**EVENT ABSTRACT**

A Big-Data Approach to Automated EEG Labeling

Iyad Obeid[1*], Amir Harati[1], Mercedes Jacobson[2] and Joseph Picone[1]

[1] Electrical and Computer Engineering, Temple University, USA
[2] Temple University Hospital, USA

Electroencephalograms (EEGs) are valuable indicators of neural activity, both because of their non-invasive and inexpensive nature, and because a wealth of prior knowledge exists on their interpretation. Software for automatically reading EEGs has long been a focus of neuroinformatics research, since such a tool would be a boon to neuroscientists studying brain function as well as to neurologists who must manually scan hours of patient recordings. Prior research has relied either on heuristic rules for signal interpretation or on pattern recognition algorithms trained on insufficient datasets. Owing to the variability and complexity of neural function, both approaches are fundamentally limited, and resulting tools have failed to be transformative to both clinicians and researchers. In response, we have created a new data-rich EEG resource by amassing archival clinical EEG data recorded Temple University Hospital over the past decade. The resulting data corpus (TUH-EEG) comprises some 22,000 EEG records from approximately 15,000 unique patients, and includes medical histories and clinical diagnoses along with the raw EEG traces. Data have been de-identified appropriately, and all work has been approved by the Temple University IRB. The size and scope of the TUH-EEG dataset is enabling us to apply a new generation of machine learning technology based on deep learning. This technology automatically self-organizes knowledge in a data-driven manner and learns to emulate a physician's decision-making process. We are combining deep learning with unsupervised training so that detailed transcriptions of the data are not required. Performance of unsupervised training on vast amounts of data has recently been shown to approach or even exceed supervised training on much less data, giving rise to the notion of big data – learning from vast archives of noisy, poorly transcribed data. We have validated this approach on the CHB-MIT scalp EEG database and have achieved a 94% seizure detection rate, which compares favorably with state of the art on this task. We are presently using unsupervised deep learning on the TUH-EEG corpus to train a system that can differentiate between six so-called signal primitives: (1) focal epileptiform, (2) general epileptiform, (3) focal abnormal, (4) general abnormal, (5) artifacts, (6) background. Once these primitives can be reliably detected, they can be used to assess the presence of higher-level phenomena specific to certain disease states or conditions. We are implementing an unsupervised training technique in which we iteratively annotate the data using the previous iteration of the technology. This is the approach we believe will be most promising for TUH-EEG since we do not have access to manually transcribed labels. It will not only demonstrate our ability to learn from data automatically, but also provide timealigned marks for physicians to review. Preliminary results suggest a primitive sensitivity of 74% with a false alarm rate of 0.6/session; our goal for operational performance is 95% sensitivity.

Figure 1
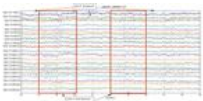
Acknowledgements

Manage Abstracts

My Submissions

  Submitted

  In Review

  Reviewed

  Accepted

  Published

  Rejected

  All Abstracts

Temple University's College of Engineering and Office of the Senior Vice-Provost for Research, the University City Science Center (QED Award No. S1313) and the NSF (Grant No. CNS-09-58854, Grant No. CNS-1305190).

Conflict of Interest

Portions of this were developed under the auspices of the University City Science Center's QED program that provides business development support for academic researchers developing early-stage healthcare IT technologies.