Seizure Detection using Deep Multiset Canonical Correlation Analysis and Bayesian Optimization

Xuefeng Bai∗, Lijun Yan†

1 School of Computer, Shenzhen Institute of Information Technology, Shenzhen, Guangdong, 518172, China

∗Corresponding author’s e-mail: xuefeng.bai@outlook.com

Abstract. As one of the most common neurological diseases in the world, epilepsy seizure is difficult to ignore. Seizure detection is receiving more and more attention from researchers. Feature extraction is one of the key steps in automatic seizure detection. Lots of features have been proposed to detect seizure using EEG signal. However, few works focus on feature fusion. In this paper, deep multi set CCA is explored for seizure detection. Since deep neural network architecture has a great impact on performance of deep multiset CCA, bayesian optimization is employed to search architecture parameters automatically. Preliminary experiments show it is effective for seizure detection using deep multiset CCA and bayesian optimization. Satisfactory seizure classification results are achieved with little manual intervention.

1. Introduction

Epileptic seizure is a common disease in neurology. It has been estimated that there are approximately 65 million people in the world are suffer from epileptic [1]. At the same time, it is a time-consuming process for neurologists to examine electroencephalograms (EEGs) to diagnose epilepsy seizure, no mention there is a strong demand from the patients that monitor the seizure in real time. In order to facilitate neurologists and patients, some works on seizure detection is proposed.

A typical seizure detection process consists of data acquisition, signal processing, feature extraction and classification. AS lots of time-frequency characteristics in EEG signal, signal processing method such as Fourier transform [2], short time Fourier transform [3], Wavelet analysis [4]. Recently, adaptive signal decomposition method such as Empirical Mode Decomposition (EMD) [5], Empirical Wavelet Transform (EWT) [6] and Variational Mode Decomposition (VMD) [7] to make usage of nonlinear and nonstationary trait of seizure signal. In additional, variety of feature extraction methods are proposed in the literature [8], such as statistical features (mean, variance, skewness, kurtosis, etc.), amplitude-based features (energy, maximum values, minimum values, etc.) and entropy-based features.

As described above, a variety of signal decomposition and feature extraction methods have been used in seizure detection. Thus, there is a natural problem that fusing features to improve detection performance. It’s not a straightforward task to utilize features extracted with different parameters and variety of signal decomposition algorithm. Feature fusion is an effective approach to solve this problem.

Canonical Correlation Analysis (CCA) [9] is able to find projections of two features that maximizes the correlation between them. It is widely used to fuse features. CCA makes a linear projection of two features, which hinders the use of nonlinear information. To solve this problem, Kernel Canonical Correlation Analysis (KCCA) is proposed [10], in which a kernel is used to learn nonlinear information.

Although KCCA is able to handle nonlinear information, it is difficult to design the kernel function. What’s more, the designed kernel function cannot effectively adapt to the data to be analysed. In order
to overcome this limitation, Deep CCA (DCCA) is proposed [11]. A CCA based loss function is put on top of a deep neural network to learn nonlinear transformation of the data that maximize the correlation between them.

Another disadvantage of CCA and DCCA is that they can only deal with two features. Multiple features cannot be handled using these methods. One of the works that extend CCA for several input ‘set of variables’ is Multiset CCA (MCCA) [12]. Similar as CCA, nonlinear information is not handled in MCCA. A deep version of MCCA that named Deep Multiset CCA (dMCCA) is proposed [13] to overcome this deficiency.

To sum up, feature fusion may be an effective method for epilepsy detection. In this paper, dMCCA is explored to fuse features for seizure detection. To make the training process more automatic, bayesian optimization is employed to search architecture parameters of dMCCA.

The rest of this paper is organized as following. In section 2, features extracted from EEG signal are described first. Then, dMCCA are analysed to determine hyper parameters that needed to be optimized. Next, bayesian optimization is performed to determine architecture parameters of dMCCA. Last, experiments are performed on a publicly available dataset to validate effectiveness of the proposed method.

2. Methodology

Pipeline of seizure detection method in this paper is shown in figure 1. First, signal pre-processing methods including EMD, EWT, VMD are used to decompose EEG signal into intrinsic mode functions (IMFs). After that, features are extracted from these IMFs as in section 2.1. Then, hyper parameters of dMCCA are generated by bayesian optimization. Next, these parameters are used in dMCCA to perform feature fusion, and a classification experiment is performed using fused features. After that, another iteration starts. Repeat several times, the optimized hyper parameters are obtained. Finally, dMCCA using optimized hyper parameters are used to fuse feature and seizure detection can be performed.

