#293

MAPPING ERP MEASURES FROM HETEROGENEOUS DATA: A SIMULATION STUDY

Talk on Sat, 8:00 AM "Neurocognitive Ontologies"

H. Liu[1] G. A. Frishkoff[1, 2] R. M. Frank[1] D. Dou[1]

[1] University of Oregon; Eugene, OR [2]Medical College of Wisconsin; Milwaukee, WI

OBJECTIVES

We describe an ontology-based method for representation and integration of neuroscience data, with an application to event-related brain potentials (ERPs). ERP datasets from different labs are highly heterogeneous, in that they are acquired in different experiment contexts and may be analyzed and described using different methods. To address this issue, we use data mining technique coupled with ontology as a semantic bridge to discover mappings across different spatial an temporal metrics characterizing datasets from different laboratories. A novel application of sequence similarity search techniques has also been proposed in the process of mapping discovery. The mappings and ontologies are being used to support data sharing and cross-laboratory analysis of ERP experiment results within our NEMO (Neural Electromagnetic Ontologies) consortium [2].

METHODS I: SIMULATED ERP DATASETS

Simulated ERP data were designed to model ERPs acquired from 128 EEG channels, 40 "subjects," and two "conditions." The data were derived from a 7-dipole, 3-shell spherical model using Dipole Simulator. Dipoles were located in occipital, posterior-temporal, and mid-frontal regions (**Fig. 1A**), representing 4 patterns: 1) ***P100***, 2) ***N100***, 3) ***N1/N2***, & 4) ***P2*** (**Fig. 1B**). The data were projected to 129 scalp locations, with time series distributed over ~600 ms. (**Fig. 1C**). Individual "subject" ERPs were designed by adding random offsets to the base intensities. Noise was superposed on the simulated ERP data for each of the 80 individual datasets (courtesy of P. Berg).

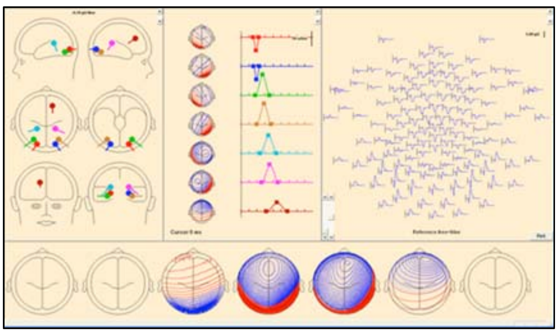

Figure 1. A) Dipole locations. B) Base model intensities for each dipole. C) Scalp data

METHODS II: ERP PATTERN ANALYSIS

The data were decomposed using spatial Independent Components Analysis (sICA), which yielded a small set of discrete temporally independent latent patterns (Fig. 2). Patterns were automatically classified and labeled (e.g., as "P100" or "N100") using data-driven methods described in [1].

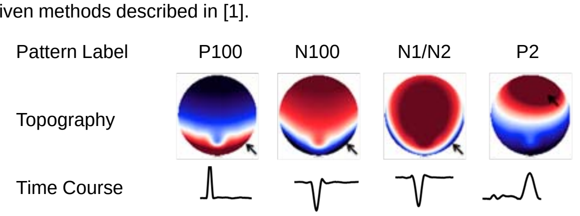

Figure 2. Spatial ICA decomposition (grand average) results. Top row, scalp topographic projection of each latent component. Bottom row, time course of latent component.

METHODS III: EXTRACTION OF ALTERNATIVE METRICS

For each latent pattern, we extracted an alternative set of summary metrics that provided an equivalent characterization of its spatiotemporal dimensions. The alternative measures simulate the case where research groups use different, yet comparable, measures to quantify the temporal and spatial attributes of their data.

SuperClass	Metric Set 1	Metric Set 2	Brief Description	
			Metric Set 1	Metric Set 2
temporal	TI-max1	TI-max2	Peak latency (ms) w.r.t. time-zero marker	Peak latency (ms) w.r.t pattern onset
spatial	IN-LOCC	IN-O1	Intensity over LOCC scalp region	Intensity at O1electrode
spatial	IN-ROCC	IN-O2	Intensity over ROCC scalp region	Intensity at O2 electrode
spatial	IN-LPAR	IN-C3	Intensity over LPAR scalp region	Intensity at C3 electrode
spatial	IN-RPAR	IN-C4	Intensity over RPAR scalp region	Intensity at C4 electrode
spatial	IN-LPTEM	IN-T7	Intensity over LPTEM scalp region	Intensity at T7 electrode
spatial	IN-RPTEM	IN-T8	Intensity over RPTEM scalp region	Intensity at T8 electrode
spatial	IN-LATEM	IN-F7	Intensity over LATEM scalp region	Intensity at F7 electrode
spatial	IN-RATEM	IN-F8	Intensity over RATEM scalp region	Intensity at F8 electrode
spatial	IN-LORB	IN-Fp1	Intensity over LORB scalp region	Intensity at Fp1 electrode
spatial	IN-RORB	IN-Fp2	Intensity over RORB scalp region	Intensity at Fp2 electrode
spatial	IN-LFRON	IN-F3	Intensity over LFRON scalp region	Intensity at F3 electrode
spatial	IN-RFRON	IN-F4	Intensity over RFRON scalp region	Intensity at F4 electrode

