Neurocomputing 92 (2012) 156–169

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Sharing and integration of cognitive neuroscience data: Metric and pattern matching across heterogeneous ERP datasets

Haishan Liu [a,*], Gwen Frishkoff [b], Robert Frank [c], Dejing Dou [a]

[a] Department of Computer & Information Science, University of Oregon, Eugene, OR 97403, United States
[b] Department of Psychology & Neuroscience Institute, Georgia State University, Atlanta, GA 30303, United States
[c] NeuroInformatics Center, University of Oregon, Eugene, OR 97403, United States

ARTICLE INFO

Available online 13 March 2012

Keywords:
Schema matching
Cluster comparison
Assignment problem
Hungarian method
Neuroscience
Electrophysiology

ABSTRACT

In the present paper, we use data mining methods to address two challenges in the sharing and integration of data from electrophysiological (ERP) studies of human brain function. The first challenge, *ERP metric matching*, is to identify correspondences among distinct summary features ("metrics") in ERP datasets from different research labs. The second challenge, *ERP pattern matching*, is to align the ERP patterns or "components" in these datasets. We address both challenges within a unified framework. The utility of this framework is illustrated in a series of experiments using ERP datasets that are designed to simulate heterogeneities from three sources: (a) *different groups of subjects* with distinct simulated patterns of brain activity, (b) *different measurement methods*, i.e, alternative spatial and temporal metrics, and (c) *different patterns*, reflecting the use of alternative pattern analysis techniques. Unlike real ERP data, the simulated data are derived from known source patterns, providing a gold standard for evaluation of the proposed matching methods. Using this approach, we demonstrate that the proposed method outperforms well-known existing methods, because it utilizes cluster-based structure and thus achieves finer-grained representation of the multidimensional (spatial and temporal) attributes of ERP data.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Over the last few decades, neuroscience has witnessed an explosion of methods for measurement of human brain function, including high-density (multi-sensor) electroencephalography (EEG) and event-related potentials (ERP), or so-called "brainwave" methods. In comparison with other neuroimaging techniques, the ERP method has several advantages: it is completely noninvasive and, unlike fMRI (which measures blood flow), it is an inexpensive, portable, and direct measure of neuronal activity. Moreover, it has excellent (millisecond) temporal resolution, which is critical for accurate representation of neural activity. Furthermore, ERP studies have given rise to many complex neural patterns that can be used to predict human behavior and to detect clinically relevant deviations in behavior, cognition, and neural function [1,2]. Dozens of these patterns have been proposed over the past several decades. Yet there is remarkably little agreement in how these patterns should be identified and described. Further, tens of thousands of large and information-rich datasets have been collected and analyzed. Yet there are few (arguably, no) quantitative comparisons ("meta-analyses") of ERP data from different studies. Given the unique importance of ERP research in human neuroscience, this lack of integration may be the central obstacle to a robust science of human behavior and brain function.

To address these challenges, we have designed a system called Neural ElectroMagnetic Ontologies, or NEMO [3–6]. NEMO includes a suite of computational methods and workflows that are built around formal ontologies (description logic representations for the ERP domain) and can be used to facilitate ERP data sharing, analysis, and integration.

In the present paper, we introduce a new component of the NEMO workflow, which uses data mining methods to address two key problems—what we term the *ERP metric matching* and *ERP pattern matching* problems. In both cases, our goal is to align variables across multiple, heterogeneous ERP datasets. This would provide a data-driven alternative to top-down (knowledge-driven) methods, such as advocating the use of restricted methods for analysis, or a controlled vocabulary for data annotation. While these top-down approaches are of considerable value [5,6], we believe that data-driven approaches may provide a complementary approach that can lead to new insights into complex ERP data structures.

The remainder of Section 1 describes the ERP metric and pattern matching problems and summarizes our approach to these two problems. Section 2 presents a theoretic framework, along with a

* Corresponding author.
E-mail address: ahoyleo@cs.uoregon.edu (H. Liu).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2012.01.028

description of our approach to the two matching problems. In Section 3, we describe two case studies in which we used our approach to align different variables (metrics and patterns) across simulated ERP data. These data were designed to mimic three sources of variability that are present in real ERP data. Section 4 presents results from the two case studies. Section 5 compares the proposed method with existing methods, summarizes some assumptions and limitations of our study, and suggests some directions for future work. Finally, Section 6 summarizes the contributions of the present work.

1.1. ERP metric matching problem

ERP patterns are characterized by three features: time course (e.g., early or late), polarity (positive or negative), and scalp distribution, or topography [4,7,8]. For example, the visual-evoked ''P100 pattern'' (Fig. 1A) has a peak latency of approximately 100 ms (Fig. 1B) and is positive over occipital areas of the scalp (Fig. 1C), reflecting generators in visual regions of the cerebral cortex.

Researchers use a variety of metrics to describe the three features. These metrics reflect different ways that temporal and spatial features of ERPs can be operationally defined. For example, one research group might use a measure of peak latency (time of maximum amplitude) to summarize the timecourse of the ''P100'' pattern in a visual object recognition experiment [9,10], while another group might use measures of pattern onset and offset to operationalize the concept of latency for the same dataset.

