Development of NeuroElectroMagnetic Ontologies (NEMO): A Framework for Mining Brainwave Ontologies

Jiawei Rong[1] Dejing Dou[1] Gwen Frishkoff[2,3] Robert Frank[3] Allen Malony[1,3] Don 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 brain electrophysiological patterns created by averaging electroencephalographic (EEG) data, time-locking to events of interest. In this paper, we propose a generic framework for mining and developing domain ontologies and especially its application to ERP data.

The concepts and relationships of domain (e.g. ERP) ontologies can be mined according to the following steps: pattern decomposition, extraction of summary metrics for concept candidates, hierarchical clustering of patterns for the class and class taxonomy, and clustering-based classification and association rules mining for relationships (axioms) of concepts. We have applied this process to several dense-array ERP datasets. Results suggest good correspondence between mined concepts and rules, on one hand, and patterns that were independently formulated by domain experts, on the other hand. The next goal of our ERP ontology framework is to address some long-standing challenges in conducting large-scale comparison and integration of results across ERP research paradigms and laboratories.

DATA REPRESENTATION

- Multiple representational spaces
  - Scalp topographic space (Fig. 1 A, C)
  - Latent factor space (Fig. 1 B)

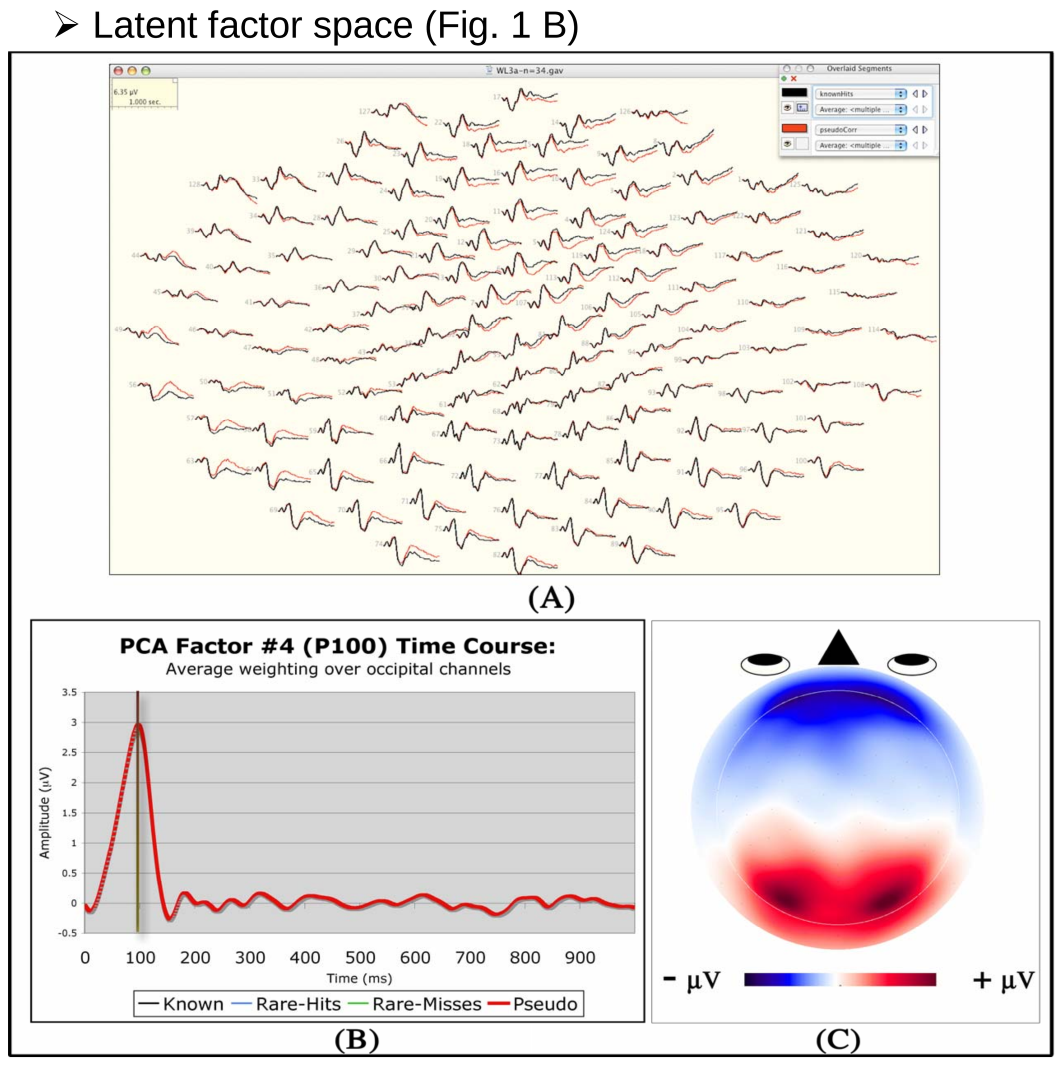

Figure 1. A. 128-channel ERP data showing brain electrical response to word and nonword stimuli. B. Latent temporal principal component analysis (PCA) representation of classical "P100" potential. C. Scalp topography for P100 potential shown in B.