2.1. Feature extraction

As analysed in section 1, features can be extracted using decomposed signals. Since different signal decomposition methods reflect different characteristics of the signal, several signal decomposition methods are used in this paper. More specifically, EEG signal is decomposed into intrinsic mode functions (IMFs) with EMD, EWT and VMD.

\[
SE = \frac{1}{N} \sum_{f=0}^{L} P_{XX}(f)
\]  

(1)
Based on the decomposed IMFs, both spectral domain based features and time domain based features are extracted. Part of features are extracted from spectral domain. The first feature is Spectral Energy (SE), extracted as Eq.1:

\[ SE = \sum_{f=0}^{\frac{f_s}{2}} P_{XX}(f) \log [\tilde{P}_{XX}(f)] \]

where \( N \) is the total number of spectral coefficients, \( \tilde{P}_{XX} \) is the PSD estimated by Welch’s method.

Then, Spectral Entropy (SEP) is calculated as Eq.2:

\[ SEP = -\sum_{f=0}^{\frac{f_s}{2}} \tilde{P}_{XX}(f) \log [\tilde{P}_{XX}(f)] \]

where \( \tilde{P}_{XX} \) is the normalized PSD.

Main frequency is an important characteristic of signal. Thus, Spectral Peak (SP) of PSD is computed. Beside SP, Spectral Centroid (SC) is also extracted as feature, which is calculated by Eq.3:

\[ SC = \frac{\sum_{f=0}^{\frac{f_s}{2}} \omega(f)M(f)}{\sum_{f=0}^{\frac{f_s}{2}} M(f)} \]

Bandwidth of AM and FM are also calculated as Eq.4, where \( A \) is the amplitude of the analytic signal, \( E \) is the Energy.

\[ B_{AM}^2 = \frac{1}{E} \int \left( \frac{dA(t)}{dt} \right)^2 dt \]
\[ \langle \omega \rangle = \frac{1}{E} \int \frac{d\phi(t)}{dt} A^2(t) dt \]
\[ B_{FM}^2 = \frac{1}{E} \int \left( \frac{d\phi(t)}{dt} - \langle \omega \rangle \right)^2 A^2(t) dt \]

In addition to spectral features, several time-domain features are extracted such as Hjorth parameters and statistical moments, which are defined as:

\[ Mob(x) = \sqrt{\frac{Var(\frac{d\phi(t)}{dt})}{Var(x(t))}} \]
\[ Comp(x) = \frac{Mob(\frac{d\phi(t)}{dt})}{Mob(x(t))} \]
\[ SK(x) = E \left[ \frac{x(t) - \mu}{\sigma} \right]^3 \]
\[ Std(x) = E \left[ \frac{x(t) - \mu}{\sigma} \right]^3 \]

2.2. Feature fusion using dMCCA with bayesian optimization

Schematic of dMCCA is shown in figure 2. In dMCCA, features from different modalities are transformed with different deep neural network. According to the principle and schematic of dMCCA, we can see that there are two key components: target loss function and deep neural network.
Loss function of dMCCA is derived from Multiset CCA (MCCA). MCCA tries to find projection vectors $\mathbf{v}_d$ to maximize the ratio between sum of between-set and within-set covariances by solving a generalized eigenvalue problem. Target of MCCA is named as inter-set correlation, which is formulated as:

$$\rho_d = \frac{1}{N-1} \frac{\mathbf{v}_d^T \mathbf{R}_B \mathbf{v}_d}{\mathbf{v}_d^T \mathbf{R}_W \mathbf{v}_d}, \quad d = 1, \ldots, D$$

where $N$ is number of modalities, $\mathbf{R}_B$ and $\mathbf{R}_W$ are the between-set and within-set covariance matrices that is calculate as:

$$\mathbf{R}_B = \sum_{l=1}^N \sum_{k=1, k \neq l}^N \mathbf{X}^l \left( \mathbf{X}^k \right)^T$$

$$\mathbf{R}_W = \sum_{l=1}^N \mathbf{X}^l \left( \mathbf{X}^l \right)^T$$

where $\mathbf{X}^l \in \mathbb{R}^{T \times D}, l = 1, \ldots, N, \mathbf{X} = \mathbf{X} - E(\mathbf{X})$.