The use of different metrics has a long history in the ERP research. While these diverse practices present a nuisance for data sharing and integration, there are reasons to embrace this heterogeneity, since different metrics may yield distinct and complementary insights [11]. The challenge then becomes how to find valid correspondences between these metrics. In the previous work, we have described top-down (knowledge-driven) methods, that is, annotation of data using a formal ontology [4–7]. This approach minimizes heterogeneity that arises from the use of different labels (e.g., ''latency'' vs. ''peak latency'' for time of maximal amplitude). It does not, however, address heterogeneities that reflect different operational definitions of time (e.g., peak latency vs. duration of a pattern), as described above. For this reason, we have also explored the use of bottom-up (data-driven) methods [11] to align different metrics across ERP datasets. In the present paper, we extend our bottom-up approach by developing and testing a more general formulation of the metric matching problem. Specifically, we view metric matching as an assignment problem and articulate a more general solution that can also be used to address a second problem—that of ERP pattern matching.

1.2. ERP pattern matching problem

The ERP pattern matching problem is the problem of finding correspondences among ERP patterns from different datasets. This problem is challenging for several reasons. The most trivial reason is that authors use a variety of labels to refer to the same pattern [4,7], just as they use different names for the same or similar metrics. This issue is readily addressed by the use of a standard ontology (or controlled vocabulary), although there is no guarantee that such a would be adopted by all research labs. The second reason is related to the metric matching problem (Section 1.1): when two different measures are used to characterize the timecourse of a pattern, they may capture subtly different views of the same data. Accordingly, they may introduce additional variability into the pattern matching equation. Finally, the most profound challenge is a consequence of the physics and physiology of signal generation: scalp-measured ERPs reflect a complex and unknown mixture of latent patterns. The reason is that neuroelectric signals are generated in cortex and are propagated to the scalp surface. Moreover, at each moment, multiple regions of cortex are co-active. Thus, at every timepoint and at every point in the measurement (i.e., scalp) space, a pattern in the measured data actually corresponds to multiple overlapping patterns, that is, different underlying sources. This overlap or ''superposition'' is exacerbated by volume conduction of these signals through the resistive skull.

Given these complexities, ERP researchers have adopted a variety of solutions for identification and extraction of ERP patterns (e.g., [1,2]). It can therefore be hard to compare results from different studies, even when the same experimental stimuli and task are used. Nonetheless, alternative analysis methods, like alternative metrics, may provide different, and equally informative views, of the ''same'' data. Thus, we propose to embrace this heterogeneity, rather than forcing researchers to use a restricted set of solutions for data analysis. As a consequence, our approach to data integration will require pattern matching, as well as metric matching, across different ERP datasets. Moreover, this matching should ideally be robust to differences among patterns that arise from the use of alternative pattern analysis methods.

1.3. Study goals and hypotheses

In this paper, we address the ERP metric and pattern matching problems by transforming them into two more general problems,

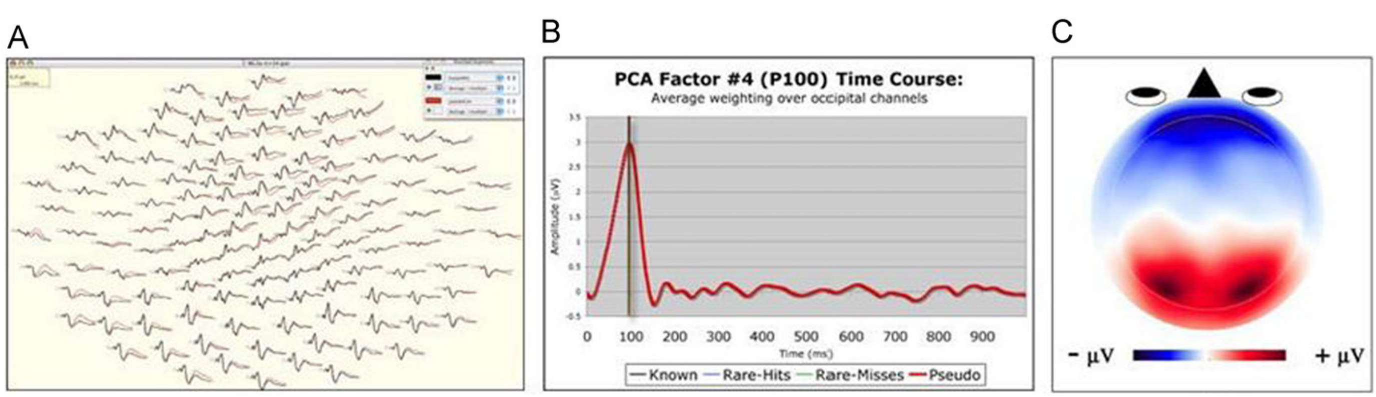

Fig. 1. (A) 128-Channel EEG waveplot; positive voltage plotted up. Black, response to words; red, response to non-words. (B) Time course of P100 pattern for same dataset, extracted using Principal Components Analysis. (C) Topography of P100 factor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Download English Version:

https://daneshyari.com/en/article/412516

Download Persian Version:

https://daneshyari.com/article/412516

Daneshyari.com