A Semi-automatic Framework for Mining ERP Patterns

J. Rong[1] D. Dou[1] G. A. Frishkoff[2,3] R. Frank[3] A. D. Malony[1,3] D. M. Tucker[3,4]

[1]Computer and Information Science, University of Oregon; Eugene, OR
[2]Learning Research and Development Center (LRDC); Pittsburgh, PA
[3]NeuroInformatics Center; Eugene, OR
[4]Electrical Geodesics, Inc. (EGI) ; Eugene, OR

OBJECTIVE

Event-related potentials (ERP) are created by averaging across segments of Electroencephalogram (EEG) data in different trials. ERP data is a mixture of artifacts, noise, and components that are related to brain operations. To reduce the labeling efforts of neuroscientists, we propose a semi-automatic framework for mining ERP patterns. These patterns will be used as pattern elements for Neural ElectroMagnetic Ontology (NEMO) [1] development.

Our framework[7] includes data preprocessing, unsupervised learning and supervised classification rule learning. Component decomposition can assist in ERP artifacts detection as well as discovery of real pattern factors of brain activities. We use temporal PCA to decompose the raw ERP data and then further reduce dimension by extracting 25 attributes from each PCA factors. Expectation maximization (EM) clustering algorithm [4] is applied to these PCA factors to separate different ERP patterns. Then, a decision tree learner [3] is built based on the clustering results to obtain the rules for differentiating different ERP patterns. These rules can be used to help refine the rules defined by the domain experts.

ERP PATTERN IDENTIFICATION AND LABELING

- Our framework aims at discovering important concepts, properties and their relationships in ERP data. These concepts and properties will be used in building ERP ontologies. We integrate expert knowledge about ERP patterns and the data mining analysis of ERP data in our framework.

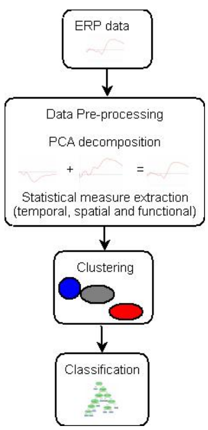

Figure 1. A semi-automatic framework for mining ERP patterns

DATA REPRESENTATION

- Multiple representational spaces
  - Scalp topographic space (Fig 2 A, C)
  - Latent factor space (Fig. 2 B)
  - Neural source space (Fig. 3)

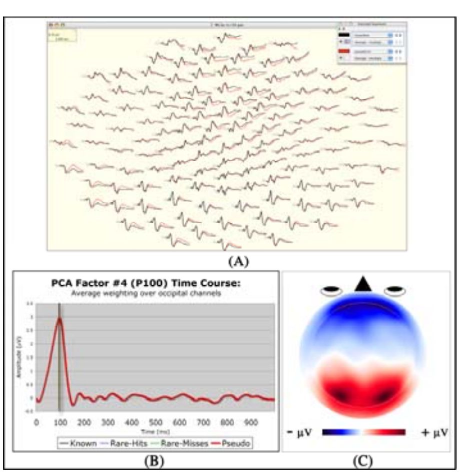

Figure 2. A. 128-channel ERP data showing brain electrical response to word and nonword stimuli. B. Latent temporal principal component analysis (PCA) representation of classical "P100" potential. C. Scalp topography for P100 potential shown in B.

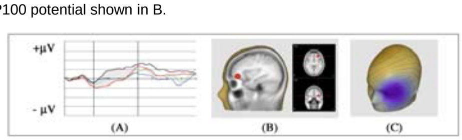

Figure 3. Representation of ERP in source space (from Ref. [2]).

DATA PREPROCESSING

- We analyzed data collected in three visual word studies (Experiment 1, 2, 3). Data were acquired using a 128-channel EEG sensor net. Sampling rate was 250hz. Together the Experiment 1 and Experiment 2 datasets comprise 89 subjects and 6 experimental conditions (#observations = 534) in total. Experiment 3 dataset consists of 36 subjects and 4 experiment conditions that were acquired in a lexical decision task in visual word study (#observations=144).

- ERP data represent a mixture of "signal" (functional brain patterns) and "noise", artifacts and brain activity that is not related to the events of interest). Data decomposition methods can help separate signal from noise and disentangle overlapping patterns. We used temporal PCA. The data set used as input to the PCA is organized with the variables corresponding to time points.

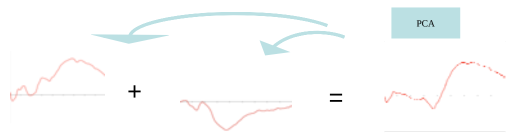

Figure 4. PCA decomposition

- Summary metrics extraction
  - ✓ Temporal metric (TI-max, TI-begin…)
  - ✓ Spatio metrics (ROI, SP-max, SP-min…)
  - ✓ Intensity metrics (IN-mean, SP-cor…)
  - ✓ Functional metrics (Modality, Event….)

CLUSTERING

- Clustering is the partition of data into subsets, or clusters, such that the data in each cluster share a common trait. Clustering is a knowledge discovery process. Ideally, ERP patterns of the same type will be clustered into one cluster.