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.

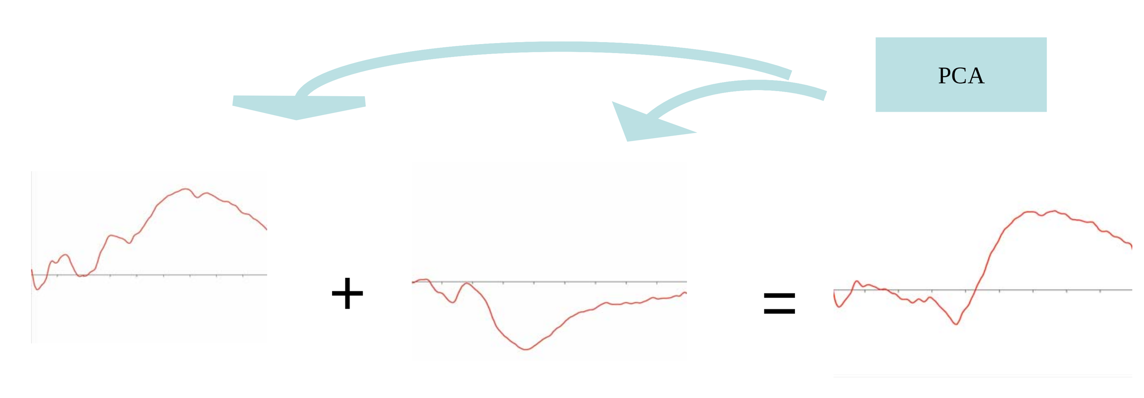

Figure 2. PCA decomposition

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

A SEMI-AUTOMATIC FRAMEWORK FOR MINING DOMAIN ONTOLOGIES

- Classes 🡸 Clustering-based Classification
- Class Taxonomy 🡸 Hierarchical Clustering
- Properties 🡸 Classification
- Axioms 🡸 Association Mining and Classification

