Journal of Neuroscience Methods 274 (2016) 179–190

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Rapid annotation of interictal epileptiform discharges via template matching under Dynamic Time Warping

J. Jing[a,*], J. Dauwels[a], T. Rakthanmanon[b], E. Keogh[c], S.S. Cash[d], M.B. Westover[d]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[b] Department of Computer Engineering, Kasetsart University, Thailand
[c] Department of Computer Science and Engineering, University of California, Riverside, CA, USA
[d] Department of Neurology, Massachusetts General Hospital, Harvard Medical School, MA, USA

HIGHLIGHTS

- An integrated system *NeuroBrowser* is proposed for EEG review and rapid annotation.
- *NeuroBrowser* lies on ultrafast template matching to accelerate the task of annotation.
- *NeuroBrowser* saves EEG experts approximately 70% on average in time efficiency.

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form 26 February 2016
Accepted 29 February 2016
Available online 2 March 2016

Keywords:
EEG
Interictal discharges
Spikes
Graphical user interface
Rapid annotation
Template matching
Dynamic Time Warping

ABSTRACT

Background: EEG interpretation relies on experts who are in short supply. There is a great need for automated pattern recognition systems to assist with interpretation. However, attempts to develop such systems have been limited by insufficient expert-annotated data. To address these issues, we developed a system named *NeuroBrowser* for EEG review and rapid waveform annotation.
New methods: At the core of *NeuroBrowser* lies on ultrafast template matching under Dynamic Time Warping, which substantially accelerates the task of annotation.
Results: Our results demonstrate that *NeuroBrowser* can reduce the time required for annotation of interictal epileptiform discharges by EEG experts by 20–90%, with an average of approximately 70%.
Comparison with existing method(s): In comparison with conventional manual EEG annotation, *NeuroBrowser* is able to save EEG experts approximately 70% on average of the time spent in annotating interictal epileptiform discharges. We have already extracted 19,000+ interictal epileptiform discharges from 100 patient EEG recordings. To our knowledge this represents the largest annotated database of interictal epileptiform discharges in existence.
Conclusion: *NeuroBrowser* is an integrated system for rapid waveform annotation. While the algorithm is currently tailored to annotation of interictal epileptiform discharges in scalp EEG recordings, the concepts can be easily generalized to other waveforms and signal types.

1. Introduction

Epilepsy refers to a group of chronic brain disorders characterized by recurrent unprovoked seizures. Epilepsy affects approximately 65 million people worldwide (Epilepsy Foundation of America). Electroencephalography (EEG) is an electrophysiological monitoring method to record the electrical activity of the brain, measuring voltage fluctuations resulting from ionic current within the neurons of the brain (Niedermeyer and da Silva, 2005). Both excitatory postsynaptic potentials and inhibitory postsynaptic potentials in cortical pyramidal cells contribute to the synaptic activity recorded as EEG (Olejniczak, 2006). EEG provides a continuous measure of cortical function with excellent temporal resolution.

Significant efforts (manpower, money, and time) are spent on interpreting EEG data for clinical purposes. Approximately 10–25 million EEG tests are performed annually worldwide (Encyclopedia of Surgery). The duration of EEG recordings ranges from 30 min to

* Corresponding author at: Research Staff Office II, S2-B2c-118, School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639789, Singapore.
E-mail addresses: jingjin@ntu.edu.sg (J. Jing), justin@dauwels.com (J. Dauwels), thanawin.r@ku.ac.th (T. Rakthanmanon), eamonn@cs.ucr.edu (E. Keogh), scash@mgh.harvard.edu (S.S. Cash), mwestover@mgh.harvard.edu (M.B. Westover).

http://dx.doi.org/10.1016/j.jneumeth.2016.02.025
0165-0270/© 2016 Published by Elsevier B.V.

Table 1
Spike detection methods.

Literature	Patient number	Spike count
Witte et al. (1991)	1	50
Gabor and Seyal (1992)	5	752
Gotman and Wang (1992)	20	Unknown
Hostetler et al. (1992)	5	1393
Sankar and Natour (1992)	11	Unknown
Pietilä et al. (1994)	6	Unknown
Webber et al. (1994)	10	927
Senhadji et al. (1995)	1?[b]	982
Feucht et al. (1997)	3	1509
Park et al. (1997)	32	n.a.[c]
Özdamar and Kalayci (1998)	5	n.a.[c]
Dümpelmann and Elger (1999)	7	2329
Hellmann (1999)	10	n.a.[c]
Ramabhadran et al. (1999)	6/18[a]	?[b]/982[a]
Wilson et al. (1999)	50	1952
Black et al. (2000)	521	Unknown
Goelz et al. (2000)	11	298
Acır and Güzeliş (2004)	18/7[a]	Unknown/139[a]
Xu et al. (2007)	12	957
Oikonomou et al. (2007)	13	333
Ji et al. (2011)	17	>780
Liu et al. (2013)	12	142
Lodder and van Putten (2014)	8	2973

[a] *N*/*N* format is applied when both training and test sets exist.
[b] A "?" after a value denotes uncertainty.
[c] The "n.a." values indicate that the data was presented in a non-standard manner that precluded comparison.

several weeks. While in current clinical practice, visual inspection and manual annotation are still the gold standard for interpreting EEG, this process is tedious and ultimately subjective. For instance, the agreement rate for interictal discharges has been found as low as 60% between electroencephalographers for certain cases (Wilson and Emerson, 2002). Moreover, experienced electroencephalographers are in short supply (Racette et al., 2014). Therefore, a great need exists for automated systems for EEG interpretation.

