Lecture Notes in Computer Science 3696 (2005) 683-692 (Springer)

Early Detection of Alzheimer's Disease by Blind Source Separation and Bump Modeling of EEG Signals.

François Vialatte[1], Andrzej Cichocki[2], Gerard Dreyfus[1], Toshimitsu Musha[3], Rémi Gervais[4]

[1] ESPCI (ParisTech), Laboratoire d'Electronique (CNRS UMR 7084), 10 rue Vauquelin, 75005 Paris, France
francois.vialatte@espci.fr, gerard.dreyfus@espci.fr
[2] BSI RIKEN ABSP Lab 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
cia@brain.riken.jp
[3] Brain Functions Laboratory Inc., KSP Building E211, Sakado, Takatsu Kawasaki-shi, Kanagawa, 213-0012, Japan
musha@bfl.co.jp
[4] Equipe neurobiologie de la mémoire olfactive, Institut des Sciences Cognitives (UMR 5015 CNRS UCB), 67 Boulevard Pinel, 69675 Bron Cedex, France
gervais@isc.cnrs.fr

Abstract. The early detection Alzheimer's disease is an important challenge. Using blind source separation, wavelet time-frequency transforms and "bump modeling" of electro-encephalographic (EEG) recordings, a set of features describing the recordings of mildly impaired patients and of controls subject is built. Feature selection by the random probe method leads to the selection of a few reliable features, which are fed to a neural network classifier. This leads to a sizeable performance improvement over detection results previously published on the same data.

1 Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disorder. Since the number of individuals with AD is expected to increase in the near future, early diagnosis and effective treatment of AD are critical issues in dementia research [1]. Finding a method for early identification of patients who are to progress toward Alzheimer's disease, but do not exhibit clinical signs of AD at the time of the test, is thus an important challenge. Furthermore, an early detection method should be inexpensive, in order to allow mass screening of that disease [2]. Electroencephalography (EEG) is one of the most promising candidates in that respect.

The "at-risk" state is commonly referred to as Mild Cognitive Impairment (MCI) [3]. In [4], two data sets of EEG recordings, obtained in the course of a clinical study [5], were created:

- the MCI set, featuring 22 EEG recordings of patients matching the criteria of mild cognitive impairment, who developed AD two years later;
- a control set, featuring 38 recordings from age-matched family members of the patients.

In the present paper, we show that the combined use of blind source separation for dimensionality reduction, bump modeling for feature generation, and statistical feature selection, provides a substantial improvement over early AD detection results obtained previously [4] on the same data set.

2 Methods

2.1 Blind Source Separation for Signal Filtering

According to the currently prevailing view of EEG signal processing, that signal can be modeled as a linear mixture of a finite number of sources, with additive noise. Therefore, blind source separation techniques can be used advantageously for filtering EEG signals. In a previous investigation of the same data set [4], the AMUSE (Algorithm for Multiple Unknown Signals Extraction [6], [7], [8], [9]) algorithm, a blind source separation technique that relies on second-order statistics for spatio-temporal decorrelation, was used in order to select the five components of the signal that had the lowest linear predictability. In the present paper, that method was used as a baseline for assessing the efficiency of the detection obtained by the present method: a data base (hereinafter referred to as D_1) containing those five components was generated[1].

In addition, a second database (hereinafter referred to as D_2) was generated, containing the recordings of three electrodes[2] only, which were conjectured to be the most relevant for early AD detection.

2.2 Time-Frequency Maps and Bump Modeling for Feature Generation

In order to obtain a compact representation of the signals of D_1 and D_2, suitable for automatic discrimination of MCI patients from control individuals, the signals were first analyzed in the time-frequency domain by a wavelet transformation, and the resulting time-frequency maps were modeled by bumps, as described below.

2.2.1 Wavelet Transformation and Time-Frequency Map Generation

EEG signals were first transformed to time-frequency maps using wavelets. There is a wide variety of wavelets (see [11] for details). In the present study, complex Morlet wavelets [12] were used. Complex Morlet wavelets are appropriate for time-frequency analysis of electroencephalographic signals ([13], [14], [15],[16]). Complex Morlet wavelets of Gaussian shape in time (deviation σ_t) are defined as:

$$w(t) = A.\exp\left(-t^2 / 2\sigma_t^2\right).\exp\left(2i\pi ft\right) \qquad \textbf{(1)}$$

where σ_t and f are appropriately chosen parameters; they cannot be chosen independently, since the product $\sigma_t f$ determines the number of periods that are present in the wavelet. In the present investigation, the wavelet family defined by $2\pi\sigma_t f = 7$ was chosen, as described in [13].