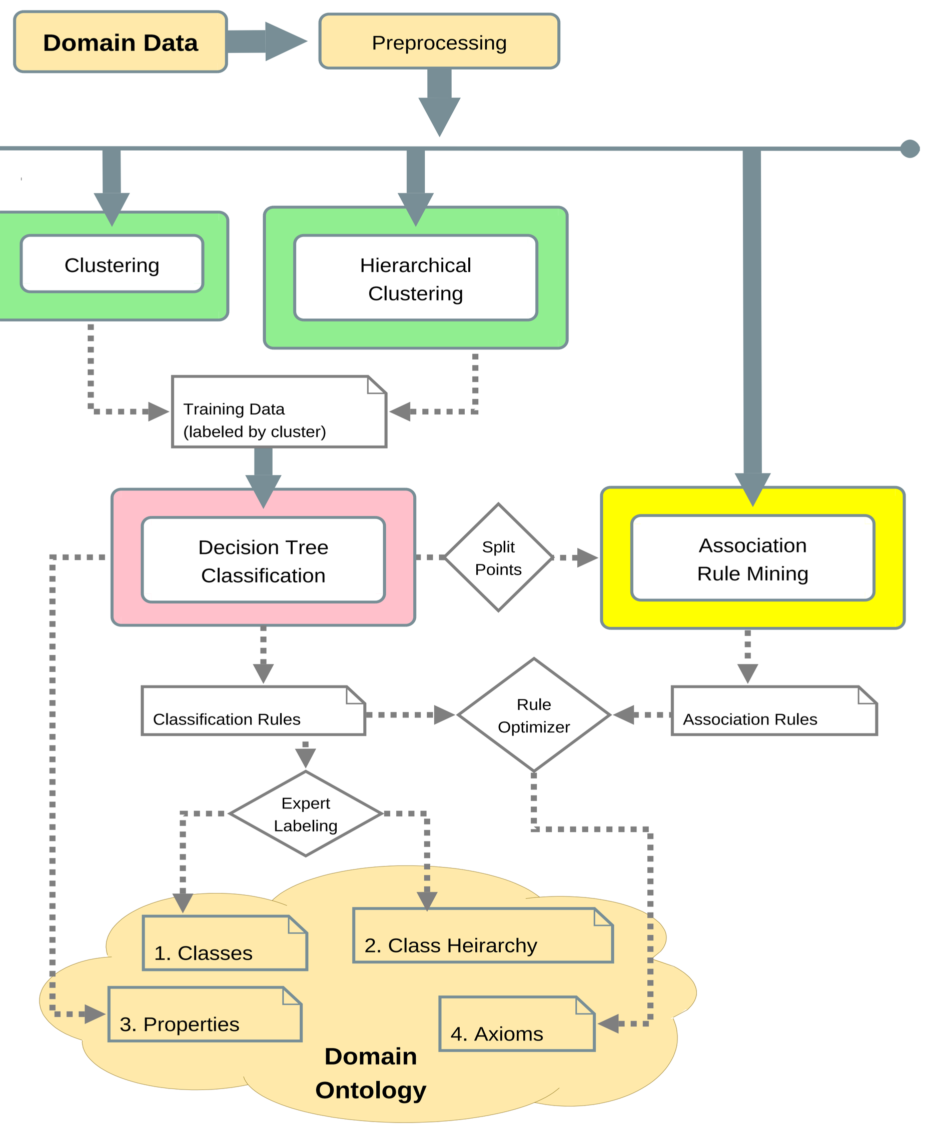

Figure 3. A semi-automatic framework for mining domain ontologies

MINING ERP CLASSES WITH CLUSTERING

- We use Expectation Maximization (EM) clustering, to automatically separate ERP patterns. EM algorithm is often used to approximate distributions using mixture models. It is a procedure that iterates around the expectation and maximization steps. The input to the EM clustering algorithm is the 25 dimension summary metric vectors. Ideally, ERP patterns of the same type will be clustered into one cluster. After the clustering, we use expert-defined rules to evaluate the clustering results (Table 1).

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

MINING ERP CLASS TAXONOMY WITH HIERARCHICAL CLUSTERING

- We used EM clustering in both divisive and agglomerative ways to find the class taxonomy. In the divisive approach, we first put all the data from 8 ERP patterns together in one cluster. Then we repeatedly sub-divide each cluster until the majority of data rows from each event forms a cluster. In the agglomerative approach, we put the data rows into 8 clusters which our neuroscientists wish. Then we further merge them into fewer clusters until all data rows can be put into one cluster. It turns out that the divisive and agglomerative results match each other (Figure 4)

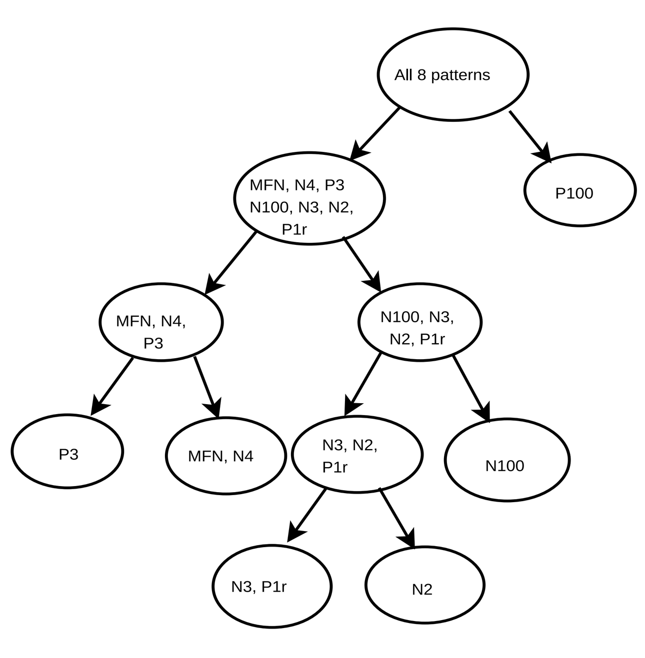

Figure 4. ERP class taxonomy

MINING AXIOMS AMONG PROPERTIES AS ASSOCIATION RULES

- Association rule mining aims at finding frequent patterns in certain datasets. In our case, we use Apriori algorithm to seek the properties that frequently co-occur for the specific ERP pattern factors.

➢ ∀f Factor(f) ∧ SP-cor_minvalue(f, 0.2295)
∧ RareMisses-RareHits_maxvalue(f, 0.5154)
→ IN-LATEM_minvalue(f, -2.0743)

- These association rules can be used in rule optimization process ((A→B) ∧ (A∧B)→C)=>(A→C)) for ERP patterns.

MINING PROPERTIES AND AXIOMS WITH CLUSTERING-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. We use cluster labels as classification labels and the resulting decision tree classifier can achieve the accuracy of 95%. We use OWL to represent datatype properties which are based on those attributes with high information gain (e.g., top 6). We can also use SWRL to represent axioms (classification rules). In FOL, it looks like:

∀f Factor(f) ∧ TI-max_minvalue(f, 128) ∧ IN-mean(ROI)_minvalue(f, 2.896) ∧ SP-cor_maxvalue(f, 0.549) → labeled_as(f, LateN1/N2)

ERP ONTOLOGY DESIGN

- Our initial ERP ontology consists of 16 classes, 1 class taxonomy, 57 properties and their relationships. Figure 5 shows a partial view of a preliminary ERP ontology.

