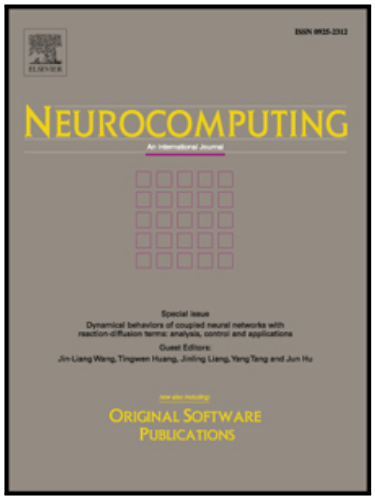
Communicated by Liang Lin

Accepted Manuscript



Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier

Md Mursalin, Yuan Zhang, Yuehui Chen, Nitesh V Chawla

PII: S0925-2312(17)30344-2
DOI: 10.1016/j.neucom.2017.02.053
Reference: NEUCOM 18127

To appear in: *Neurocomputing*

Received date: 29 June 2016
Revised date: 29 December 2016
Accepted date: 19 February 2017

Please cite this article as: Md Mursalin, Yuan Zhang, Yuehui Chen, Nitesh V Chawla, Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.02.053

Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier

Md Mursalin[1], Yuan Zhang[1]*,Yuehui Chen[1], Nitesh V Chawla[2]
[1]Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China
[2]Department of Computer Science and Engineering, University of Notre Dame, USA
yzhang@ujn.edu.cn

***Abstract*—Analysis of electroencephalogram (EEG) signal is crucial due to its non-stationary characteristics, which could lead the way to proper detection method for the treatment of patients with neurological abnormalities, especially for epilepsy. The performance of EEG-based epileptic seizure detection relies largely on the quality of selected features from an EEG data that characterize seizure activity. This paper presents a novel analysis method for detecting epileptic seizure from EEG signal using Improved Correlation-based Feature Selection method (ICFS) with Random Forest classifier (RF). The analysis involves, first applying ICFS to select the most prominent features from the time domain, frequency domain, and entropy based features. An ensemble of Random Forest (RF) classifiers is then learned on the selected set of features. The experimental results demonstrate that the proposed method shows better performance compared to the conventional Correlation-based method and also outperforms some other state-of-the-art methods of epileptic seizure detection using the same benchmark EEG dataset.**

***Index Terms*—Electroencephalogram (EEG), Discrete Wavelet transformation (DWT), Correlation-based Feature Selection (CFS), Improved Correlation-based Feature Selection (ICFS), Random Forest (RF)**

I. Introduction

Epilepsy is one of the most common disorders of the nervous system and affects people of all ages, races and ethnic backgrounds. Epileptic seizures are characterized by an unpredictable occurrence pattern and transient dysfunctions of the central nervous system due to excessive and synchronous abnormal neuronal activity in the cortex [1]. This activity could include several neurons of different locations and sizes. The clinical symptoms of epileptic seizures might affect the motor, sensory, and automatic functions of the body along with the consciousness, cognition, and memory of the patient [2]. To diagnose epilepsy, EEG signal interpretation is considered as the most prominent testing tools due to the fact that it is painless, low cost, and has efficient temporal resolution of long-term monitoring [3]. However for long EEG recording the visual interpretation becomes an expensive, intensive and tedious error-prone exercise and also result can vary from different neurophysiologists in the same recording [4]. Thus, there is an ever-increasing need for developing an effective method for automatic seizure detection in order to prevent the possibility of any missing information so as to give a proper diagnosis and possible treatment plan for epilepsy. There are several factors which affect a classifier's behavior, such as the dimensionality of the feature space, the number of available patterns to learn from, imbalanced class labels and so on. However, to develop an efficient method, the main challenges are designing an appropriate feature extraction method and selecting the most prominent features because the quality of feature set plays an important role on the classification accuracy. An optimized feature selection using the gentle adaboost algorithm was recently proposed by Peng et al. [5], which also emphasizes on the quality of the features.

In this study we have developed a method to detect seizure from EEG signal using improved correlation based feature selection (ICFS) method which outperforms the conventional correlation based feature selection (CFS) methods. Our study shows that, ICFS requires on average three fewer features than the conventional CFS method. The Random Forest (RF) classifier is used in this study for detecting the seizure. The contributions and novelties of the proposed method are,

i) Extracting the most prominent features from time domain, frequency domain, and entropy based features.

ii) Developing an Improved Correlation based Feature Selection (ICFS) method for selecting the most distinguishing features.

iii) Classifying the output by applying random forest classifier based on ICFS.

iv) Analyzing the performance of both conventional CFS and ICFS method.

The rest of this paper is organized as follows. Section II discusses the review of prior work related to the different feature extraction and classification methods. The details of the EEG processing pipeline for our approach and its components are described in section III. In section IV the evaluation procedure and the experimental results are presented and concluded in section V.

II. Related Work

Significant research has been done to detect the epileptic seizure from EEG signal. M. Z. Parvez et al. [6] proposed a generic seizure detection approaches for feature extraction of ictal and interictal signals using various established transformations and decompositions and least square support vector machine was applied on the features for classifications. They