The signals present in the D_1, and D_2 databases were wavelet-transformed in the frequency range 1.5 to 31.5 Hz, discretized in 0.25 Hz steps.

2.2.2 Bump Modeling

The bump modeling technique is a 2-dimensional generalization of the Gaussian mesa function modelling technique that was initially designed for one-dimensional signals (electrocardiogram analysis) [17]. In the present study, it was used for extracting information from the time-frequency maps (see [18] for a detailed description of the method). This method was successfully applied to the analysis of local field potential signals, gathered from electrophysiological (invasive) measurements [19]; the present paper reports the first application of bump modeling to surface EEG signals.

The purpose of this method is to approximate a time-frequency map as a sum of known elementary parameterized functions called bumps; therefore, the map is represented by the set of parameters of the bumps, which is a very compact encoding of the map, resulting to information compression rates ranging from one hundred to one thousand. The parameters of the bumps are computed in order to minimize the modelling error. The method is somewhat similar in spirit to the matching-pursuit type algorithms [20]. The algorithm performs the following steps on the maps (after normalization):

- (i) window the map in order to define the zones to be modelled,
- (ii) find the zone that contains the maximum amount of energy,
- (iii) adapt a bump to the selected zone, and withdraw it from the original map,
- (iv) if the amount of information E modelled by the bumps reaches a threshold, stop; else return to (iii).

Half ellipsoids were found to be the most appropriate bump functions for the present application. Each bump is described by 5 parameters: its location on the map (2 parameters), its amplitude (one parameter)

---

[1] The AMUSE algorithm was provided by the "ICALab" software package [10] in the present investigation

[2] The conventional FP1, FP2 and FPZ electrode sites.

and the lengths of its axes (2 parameters). Figure 1 shows a typical example of bump modeling of the time-frequency map of an EEG recording.

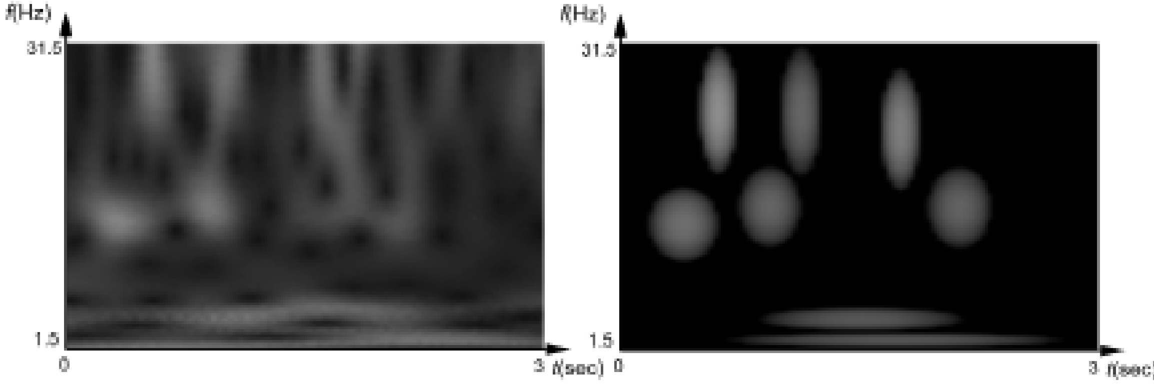

Fig. 1. Left: normalized time-frequency map of an EEG signal; right: bump modeling of the map

After completion of the modeling, the parameters of the bumps are candidate features for classification: thus, the number of candidate features for a given signal is 5 times the number of bumps required for describing its time-frequency map. Given the small size of the data set, feature selection is mandatory, in order to avoid overparameterization.

2.3 Feature Selection

After bump modeling, the signals under investigation are represented by the set of parameters that describe the bumps. Within that set, an even more compact representation was sought, based on expert knowledge on the frequency bands of interest. First, five bands were defined: θ (3.5-7.5 Hz), α_1 (7.5-9.5 Hz), α_2 (9.5-12.5 Hz), β_1 (12.5-17.5Hz) and β_2 (17.5-25Hz). Within each band, the following features were computed:

- F_1: the number of bumps in the band,
- F'_1: the number of high-amplitude bumps (normalized amplitude > 0.7) in the band
- F_2: the sum of the amplitudes of the bumps present in the band,
- F'_2: the sum of the amplitudes of the high-amplitude bumps present in the band,
- F_3: the maximal amplitude of the bumps present in the band.