The finding of primary importance for the diagnosis for epilepsy is the presence of interictal discharges, encompassing spikes, polyspikes, sharp waves and spike-wave complexes. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) describes spikes as a subcategory of "epileptiform pattern", in turn defined as "distinctive waves or complexes, distinguished from background activity, and resembling those recorded in a proportion of human subjects suffering from epileptic disorders..." (Chatrian et al., 1974). Spikes are the key diagnostic biomarker for epilepsy. The presence of spikes predicts seizure recurrence (van Donselaar et al., 1992), and allows a physician to make a confident diagnosis of epilepsy and to prescribe appropriate treatment (van Donselaar et al., 1992; Fountain and Freeman, 2006; Pillai and Sperling, 2006). In practice, physicians detect spikes by visually inspecting 10–20 second-long of an EEG signal at a time. Spike detection is difficult, due to: (i) the wide variety of morphologies of spikes (Fig. 1), and (ii) the similarity of spikes to waves that are part of the normal background activity and to artifacts, e.g., potentials from muscle, eyes, and the heart.

Automated spike detection would be faster, less expensive, more objective, and potentially more accurate. Automated spike detection would enable wider availability of EEG diagnostics and more rapid referral to qualified physicians who can provide further medical investigation and interventions. However, spikes are difficult to detect in an automated manner due to the large variability of spike waveforms within and between patients among other factors. As illustrated in Table 1, attempts have been made to create automatic spike detection systems. Unfortunately, these methods have not been validated on a large dataset and consequently are not universally accepted. One of the most critical hurdles to developing an effective algorithm for spike detection is the lack of a sufficiently large database of expert-annotated spike waveforms.

The brute-force approach to generating such a database is to manually annotate numerous EEG records. However, exhaustive manual annotation of spikes is prohibitively time-consuming, especially for EEG recordings with large numbers of spikes (up to thousands per hour). The time and labor required severely limits the willingness of EEG experts to help establish a large database of annotated spikes. At present, no technology exists to enable rapid waveform annotation in EEG recordings.

In this paper, we describe a new system that dramatically accelerates the process of acquiring expert-annotated EEG records. This system, named *NeuroBrowser*, includes a graphical user interface designed for EEG review and rapid waveform annotation. The algorithm underlying *NeuroBrowser* is based on the observation that, within the same patient, spikes typically share a similar morphology (as shown in Fig. 2). With suitable choice of similarity measure and spike templates, it is possible to extract many more similar candidate waveforms from the same EEG record in less time. Rather than annotating one spike at the time, groups of spikes (typically 10–100 spikes) can be annotated by template matching, accelerating the annotation process.

NeuroBrowser includes a fully functional EEG viewer plus a custom-built algorithm for template matching to enable rapid waveform annotation. In earlier work (Jing et al., 2014), we applied the Euclidean distance (ED) as the similarity measure. However, the ED is sensitive to small variations in waveforms, and often fails to "match" waveforms that share strong morphological similarity, limiting annotation speed.

In this study, we consider a more powerful similarity measure: Dynamic Time Warping (DTW) (Keogh and Ratanamahatana, 2005; Müller, 2007). It is a distance measure that permits non-linear distortion so as to achieve better waveform alignments. Specifically, a modification of the *Trillion* algorithm from the UCR (University of California, Riverside) suite (Rakthanmanon et al., 2012) is employed here for rapid similarity search under DTW. Annotation speed is improved drastically as a result. A related system (hereafter referred to collectively simply as the "*Self-Adapting System*") has been proposed in (Lodder and van Putten, 2014), which also integrates the concept of template matching together with user assessment to iteratively refine the detection results while populating the database of templates. The *Self-Adapting System* is not dependent on user-selected templates in the actual application, while for *NeuroBrowser*, the user does need to define a specific template for the EEG recording at hand. Rather than a specific template, the detection in *Self-Adapting System* is made based on the similarities in detected transients in naive EEGs that are subsequently compared to templates collected from other recordings. During user assessment, the system only shows the top 10 candidate spikes to the user that have the best agreement with any of the templates in the database. After that, the system learns from the user, as he/she agrees, disagrees or is uncertain about the 10 candidates presented. In addition, correlation is used as the similarity measure in *Self-Adapting System*, and the candidate spikes are assessed by the user one at a time. By contrast, we achieve much faster assessment by a variety of strategies, including preprocessing, DTW, and clustering of spikes.

Our experimental results show that *NeuroBrowser* is able to save EEG experts an average of approximately 70% of the time spent on annotating spikes, relative to conventional unassisted annotation. A database of 19,000+ spikes from 100 patient EEG recordings is constructed with *NeuroBrowser*. Each EEG was cross-annotated by 3 neurologists at Massachusetts General Hospital (MGH), and we only consider spikes that are accepted by at least 2 neurologists. To our knowledge this represents the largest expert-annotated