Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition

Cheng-Wen Ko, Hsiao-Wen Chung *

Department of Electrical Engineering, National Taiwan University, Room 238, Taipei 10764, Taiwan

Accepted 22 October 1999

Abstract

Objective: Automatic detection of epileptic EEG spikes via an artificial neural network has been reported to be feasible using raw EEG data as input. This study re-investigated its suitability by further exploring the effects of data preparation on classification performance testing.

Methods: Six hundred EEG files (300 spikes and 300 non-spikes) taken from 20 patients were included in this study. Raw EEG data were sent to the neural network using the architecture reported to give best performance (30 input-layer and 6 hidden-layer neurons).

Results: Significantly larger weighting of the 10th input-layer neuron was found after training with prepared raw EEG data. The classification process was thus dominated by the peak location. Subsequent analysis showed that online spike detection with an erroneously trained network yielded an area less than 0.5 under the receiver-operating-characteristic curve, and hence performed inferiorly to random assignments. Networks trained and tested using the same unprepared EEG data achieved no better than about 87% true classification rate at equal sensitivity and specificity.

Conclusions: The high true classification rate reported previously is believed to be an artifact arising from erroneous data preparation and off-line validation. Spike detection using raw EEG data as input is unlikely to be feasible under current computer technology. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Spike detection; Artificial neural network; Raw EEG data; Data preparation

1. Introduction

An artificial neural network (ANN) is now regarded as a potentially effective approach to the automatic detection of EEG spikes since its first appearance (Gotman and Ives, 1996; Blum, 1998). An ANN has an attractive advantage in such applications, in that the recognition of epileptic discharges does not need a clear definition of spike morphology (Gabor and Seyal, 1992), which is certainly necessary in rule-based detection algorithms (Gotman 1985; Webber et al., 1993). Furthermore, previously "unseen" patterns could be recognized by the ANN with proper "training" (Webber et al., 1994). Most of the ANN spike detection algorithms reported to date employed parameterized EEG data as input, using various ways of feature extraction for a simplification of ANN computation (Gabor and Seyal, 1992; Webber et al., 1994; Wilson et al., 1999).

As a result, due to varieties in the network architecture or data source (e.g. channel numbers or displaying montages), a direct comparison of ANN performance is difficult at this stage (Webber et al., 1994).

Spike detection with ANN using raw EEG data (Özdamar et al., 1991; Yaylali et al., 1992) has the potential advantage of avoiding possible false classification arising from data omission in EEG parameterization. Such a method, however, has not appeared as a complete report in journal article form until the recent one by Özdamar and Kalayci (1998). Results from Özdamar’s study on 3600 EEG files from 5 patients, validated off-line, had shown a sensitivity of about 93% with overall true classification rate on the order of 85% (Özdamar and Kalayci, 1998). Therefore, high success rates seemed to be achievable in ANN-based spike-detection systems using raw EEG as input. Using a similar method for ANN spike detection with raw EEG data, Webber et al. performed a comparison on 10 patients and reported a sensitivity of 74 and 46% for parameterized and raw EEG data, respectively, at the threshold value corre-
sponding to equal sensitivity and selectivity (Webber et al., 1994). The results from Webber’s study, again using off-line validation, suggested that ANN employing parameterized input performed superiorly to that using raw EEG data. Although a different number of neurons from that in Özdamar’s study was used in Webber’s work, rendering a direct and objective comparison difficult, which data source is potentially suitable for use in automatic spike detection via ANN network remains a debatable issue warranting continuous evaluation.

In this article, a re-investigation of the suitability of spike detection with raw EEG data as ANN input was attempted. An ANN trained with error back-propagation (EBP), a very popular architecture of ANN used for automatic detection of EEG spikes (Özdamar et al., 1991; Gabor and Seyal, 1992; Yaylali et al., 1992; Webber et al., 1994; Özdamar and Kalayci, 1998), was chosen to perform the recognition task. The data preparation scheme employed in previous works (Webber et al., 1994; Özdamar and Kalayci, 1998) was then intentionally modified to observe their effects on classification sensitivity and specificity. The consequent implications on the possible extension to online recognition, and the inherent restrictions of ANN spike detection using raw EEG data as input, were also addressed.