Two groups of candidate features were defined: group A contains $\{F_1, F'_1, F_2, F'_2\}$ and group B contains $\{F_1, F'_1, F_3\}$.

Thus either 3 or 4 features were computed per band, depending on the group of features under consideration. Therefore, for database D_1 (5 time-frequency maps), the number of candidate features was either 75 or 100; for database D_2 (3 time-frequency maps), the number of candidate features was either 45 or 60. Since the number of candidate features was still too large given the number of examples in the data bases, feature selection was performed [21]. A selection algorithm based on the Gram-Schmidt orthogonal forward regression (OFR) algorithm [22] was applied to select features relevant to the classification. In order to choose the number of features to be kept, the random probe method [23] was applied. 100 "probes", i.e. realizations of random variables, were computed and appended to the feature set. A risk level is defined, which corresponds to the risk that a feature might be kept although, given the available data, it might be less relevant than the probe. At each step of the selection procedure, the following steps are performed:

- (i) obtain a candidate feature from OFR,
- (ii) compute the value of the cumulative distribution function of the rank of the probe for the rank of the candidate feature,
- (iii) if that value is smaller than the risk, select the feature and return to (i);
- (iv) else, discard the feature under consideration and terminate.

Since the database is small (22 examples of mildly impaired patients only) compared to the number of features to be tested, the ensemble feature ranking method [24] was used: 60 subsets are built by iteratively removing one example from the database. OFR is then applied to these subsets. The overall distribution of features, and the average number N_k of selected features is computed; finally, the N_k overall best features are selected. Accepting a percentage P of false positive variables (i.e. irrelevant variables that are wrongly selected), we obtained finally F features, ranked as the most efficient ones.

3 Results

Each dataset was used for training and validating a neural network classifier (multilayer perceptron model, see for instance [25]). The generalization performance was estimated using the leave-one-out cross-validation method [26]. The best results, shown in Table 1, were obtained with linear classifiers (zero hidden neuron).

The best results are obtained on data set D_1 (without component 4 found by the AMUSE algorithm), with group of candidate features A. Raw signals from D_2 also provided good results for group B. For comparison, the last line of the table reports results obtained with the same EEG recordings with a different representation.

Table 1. Number of subjects correctly and incorrectly classified by neural network models, using D_1 and D_2 datasets, depending on the feature group (A or B). Results were obtained using the leave-one-out cross-validation method, validation set results are presented below.

Datasets	Misclassified		Correctly classified %		
	MCI $N = 22$	Controls $N = 38$	MCI $N = 22$	Controls $N = 38$	All $N = 60$
D_1, A group → F =12, P = 12% Components 1,2,3 and 5+Bumps	2	2	91.0	94.7	93.3
D_1, A group → F = 6, P = 9% Components 1-5+Bumps	4	4	81.8	88.6	86.7
D_1, B group → F = 11, P = 10% Components 1-5+Bumps	2	3	91.0	92.1	91.7
D_2, A group → F = 4, P = 8% FP1, FP2 and FPZ + Bumps	7	7	68.2	81.6	76.7
D_2, B group → F = 12, P = 13% FP1, FP2 and FPZ + Bumps	2	4	91.0	89.5	90.0
Previous study, best results [5]	6	6	72.7	84.2	80.0

4 Discussion

In the present paper, we reported the first application of the bump modeling method to the automatic classification of EEG data. Bump modeling is a 2-dimensional extension of mesa function modeling, which was first used successfully for electrocardiographic data. For comparison purposes, the method was applied to EEG recordings that had been classified previously by other methods, resulting in a sizeable improvement, the overall correct classification rate being raised from 80% to 93%.

The task was the discrimination of EEG recordings of normal individuals from EEG recordings of patients who developed Alzheimer's disease two years later. Therefore, the present study provides exciting prospects for early mass detection of the disease. The method is very cheap as compared to FMRI, requiring only a 21-channnel EEG apparatus, and 20 seconds of artifact-free recording.

However, a full validation of the method requires investigating more extensive databases. Furthermore, there is presumably a lot of information present in the recordings that is not yet exploited, such as the dynamics of the bumps.

