To appear, Proceedings of the International Conference on Biomedical Ontologies (ICBO'09).

Development of Neural Electromagnetic Ontologies (NEMO): Ontology-based Tools for Representation and Integration of Event-related Brain Potentials

Gwen Frishkoff[12], Paea LePendu[2], Robert Frank[2], Haishan Liu[2], & Dejing Dou[2]
[1]Medical College of Wisconsin, Milwaukee, WI; [2]Univ. of Oregon, Eugene, OR

Abstract

We describe a first-generation ontology for representation and integration of event-related brain potentials (ERPs). The ontology is designed following OBO "best practices" and is augmented with tools to perform ontology-based labeling and annotation of ERP data, and a database that enables semantically based reasoning over these data. Because certain high-level concepts in the ERP domain are ill-defined, we have developed methods to support coordinated updates to each of these three components. This approach consists of "top-down" (knowledge-driven) design and implementation, followed by "bottom-up" (data-driven) validation and refinement. Our goal is to build an ERP ontology that is logically valid, empirically sound, robust in application, and transparent to users. This ontology will be used to support sharing and meta-analysis of EEG and MEG data collected within our Neural Electromagnetic Ontologies (NEMO) project.

Introduction

In the last two decades, neuroscience has witnessed the development of some exciting new methods for research on human brain function—including high-density electroencephalography (EEG), whole-head magnetoencephalography (MEG), and functional Magnetic Resonance Imaging (fMRI). Each of these methods has contributed important insights on human brain function. At the same time, the proliferation of data has made clear the need for large-scale summary and integration of research results. To meet this need, several groups have been working to develop formal ontologies that can be used for consistent annotation, sharing, and meta-analysis of neuroscience data[1,2].

In the present paper, we describe initial steps in the development of an ERP (event-related potentials) ontology. ERPs are measures of brain electrical activity (EEG or "brainwaves") that are time-locked to experimental events (e.g., the appearance of a word). These measures provide a powerful technique for studying brain function, because they are acquired noninvasively and can therefore be used in a variety of populations —e.g., children and patients, as well as healthy adults. In addition, they provide detailed information about the time dynamics, as well as the scalp spatial distribution, of neural activity during various cognitive and behavioral tasks.

ERP research is likely to enjoy several benefits from the development of ERP ontologies. Historically, progress in this area has been hampered by debates over how to define high-level concepts[3]. As a result, it is has been hard to achieve even informal consensus, let alone quantitative syntheses of results across experiments (i.e., statistical "meta-analysis"). In this context, the process of building an ontology may prove to be a healthy exercise. Where there are debates over concepts, the need to make these concepts explicit will bring controversies into the open. Where there is mere inconsistency in labeling, the existence of a common reference may lead to standards for reporting that will better support cross-lab integration of research results.

To address these aims, we have assembled an international team of ERP researchers and computer scientists to found the Neural Elecromagnetic Ontologies (NEMO) consortium[3,4]. The major goal of our project is to address basic scientific questions in ERP research using ontology-based classification and labeling of ERP data, particularly in studies of language comprehension. The present paper gives an overview of the NEMO project and describes how it builds on and extends other efforts in bio- and neuro-ontology development.

NEMO Framework

Our framework includes the following components:

1. Top-down (knowledge-driven) specification and coding of domain concepts (*NEMO ontologies*);
2. Bottom-up (data-driven) validation and refinement of complex concepts, including tools for *ontology-based labeling of ERP data*;
3. An international *consortium of experts* in ERP methods, with a shared interest in language;
4. An *ontology database* and *inference engine* to enable semantic queries over labeled data.

Each of these components is described in the following sections.

Top-down Ontology Development

Traditional methods for ontology development can be described as top-down or *knowledge-driven*, and are largely manual. The process typically begins with knowledge capture, that is, expert identification of a relatively small set of domain concepts. In NEMO, we have focused on defining concepts that represent spatial and temporal attributes of ERP patterns, as well as some functional (i.e., cognitive) concepts that are of immediate interest for analysis of ERP experiment data (building on previous efforts in the development of cognitive ontologies[1,2]). In addition, because our goal is to use ontologies to develop tools for labeling of ERP data, we have represented data-level concepts in a separate but linked namespace. These first steps in ontology development are documented in NEMOlex, a text document that was modeled after Neurolex (formerly BirnLex[2]). NEMOlex contains natural language descriptions of concepts (classes and relations), organized by categories (e.g., spatial, temporal, and functional).