- We use Expectation Maximization (EM) clustering[4], to automatically separate ERP patterns. EM algorithm is often used to approximate distributions using mixture models. It is an iterative procedure that circles around the expectation and maximization steps. The input to the EM clustering algorithm is the 25 dimension summary metric vectors. After the clustering, we use expert-defined rules to evaluate the clustering results (Table 1, Table 2).

Cluster/ Pattern	0	1	2	3
P100	0	**76**	0	2
N100	**117**	1	0	54
lateN1/ N2	13	14	0	**104**
P300	0	61	**110**	42

Table 1. EM clustering results for Experiment 1 group 2 pattern factors

Cluster/Pattern	0	1	2	3	4	5	6	7
P100	0	1	0	**109**	0	0	0	0
N100	0	0	20	0	8	**85**	2	0
N300	0	0	34	0	14	1	5	0
lateN1N2	0	0	0	0	49	4	**79**	0
P1r	0	0	**76**	0	16	0	9	0
MFN	0	25	0	0	0	0	0	**40**
N400	0	9	0	0	0	0	0	7
P300	**108**	5	0	0	0	0	0	2

Table 2. EM clustering results for Experiment 3 pattern factors

REFERENCES

[1] Dou, D., Frishkoff, G., Malony, A., Tucker, D., & LePendu, P. (2006). Finding NEMO (Neural ElectroMagnetic Ontology): Design & implementation of a sharable EEG/MEG database system. Poster presented at the Human Brain Mapping, Toronto, Canada.

[2] Frishkoff, G. A., Tucker, D. M., Davey, C., & Scherg, M. (2004). Frontal and posterior sources of event-related potentials in semantic comprehension. *Brain Res Cogn Brain Res, 20*(3), 329-354.

[3] Han, J. and Kamber, M., Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.

[4] Dempster, A, Laird, N. and Rubin, D., Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statistical Society, 39:1-38, 1977

[5] Dou, D., Frishkoff, G. A., Rong, J., Frank, R., Malony, A. D., Tucker, D. M., Development of NeuroElectroMagnetic Ontology (NEMO): A framework for mining brain wave ontology, to appear in SIGKDD, 2007

[6] Lependu, P., Dou, D., Rong, J., Frishkoff, G.A., Semantic data modeling and query answering for brain wave ontologies, submitted to ISWC 2007.

[7] Rong, J., Dou, D., Frishkoff, G.A., Frank, R., Malony, A. D., Tucker, D. M., A semi-automatic framework for mining ERP patterns, in *Proc. of IEEE Symposium on Data Mining and Information Retrieval (DMIR), vol 1*, 375-381, Niagara Falls, Ontario, 2007

CLUSTER-BASED CLASSIFICATION

- EM clustering automatically partitions observations into clusters. A related goal is to develop rules that accurately assign observations to clusters. Therefore, after EM clustering, we use classification methods to build a decision tree learner. Decision tree[3] is a flowchart like tree structure learner with each internal node representing an attribute and the leaf node representing a category. We use cluster labels as classification labels and the resulting decision tree classifier is shown in Figure 5.

Figure 5. Decision tree classifier

- During the building of decision tree, at each level, the algorithm always chooses the attribute that is most capable of differentiating different patterns. Table 3 lists the information gain of some attributes, which provides informative information of the importance of attributes. The advantage of using decision tree classifier is that we can generate rules to classify different patterns from the tree. Table 4 compares corresponding rules defined by domain experts and generated by decision tree for N100 pattern.

Attribute	Merit	Ranking
TI-max	0.836	1
IN-mean (ROI)	0.238	2.2
IN-mean (ROCC)	0.224	3.3
SP-cor	0.215	3.6
…	…	…

Table 3. Information gain

Expert-defined rule	Decision tree rule
$\forall n$, FAn = N100 if 150 ms < TI-max <=220 $\wedge$ IN-mean (ROI) < - 0.4 $\wedge$ Event = stimon $\wedge$ Modality = visual	$\forall n$, FAn $\in$ cluster0 if TI-max > 128 $\wedge$ |IN-mean (ROI)| > 2.896 $\wedge$ SP-cor > 0.549 $\wedge$ IN-max > 3.514 …

Table 4. Comparison of expert-defined rules and decision tree generated rules

SUMMARY & CONCLUSIONS

- We introduced a semi-automatic framework for mining ERP patterns. It includes data preprocessing, EM clustering and decision tree classification of ERP patterns.
- An important feature of our framework is that it integrates top-down ERP analysis techniques with bottom–up data mining methods (unsupervised and supervised learning) during the ERP pattern identification and labeling process.
- The split of patterns into multiple clusters has several reasons. PCA decomposition process can be refined; More efficient summary metrics can improve the clustering results and the clustering algorithm can be adjusted to better partition ERP patterns.
- Once the ERP pattern identification and labeling system becomes robust, we can use it to mine the important concepts and properties and their relationship in ERP data. These classes and properties can be used to build ERP ontology. (Dou et.al 2007)
- Ontology-based ERP database design can integrate more semantic meaning of the data. (Paea et.al 2007)