References

1. Jeong, J. (2004). EEG dynamics in patients with Alzheimer’s disease. Clinical Neurophysiology, 115:1490-1505
2. DeKosky, S.T., Marek, K. (2003). Looking backward to move forward: early detection of neurodegenerative disorders. Science, 302(5646):830-834.
3. Nestor, P.J., Scheltens, P., Hodges J.R. (2004). Advances in the early detection of Alzheimer’s disease. Nature Medicine, 10 Suppl.:S34-41. Review.
4. Cichocki A., Shishkin S.L., Muash T., Leonowicz Z., Asada T., Kurachi T (in press). EEG filtering based on blind source separation (BSS) for early detection of Alzheimer’s disease. Clinical Neurophysiology.
5. Musha, T., Asada, T., Yamashita, F., Kinoshita, T., Chen, Z., Matsuda, H., Masatake, U., Shankle, W.R. (2002). A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer’s disease. Clinical Neurophysiology, 113(7):1052-1058.
6. Cichocki, A., Amari, S. (2003). Adaptative blind signal and image processing: learning algorithms and applications. New York, NY: Wiley.
7. Tong, L., Soon, V., Huang, Y.F., Liu R. (1991). Indeterminacy and identifiability of blind identification. IEEE Transactions CAS, 38:499-509
8. Tong, L. Inouye, Y., Liu, R. (1993). Waveform-preserving blind estimation of multiple independent sources. IEEE Transactions on Signal Processing, 41(7):2461-2470.
9. Szupiluk, R., Cichocki, A. (2001). Blind signal separation using second order statistics. Proceedings of SPETO, 485-488.
10. Cichocki, A., Amari, S., Siwek, K., Tanaka, T., et al. ICALAB toolboxes [available online at http://www.bsp.brain.riken.jp/ICALAB]
11. Poularikas, A.D. (1996). The transforms and applications handbook. CRC Press.
12. Kronland-Martinet, R., Morlet, J., Grossmann, A. (1987). The wavelet Transform, in Expert Systems and Pattern Analysis, C.H Chen Edt, World Scientific, pp 97-126.
13. Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. Journal of Neuroscience, 16:4240-4249.
14. Ohara S., Crone N.E., Weiss N., Lenz F.A. (2004). Attention to a painful cutaneous laser stimulus modulates electrocorticographic event-related desychronization in humans. Clinical Neurophysiology, 115:1641-1652.
15. Caplan, J.B., Madsen, J.R., Raghavachari, S., Kahana, M.J. (2001). Distinct patterns of brain oscillations underlie two basic parameters of human maze learning . Journal of Neurophysiology, 86:368-380.
16. Düzel, E., Habib, R., Schott, B., Schoenfeld, A., Lobaugh, N., McIntosh, A.R., Scholz, M., Heinze, H.J. (2003). A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory. Neuroimage ,18:185-197.
17. Dubois, R. (2004). Application de nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie. Thèse de Doctorat de l'Université Pierre et Marie Curie - Paris VI. [Available from http://www.neurones.espci.fr/Francais.Docs/dossier_recherche/bibliographie/theses_soutenues.htm#ancre41560]
18. Vialatte, F., Martin, C., Dubois, R., Quenet, B., Gervais R., Dreyfus G. A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics. Neural networks (submitted).
19. Vialatte, F., Martin, C., Ravel, N., Quenet, B., Dreyfus, G., Gervais, R. (2003). Oscillatory activity, behaviour and memory, new approaches for LFP signal analysis. 35 th annual general meeting of the European brain and behaviour society, 17-20 September 2003, Barcelona, Spain – Acta Neurobiologiae Experimentalis, Vol. 63, supplement 2003.
20. Mallat, S., Zhang, Z. (1993). Matching pursuit with time-frequency dictionaries. IEEE Transactions on Signal Processing, 41(12):3397—3414
21. Guyon I., Elisseef A. (2003). An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3:1157-1182
22. Chen, S., Billings, S. A. and Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control, 50:1873-1896.
23. Jong K., Marchiori E., Sebag M. (2004) Ensemble Learning with Evolutionary Computation: Application to Feature Ranking. 8th International Conference on Parallel Problem Solving from Nature (PPSN). Lecture Notes in Computer Science 1133-1142. Springer.

24. Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y. (2003). Ranking a Random Feature for Variable and Feature Selection. Journal of Machine Learning Research, 3:1399-1414.
25. Haykin S. (1999). Neural Networks – a comprehensive foundation. 2° edition, Prentice Hall.
26. Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). Journal of the Royal Statistical Society B, 36:111–147.