In the next step, domain experts work with ontology engineers to develop a formal conceptualization of domain-specific concepts. These concepts are subsequently coded in the Web Ontology Language (OWL), and Protégé is used to generate a set of web-accessible documents that can be viewed online (see nemo.nic.uoregon.edu for links to owl ontologies).

Throughout this process we have worked to implement recommendations of the Open Biomedical Ontologies (OBO) community[5]. Domain-specific concepts in NEMO are linked to more basic or foundational concepts, as implemented in the Basic Formal Ontology (Figure 1). Similarly, to facilitate reuse and integration of NEMO with other neuroscience ontologies, we have aligned our efforts with members of the OBO, including fBIRN and NIFSTD. For example, the NEMO concept *scalp* is defined as a proper_part_of NeuroLex class *head.* In addition, we have designed NEMO ontologies to be modular wherever possible. Concepts representing spatial, temporal, and functional objects and

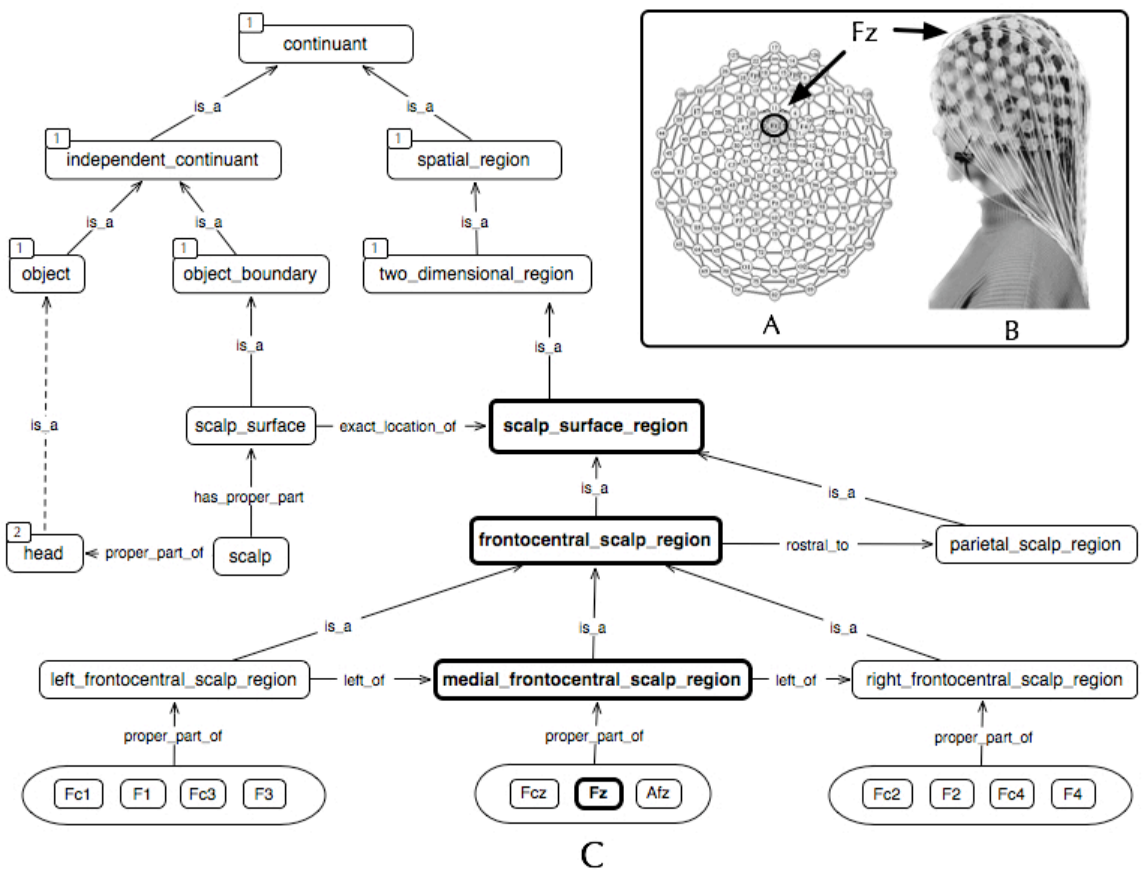

Figure 1. A) International 10-10 electrode layout (i.e., *scalp_surface_region*), with *electrode_location Fz* highlighted; B) EEG net applied to the scalp surface. C) A subset of concepts from NEMO_spatial. Concepts marked by superscript '1' are from BFO; superscript '2' denotes concepts from NeuroLex (formerly BirnLex).

properties are therefore stored in different name spaces (NEMO_spatial, NEMO_temporal, and NEMO_functional). Of key importance are ERP spatiotemporal patterns that are seen in particular experiment contexts. These patterns have distinctive spatial, temporal and functional attributes as described in the following section. Pattern definitions are represented as first-order rules in our merged NEMO_erp ontology.