Table 1. Two alternative sets of summary spatial and temporal metrics

ONTOLOGY-BASED MAPPING OF ALTERNATIVE METRICS

The input to the data mining consisted of spatial and temporal metrics for each ERP observation (n=80) & each metric (n=13). Figure 3 illustrates the workflow of the proposed approach:

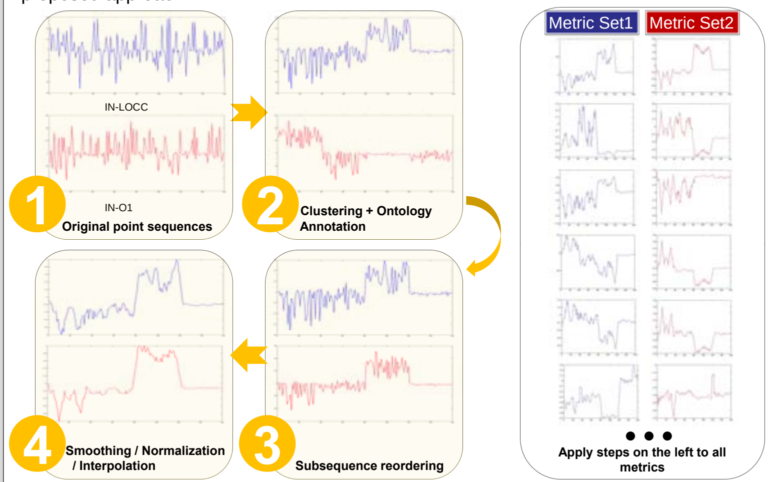

Figure 3. Workflow of ontology-based mapping discovery

- Step 1: starts from two point sequences representing two metrics.
- Step 2: identify meaningful grouping of the points:
  - Clustering: groups metric values by assigning clustering labels in a data-driven way.
  - Ontology annotation: ontologies are used to annotate clusters.
- Step 3: align clusters (subsequences) using annotated pattern labels
- Step 4: the following pre-processing steps are carried out:
  - Normalization: scales sequence values to unit range so that two sequences can be compared on a common basis.
  - Smoothing: the moving average method is used to reduce within cluster invariance.
  - Interpolation: interpolates the sequences if the number of points is different.

MAPPING RESULTS

Table 2. mapping result from cross-spatial join

- Experiments were performed on two simulated datasets, each using both metric sets.
- Cross spatial join: calculate the Euclidean distance between all pairs of metric sequence curves. The smaller the value of the distance is, the closer the two curves are.
- Applying heuristics: the following two heuristics are applied to selected the values that indicates most similarity between curves/metrics.
- 1-to-1 matching: assume a 1-to-1 matching relationship across the metric set.
- Global Minimum: Selected cells has the minimum sum.
- Ensemble result: Using the ensemble with majority vote can enhance the overall performance including both precision and recall.

SUMMARY & CONCLUSIONS

- We describe a framework for mapping cross-laboratory ERP data. Contributions include:
  - Using ontology annotation as a semantic bridge between heterogeneous ERP data.
  - Recasting the metric mapping problem to the sequence similarity search problem
- Results suggest that the proposed method is accurate and robust:
  - The ensemble method achieved an performance of 76% precision and 100% recall.
  - The mapping result is consistent throughout multiple samples of random ordering.

REFERENCES

[1] Frishkoff, G., Frank, R., Rong, J., Dou, D., Dien, J., & Halderman, L. (2007). A framework to support automated ERP pattern classification and labeling. Computational Intelligence and Neuroscience, vol. 2007, Article ID 14567, 13 pages.

[2] Frishkoff, G., & LePendu, P., Dou, D., . (2009), Development of NeuroElectroMagnetic Ontologies (NEMO): Ontology-based tools for representation and integration of event-related brain potentials, Proceedings of the International Conference on Biomedical Ontologies (ICBO-09), Buffalo, NY.

[3] Dou, D., Frishkoff, G., Rong, J., Frank, R., Malony, A., & Tucker, D. (2007). Development of NeuroElectroMagnetic Ontologies (NEMO): A framework for mining brain wave ontologies, Proceedings of the Thirteenth International Conference on Knowledge Discovery and Data Mining (KDD2007), pp. 270-279, San Jose, CA.

ACKNOWLEDGMENTS

This work was supported by an R01 grant to the University of Oregon by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) at the National Institutes of Health (NIH), Award #R01 EB007684-01. We thank our NEMO consortium for helpful comments and ongoing support of this project.