, C.W., Castro, C.A., Messer, S.C., McGurk, D., Cotting, D.I., Koffman, R.L.: Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N. Engl. J. Med. 351(1), 13–22 (2004)
4. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 888–905 (2000)
5. Wang, X., Davidson, I.N.: Active Spectral Clustering. In: Proceedings of the IEEE International Conference on Data Mining, ICDM 2010 (2010)
6. Bach, F.R., Jordan, M.I.: Learning spectral clustering. In: NIPS (2003)
7. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: Saul, L.K., Weiss, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems 17, pp. 1601–1608. MIT Press, Cambridge (2005)
8. Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., Rosen, B.R.: Functional mapping of the human visual cortex by magnetic resonance imaging. Science 254, 716–719 (2006)
9. Stam, C.J., Reijneveld, J.C.: Graph Theoretical Analysis of Complex Networks in the Brain. Nonlinear Biomedical Physics 1, 3 (2007)
10. Basu, S., Davidson, I., Wagstaff, K. (eds.): Constrained Clustering: Advances in Algorithms, Theory and Applications. Data Mining and Knowledge Discovery, vol. 3. Chapman & Hall/CRC.Morup (2008)
11. Wang, X., Davidson, I.N.: Flexible Constrained Spectral Partitioning. In: KDD, pp. 563–572 (2010)
12. Wang, X., Qian, B., Davidson, I.: On Constrained Spectral Clustering and Its Applications. Journal of Knowledge Discovery and Data Mining (in press)
13. Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L.: The brain's default network. Annals of the New York Academy of Sciences 1124(1), 1–38 (2008)
14. Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V.: Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proceedings from the National Academy of Sciences U.S.A. 101, 4637–4642