2. Materials and methods

In an attempt to reproduce the results of spike detection on raw EEG data by Webber et al. and Özdamar and Kalayci, we collected a total of 600 single-channel interictal EEG files, each with 230 ms in duration, for the evaluation of ANN performance. These files, consisting of 300 spike and 300 non-spike waveforms, were taken from interictal EEG examinations recorded on 20 patients with epilepsy history. The sampling frequency was fixed at 128 Hz. All files were pre-selected and classified as spikes or non-spikes by one experienced neurologist. Upon the neurologist clicking on the peak of a spike waveform, the computer generated a file by extending from the peak 10 points to the front and 20 points to the end. Each file thus contained 30 points and was 230 ms in length. Non-spike waveforms were selected in a similar manner by the neurologist randomly clicking on places far from a spike in an EEG record. Thus the classification performed by the neurologist can be used as the standard to be compared with the ANN output. Note that the above procedure is the same as that used for off-line validation in previous reports (Webber et al., 1994; Özdamar and Kalayci, 1998), except that the spike files in the work by Webber et al. had 100 ms before and 300 ms after the waveform peak (Webber et al., 1994). The difference in duration is not expected to affect the results, for reasons that will be illustrated later.

The EBP ANN architecture as shown in Fig. 1 was used for spike identification. The network contained 30 input-layer neurons as needed for raw data input, and 6 neurons in the hidden layer. Each input-layer neuron receives the instantaneous amplitude and passes to the hidden layer with multiplication by weight \( W1(i,j) \), where the indices represent the source and the destination neurons, respectively.

**Fig. 1.** A popular three-layered ANN (input, hidden, and output layers), where each node in the hidden and output layers receives multiple data from its upper-layer neurons in the form of linear combinations. In this study, each input-layer neuron simply receives the instantaneous EEG amplitude and passes to the hidden layer with multiplication by weight \( W1(i,j) \).
spikes can quickly be identified as the data containing a large amplitude at the 10th point. In order to investigate this effect, we displayed the weights $W_{1(i,j)}$ after training as a $6 \times 30$ image to examine the relative weighting of the 10th input-layer neuron.

To further explore the effects of data preparation, we simulated the online spike detection by randomly varying the peak location of the spike EEG files used for performance testing. Here we intentionally challenged the ANN by using all 600 EEG files both for training and testing of ANN, with the only difference being that the spike files used for training had their peak placed at the 10th point, while the testing data were randomized in peak location. Likewise, the ANN performance was assessed by ROC analysis.

Finally, we looked into the ANN performance in case the preparation effects were to be eliminated by randomizing the peak locations of the training data. Similar to the above procedure, we presented the same 600 data sets to the ANN both for training and testing. As a well-trained ANN should theoretically give a 100% true classification rate in such a case, the performance obtained here should represent the maximum true classification rate under the chosen ANN architecture.

### 3. Results

Fig. 2 shows the $6 \times 30$ image of $W_{1(i,j)}$ trained using EEG files among which the spike files had their peaks located at the 10th point. It is clearly seen that several weights in the 10th column exhibit significantly higher values than all the other elements. This result indicates that the classification process is essentially dominated by the amplitude of the 10th data point. Clearly, this instantaneous amplitude should not be regarded as an important characteristic of an epileptic spike. In addition, as the dominance originated from a placement of the waveform peak location rather than from the EEG file duration, we infer that similar phenomena should also be observable in the work by Webber et al. (1994) which used a significantly longer file duration than ours.

Fig. 3 shows the ROC curves obtained for the ANN trained in similar manners, but with different methods of off-line performance testing. The uppermost curve (dotted line) shows the performance of the ANN trained and tested according to the methods reported by Özdamar and Kalayci and Webber et al.. The area under the ROC curve (0.9997) corresponds to about 97% at equal sensitivity and specificity, higher than, but in qualitative agreement with, the 93% sensitivity reported by Özdamar and Kalayci. Note, however, that if one changes only the testing EEG data by randomizing the peak location of the spike files, which mimics the actual situation during online spike detection, the performance degrades significantly to the central curve (solid line) in Fig. 3. One notices that the area under the ROC curve is 0.49, equivalent to the performance of random assignments. This result is not surprising, as we have pointed out that the ANN trained using the prepared spike files classifies waveforms predominantly by using the information from the 10th data point.

One may think of a remedy by implementing an additional pre-processing algorithm to search for local peaks in a continuous EEG record, such that all testing EEG files sent to the ANN have their peaks located at the 10th point. In such a case, however, since all files, including non-spike waveforms, exhibit relatively large amplitudes at the 10th point, the strong weight of $W_{1(10,j)}$ would substan-
tially increase the chance of false-positive detection. The situation is shown in Fig. 3 as the lowest curve (dashed line), obtained by intentionally locating the peaks in all our non-spike testing data at the 10th point. The area under the ROC curve was found to be 0.40, showing that the increase in false positives renders the ANN performance even worse than random assignments. Hence this remedy is certainly unacceptable.