Bottom-up Validation & Refinement

While ontologies are intended to capture expert (i.e., domain) knowledge, knowledge in certain areas may be uncertain or changing. For example, spatiotemporal ERP patterns, which are the main concepts of interest in the ERP domain, are often ill-defined. The same label (e.g., "N400") can be used to pick out manifestly different entities[3]. Conversely, the same pattern may be called by different names in different experiment paradigms or research groups.

The existence of a standard ERP ontology can help to address this lack of consistency, but there is no guarantee that concepts defined using "top-down" methods will be optimal for classification, labeling, and annotation of actual ERP data. To address this concern, NEMO has adopted a *data-driven* strategy for validating and refining high-level patterns before encoding this knowledge in our ontologies. This strategy is used to augment first-pass ontology engineering steps described in the previous section.

Our approach is outlined in Figure 2. It begins with expert specification of spatial, temporal, and functional concepts, including definitions of patterns that are commonly found in ERP data. These definitions represent expert hypotheses about (a) the ERP patterns that exist and (b) the spatial and temporal attributes that define these patterns (see Ref. [3], Appendix B for concrete examples). To test these hypotheses, we encode these pattern rules in an automatic data classification and labeling tool. ERP datasets are summarized by extracting attribute vectors that constitute a compact summary of the measured data. The values of these spatiotemporal metrics are then compared to rule-specific thresholds for each ERP pattern of interest. Results are recorded in a true/false table, and observations meeting pattern criteria are flagged as instances of that pattern.

Next, we perform clustering on the spatial and temporal values of these summary metrics using the Expectation-Maximization (EM) algorithm[3,4]. The resulting clusters represent candidate pattern classes, which are characterized by the central tendencies of their cluster attributes (e.g., mean latency and amplitude over scalp regions of interest). Based on these results, we refine our initial hypotheses about the number of pattern classes in the ontology and the definitions of these patterns. If similar results are obtained across multiple datasets, this leads in turn to a revision of NEMO ontologies and ontology-based tools for pattern classification and labeling.

We have applied these methods to several datasets[3,4], and results have led to refinements of our methods for ontology-based labeling. In our current ERP labeling tools, for example, we have omitted reference to high-level ERP pattern concepts, such as the "N400." Concepts are still coded in the NEMO_erp ontology, but with provisional notes that indicate they are based on working hypotheses that are awaiting robust empirical testing and validation.