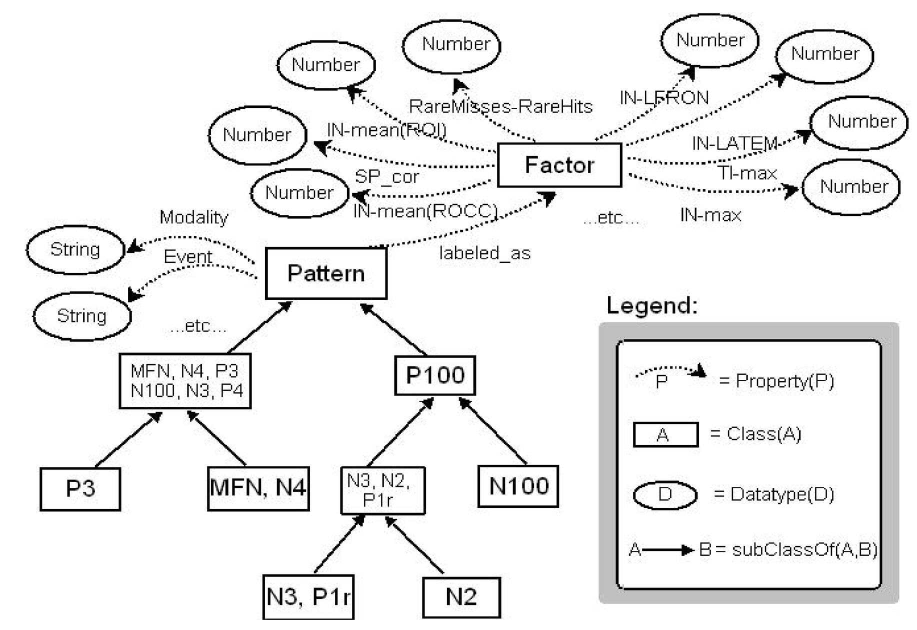

Figure 5. Partial view of a preliminary ERP ontology

ONTOMINER

- OntoMiner (Figure 6) is an ontology mining tool developed by us as the implementation of our framework. At the current version, it integrates EM clustering, clustering-based classification and association rule mining. It is developed based on WEKA.

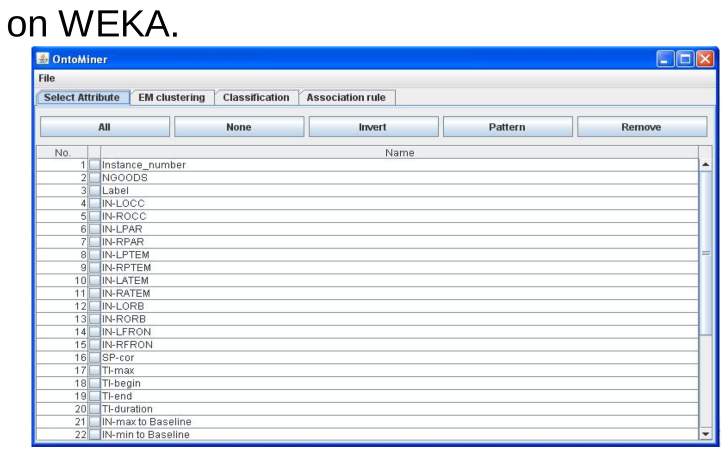

Figure 6. OntoMiner interface

SUMMARY & FUTURE WORK

- We introduced a generic framework for mining domain ontologies and its application to ERP ontology. OntoMiner implements this framework and can be used to mine ERP ontologies. We will continue to enrich the functionality of OntoMiner and integrate it with OntoGUI to facilitate the creation and modification of ontologies.
- For the study of brain activities, there are various data collection methods and data processing methods:
  - Principal Component Analysis v.s. Microstate analysis for the decomposition of ERP data
  - Scalp space v.s. Source space. They represent brain signals at different locations of the brain.
  - MEG v.s. EEG. They are different brain signal collection methods. MEG collects data from magnetic field while EEG from electronic field.
- Each part of our ontology mining framework can be replaced by alternative methods mentioned above and thus generate different brain wave ontologies. Finding the mappings between these ontologies is of great interest to both computer scientists and neuroscientists. For instance, finding the mappings between the dipoles in the dipole model in source space and ERP patterns in the scalp space is a very intriguing topic in the research area. The challenge lies in that the current ontology matching or mapping algorithms rely on text similarity information while for the ontologies generated from time series data, text similarity is not accessible.
- Ontology-based data modeling can represent as much as semantics when storing ERP data. The next goal of NEMO project is an ontology-based integration system which will facilitate the representation and dissemination of ERP data across studies and labs.