It should be clear now that the ANN trained in the above manner would not give satisfactory performance for online spike detection. Training of the ANN, therefore, should be accomplished by presenting EEG files with peak locations randomized. Such an investigation was undertaken by plotting the ROC curves (Fig. 4) of the same ANN trained by our 600 EEG files, but randomized in location. Three different types of performance testing were attempted: the first with all EEG files randomized in location, the second with the peaks of the spike files prepared at the 10th point, and the third with non-spike files prepared. Note that since the same 600 EEG files are used both for ANN training and for testing, the area under the ROC curve should ideally approach unity. One can see from Fig. 4 that these three curves visually yield insignificant difference. The best achievable performance of the chosen ANN architecture, however, is characterized by the areas of about 0.93 under the ROC curves, corresponding to 87% true classification rate at equal sensitivity and specificity.

4. Discussion

The results from our study have implications in the inherent limitations of automatic spike detection via ANN using raw EEG as input. Fig. 2 shows that if the ANN were to be trained by erroneously prepared spike files, there would exist extraordinarily strong weighting for one single neuron, in this case the 10th input neuron. An inference drawn from the above statement is the expected negligible improvements in ANN performance with further increase in the number of hidden-layer neurons, as reported in Özdamar and Kalayci (1998). In actual situations, however, spike peaks are not automatically located at the same place in a data segment. This is particularly true during online spike detection. Even if a peak-searching algorithm is appended to pre-process the EEG data, the resulting increase in false-positive detection renders the ANN inferior to random assignments (Fig. 3). The high true classification rate reported by Özdamar and Kalayci is therefore concluded to be an artifact arising from erroneous data preparation and off-line validation.

The optimal number of hidden-layer neurons needed to perform a pattern recognition task has been shown to be \( \log_2(M) \) (Mirchandani and Cao, 1989), where \( M \) is the number of unconnected domains in the input data hyperspace. In the case of spike detection using 30-point raw EEG data as input, the number of isolated regions in the 30-dimensional hyperspace is expected to be quite large. In fact, if the spike detection ANN is to simulate the process of human visual recognition, information such as the rate of change or the peak curvature should be included in the ANN. Along with the fact that the spike peak should be allowed to appear at any of the 30 input data points, the need for hidden-layer neurons to produce these combinations can be enormous. We therefore infer that the number of hidden-layer neurons needed to achieve satisfactory spike detection performance is perhaps substantially larger than...
the number of input-layer neurons, if the raw EEG data were to be used as input. The architectures reported in Özdamar’s and Webber’s works were probably over-simplified for online classification of epileptic spikes using raw EEG as input.

Our results also imply that parameterized EEG data are a better choice for spike detection via ANN. This inference is consistent with results from Webber et al. (1994), although their comparison was nevertheless biased somehow toward raw EEG input. Fig. 4 demonstrated that ANN using raw EEG input could achieve no more than 87% for sensitivity and specificity simultaneously, if the small number of 6 hidden-layer neurons is used. The results of Fig. 3 even showed that the performance might become worse than random classification if the training data were erroneously prepared. As we conclude that the number of hidden-layer neurons needed to achieve satisfactory true classification rate is much larger than the number of input-layer neurons, constructing an ANN using fewer neurons with parameterized EEG input should be computationally much more economic.

Finally, note that the spike detection algorithms discussed above are purely morphological, i.e. the ANN detects single-channel spike waveforms only. Naturally, none of the spatial or temporal characteristics were taken into account. As the diagnostically useful spikes are those related to the source of abnormal interictal discharges, they are usually paroxysmal in time and present inter-channel gradient in peak amplitude and peak width. Under this circumstance, the ANN employing multi-channel parameterized information based on morphology as well as context, such as that reported by Webber et al. (1994), may exhibit stronger potential in providing diagnostic assistance than the purely morphological spike detection.

In conclusion, although the concept of artificial neural network is helpful in performing pattern-recognition tasks such as EEG spike detection, both its application and validation need to be undertaken with clear understanding of its physical meanings. In the particular case of spike detection using raw EEG data as input, the feasibility under current computer technology is likely to be low. Parameterized EEG data as input is perhaps a better choice.

Acknowledgements

The authors are indebted to Dr. Cheng-Ming Sung at Taipei City Psychiatric Center for providing the patient data and assistance in off-line validation of ANN performance. This work is supported in part by the National Science Council under grants NSC 87-2213-E002-084 and NSC 88-2213-E002-020.

References