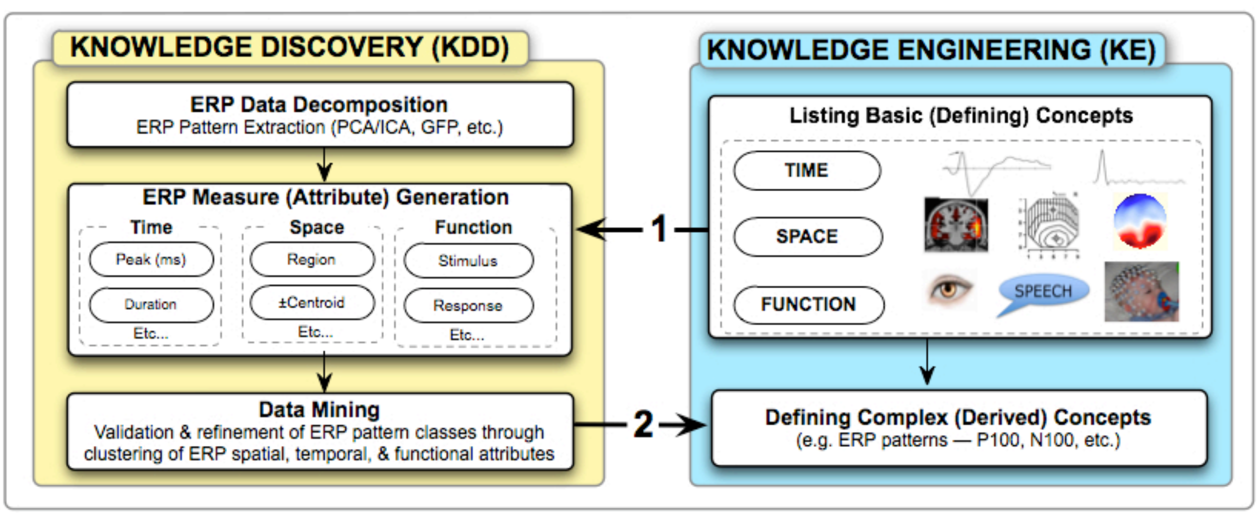

Figure 2. NEMO framework for deriving complex concepts for ERP ontology.

The NEMO Consortium

The application domain for our project is language processing. We have established an international consortium of experts in this area who contribute ERP data from experiments and collaborate on the design and testing of ontology-based tools developed for NEMO. Consortium members include John Connolly (McMaster U.), Timothy Curran (U. Colorado), Dennis Molfese (U. Louisville), Charles Perfetti (U. Pittsburgh), Joseph Dien (U. Maryland), and Kerry Kilborn (Glasgow U.).

The NEMO Ontology Database

The NEMO database will store large numbers of ERP datasets collected from multiple research sites (e.g., from members of our research consortium). As described above, we have developed MATLAB scripts that automatically decompose, classify, label, and annotate ERP data using ontological terms. On the backend, we will support ontology-based querying and reasoning by using specialized databases designed to model the class (subsumption) hierarchy as well as most integrity and cardinality constraints. These databases will be coupled with a reasoning engine (OntoEngine[6]) to support efficient and scalable knowledge-based query answering. For example, consider the following database query:

Return all data instances that belong to ERP pattern classes which have a surface positivity over frontal regions of interest and are earlier than the N400.

In this query, "frontal region" is a clear generalization that can be unfolded into constituent parts (e.g., right frontal, left frontal; see Figure 1). At an even more abstract level, the "N400" is a pattern class that is associated with spatial, temporal, and functional properties (Figure 2). As described above, these three types of concepts are encoded in separate namespaces, and linking concepts are used to combine them for definition of high-level pattern concepts in NEMO. This design allows for a rich and flexible range of queries, which we refer to as *ontology-based queries*[7].

NEMO has investigated several methods of using databases to support ontology-based queries. A view-based approach is commonly used to simplify instance-checking and subsumption-based reasoning by unfolding views at query time. By contrast, we have developed a new method that uses asynchronus, event-driven triggers to forward-propagate the knowledge model so that queries are answered more quickly and efficiently[7].

Summary and Conclusions

In conclusion, we have described a first-generation ontology for representation and integration of event-related brain potentials. The ontology is designed following OBO "best practices" and is augmented with tools to perform ontology-based labeling and querying of ERP data.

We have further described how data mining (i.e., clustering) is used to help validate and refine top-down ERP ontologies. These ontologies will be used to support sharing and meta-analysis of cognitive neuroscience data collected within the NEMO project.

Acknowledgements

This work is supported by a grant from the National Institutes of Health (NIH), award #1R01EB007684. We thank Maryann Martone, Jessica Turner and Angela Laird for helpful discussions.

References

1. Laird A, Lancaster J, Fox P. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics. 2005; 3:65-78.

2. Bug W, Ascoli G, Grethe J, Gupta A, Fennema-Notestine C, Laird A, Larson S, Rubin D, Shepherd G, Turner J, Martone M. The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. Neuroinformatics. 2008; 6:175-194.

3. Frishkoff G, Frank R, Rong J, Dou D, Dien J, Halderman L. A framework to support automated classification and labeling of brain electromagnetic patterns. Comput Intell Neurosci. 2007; 14567.

4. Dou D, Frishkoff G, Rong J, Frank R, Malony A, Tucker, D. Development of NeuroElectro-Magnetic Ontologies (NEMO): A Framework for Mining Brainwave Ontologies. Proc KDD. 2007; 270-279.

5. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg L, Eilbeck K, Ireland A, Mungall CJ, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007; 25:1251-1255.

6. Dou D, McDermotte V, Qi P. Ontology translation on the semantic web. Journal of Data Semantics. 2005; 2: 35-57.

7. LePendu P, Dou D, Rong J, Frishkoff G. Ontology Database: A New Method for Semantic Modeling and an Application to Brainwave Data. Proc SSDBM. 2008; 313-330.