Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 50 (2018) 36–51

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Measuring concept semantic relatedness through common spatial pattern feature extraction on EEG signals

Action editor: Yi Zeng

Hiram Calvo *, José Luis Paredes, Jesús Figueroa-Nazuno

Centro de Investigación en Computación, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz e/M.O. Mendizábal s/n, Nva. Ind. Vallejo, 07738, Mexico City, Mexico

Received 20 August 2017; received in revised form 20 February 2018; accepted 13 March 2018
Available online 24 March 2018

Abstract

We study the semantic relationship between pairs of nouns of concrete objects such as "HORSE - SHEEP" and "SWING - MELON" and how this relationship activity is reflected in EEG signals. We collected 18 sets of EEG records; each set containing 150 events of stimulation. In this work we focus on feature extraction algorithms. Particularly, we highlight Common Spatial Pattern (CSP) as a method of feature extraction. Based on these latter, different classifiers were trained in order to associate a set of signals to a previously learned human answer, pertaining to two classes: semantically related, or not semantically related. The results of classification accuracy were evaluated comparing with other four methods of feature extraction, and using classification algorithms from five different families. In all cases, classification accuracy was benefited from using CSP instead of FDTW, LPC, PCA or ICA for feature extraction. Particularly with the combination CSP-Naïve Bayes we obtained the best average precision of 84.63%.

Keywords: Semantic concept similarity; Electroencephalogram; EEG; Common spatial pattern; CSP; Signal classification; Oddball; Emotiv EPOC

1. Introduction

The study of how human brain works has become a multidisciplinary task integrated by several research areas such as mathematics, biology, and psychology, among many others. This has led to a new knowledge area known as neuroscience. Roughly over three decades, computer science has contributed with methods and tools for analysis and visualization of information obtained through techniques to study the human brain, for example: functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) (Miyawaki et al., 2008; Chan, Halgren, Marinkovic, & Cash, 2011) and electroencephalography (EEG) (Sanei & Chambers, 2007; Tong & Thakor, 2009). Language is one of the main fields of study of neuroscience, with recent findings that suggest that the area of Broca and Wernicke is not the only area responsible for language activity. However, after these works, new questions have arisen related to how are concepts recognized, stored and organized in the brain.

Recently, the study of semantic representation of concepts in the brain has received particular interest with studies such as Maguire, Brier, and Ferree (2010), who use PCA, ANOVA and average Event-Related Potential (ERP) for finding qualitative differences in processing taxonomic versus semantic relationships. Murphy et al. (2011) sample signals in 64 regions. They use PCA, ICA and CSP techniques to decode semantic categories from EEG. They classify using a SVM. Geuze, van Gerven, Farquhar, and Desain (2013) analyze the average ERP component and

* Corresponding author.
E-mail addresses: hcalvo@cic.ipn.mx (H. Calvo), jparedes_a13@cic.ipn.mx (J.L. Paredes), jfn@cic.ipn.mx (J. Figueroa-Nazuno).

https://doi.org/10.1016/j.cogsys.2018.03.004
1389-0417/

use a logistic regression classifier. Chan et al. (2011) compare different techniques for acquiring information from brain activity: EEG, MEG and EEG/MEG. They use a SVM classifier to decode word and category representations.

Our work is motivated by the work in cognoscitive semantics in its early stages by Collins and Quillian (1969), who show that semantic organization in memory is done hierarchically. Following them, we propose to analyze the relationship between two concepts in EEG signals. For this purpose, we (i) developed an experimental model to study the semantic relationship between two concepts, that can be related or not; for example, "BABY-CRADDLE" and "TELEPHONE-FISH". (ii) implemented a tool to carry out the experimental development autonomously, considering that the subject to be tested must be in control of the experiment, the EEG should be registered continuously, and each one of the events should contain detailed information; and (iii) experimented with several preprocessing tools of the segments composing the samples. This preprocessing (prior to their classification) includes filtering, and feature extraction.

In the Section 2 we will describe works related to our proposal. In Section 3 we will describe how we acquired the EEG signals, while feature extraction and classification is explained in Section 4, and then results and discussion are presented in Section 5. Finally, conclusions are drawn in Section 6.

2. Related work

The work of Quillian (1967) is one of the earliest works related to investigating how concepts are stored in our brains. Quillian proposed a model of semantic storage in a hierarchical form, known today as a semantic network. This work was subsequently taken up by Collins and Quillian (1969) in which they demonstrate experimentally that there is a semantic distance between concepts based on reaction times measuring.

Two years later, Meyer and Schvaneveldt (1971) used this principle to study the existence of semantic relationship between words such as "nurse-doctor" and "bread-doctor" respectively. Thanks to reaction times, concept relatedness was measured. A decade later, works on modeling internal semantic representations in the brain through EEG began, with works such as Boddy and Weinberg (1981). Kutas and Hillyard (1980a, 1980b, 1980c, 1983, 1984, 1989) performed experiments on ERP and language, specifically on subjects like semantics and language coherence.

2.1. Experiments on ERP and language

Event-Related Potential (ERP) is one of the main objects of study at scientific level. For electroencephalography is the sum of a large number of action potentials, which originate from the stimulation either by external events such as sensory stimulation or internal as the result of cognitive processes or actions by the motor system. In most cases this type of activity differs very little from the base EEG signal and has slight variations in amplitude (1 and 30 μV); latency is defined as the time interval (ms) in which the response to the stimulus is carried out and may vary between 150 and 600 ms (Kutas & Hillyard, 1980c).

Language, as well as some other cognitive processes such as logic, the recognition of objects and patterns, to name a few, requires the implementation of secondary mechanisms. According to Meyer and Schvaneveldt (1971) within this type of activity, the interaction of short and medium term memory mechanisms is necessary. Language is one of the most complex processes, since it intervenes in tasks such as the evocation of memories and the logical structure of language. In most cases, these processes develop as a joint task between several areas located in both hemispheres of the brain; several works (v.gr. Bouaffre & Faita-Ainseba, 2007; Federmeier & Kutas, 1999) have discussed the existence of a hemispheric difference. Particularly, while carrying out a semantic process, the left hemisphere has a predominant role. In addition to the hemispheric difference, the language process is carried out mainly in the frontal and temporal regions of the human brain. In Section 5.2 we study the interaction of the ERP components in each hemisphere and particular regions of the brain.

Several works (Fodor, 1983; Friederici, 2002; Morton, 1969; Shallice, 1988) show that language is processed as a sequence of sub-processes: phonologic, lexical-syntactic, semantic, and deep syntactic analysis; Generally, this occurs in a span of between 300 and 600 ms after the stimulation process, as shown in Fig. 1 (Pulvermüller, Shtyrov, & Hauk, 2009).

During the study of language processing and ERP, two components are found: P300 (Positive at 300 ms) and N400 (Negative at 400 ms). These two components also play an important role for classification.

2.1.1. Component P300

This component can be found in all cognoscitive processes as a direct result of stimulation, as described by Sutton, Braren, Zubin, and John (1965). The behavior of the P300 component is very characteristic in EEG signals, and can be easily identified, since it is the first large positive variation with respect to the base signal. Its amplitude is between 5 and 20 μV and usually occurs between 250 and 350 ms after the first stimulus is presented, although this window can be extended. The parameters between amplitude and latency are subject to endogenous and exogenous of the subject. The P300 component is associated with the selective attention process, which consists in selecting a desired object within a set of options.

2.1.2. Component N400

This component is closely related to the language process and was originally identified by Kutas and Hillyard (1980a, 1980b, 1980c, 1983, 1984, 1989); recent research