In dMCCA, $\mathbf{X}$ is replaced with a hidden representation $\mathbf{H}$. To reduce the computation complexity, RB is estimated by $\mathbf{R}_B = \mathbf{R}_T - \mathbf{R}_W$. $\mathbf{R}_T$ is total covariance that is defined as:

$$\mathbf{R}_T = N^2 \left( \mathbf{H}^* - \mu \mathbf{1}^T \right) \left( \mathbf{H}^* - \mu \mathbf{1}^T \right)^T$$

where $\mathbf{H}^* = \sum_{l=1}^N \mathbf{H}^l$ is the average of all modalities; $\mu = \frac{1}{N} \sum_{l=1}^N \left( \frac{1}{N} \sum_{l=1}^N \mathbf{h}_l^T \right)$ the mean of all samples including all modalities.

Another key component in dMCCA is neural network. As shown in figure 2, features are fed into deep neural network to get a hidden representation before target function is calculated. Nonlinear processing capabilities and data adaptation are the results of this transformation.

Generally speaking, hyper parameters have great influence on the performance of deep neural network. There are many hyper parameters in deep neural network. In them, parameters related to topology is more important, which are the number of layers and the number of nodes in each layer. These parameters should be tuned carefully to achieve satisfactory performance. It is a non-trivial and time-consuming work, which requires excellent skills on parameters adjustment. Most of time, it is not realistic.

To handle this problem, bayesian optimization is employed to determine these parameters automatically in this paper. Target of bayesian optimization is finding the minimum of an objective function $f(x)$ on a bounded set $X$. To do this, a probabilistic model $p(f|D)$ is built to learn the relationship between $X$ and $f(x)$ through iteration, where $D$ is existing observations. During iteration, an acquisition function $a$ based on $p(f|D)$ that trades off exploration and exploitation is used to determine the next $x$. 

**Figure 2** Pipeline of seizure detection using dMCCA and bayesian optimization.
To determine hyper parameters of dMCCA, the main problem is how to design the observation method for bayesian optimization. In this paper, an observation is represented as \((x_i, y_i)\), where \(x_i\) is candidate hyper parameters of dMCCA, \(y_i\) is the seizure detection results using transformed features generated by dMCCA.

3. Experimental results

In order to verify the effectiveness of proposed method, a publicly available database offered by the University of Bonn [14] is used to perform experiments. There are 5 subsets in this database: Z, O, N, F, S. In each subset, there are 100 temporal series that is sampled with a frequency of 173.6 Hz and a duration of 23.6 s. The Z and O are collected from 5 health volunteers with eyes open and closed. The N, F, S are collected from epileptic patients. In particular, Set S is sampled during the seizure activity, set F and N are sampled during the seizure-free interval with electrodes placed on the epileptogenic zone and opposite hippocampus. We focus on dealing with the S, F, Z sets, which are corresponding to ictal, interictal and normal class.

For each parameter optimization experiment, 100 iterations are performed. Considering the influence of different random states on the optimization algorithm, experiments are repeated 100 times with different random states. A set of classifiers including Nearest neighbour, linear SVM, RBF SVM, Gaussian Process and multilayer perceptron are employed to perform classification on transformed features. Mean accuracy of these classifiers is recorded as seizure detection accuracy. Random search is selected as the comparison method to generate hyper parameters of dMCCA.

![Figure 3](image)

Figure 3 Sorted mean classification accuracy using bayesian optimization and random search.

For contrast, classification accuracy of each iteration is sorted. Mean value of sorted classification accuracy is collected and shown in figure 3. Although similar mean classification accuracy is obtained when all of the optimization epochs done, classification accuracy optimized by bayesian optimization outperforms that optimized by random search with a little advantage, which indicates better optimization result can be obtained using bayesian optimization.

In addition, maximum classification accuracy after each experiment is collected and sorted, which is shown in figure 4. There is ladder-like growth when random search or bayesian optimization is used as hyper parameters searching method, which is caused by uncertainty in the searching process. A better hyper parameter cannot be guaranteed after every experiment. In these experiments, higher accuracy is obtained when bayesian optimization is used, which indicates better searching ability of bayesian optimization.
At last, incumbent values of each iteration in experiments are collected and mean value of incumbent is shown in figure 5. Incumbent value is the best detection result since the start of parameters optimization. As it is replaced when a better detection result appears, incumbent values are increasing monotonously.

As can be seen from figure 5, incumbent values collected from bayesian optimization are a little bigger than those incumbent values collected from random search by a small margin. It indicates hyper parameters generated by bayesian optimization is better than parameters generated by random search.

4. Conclusion
We present a seizure detection method using hyper parameters optimized dMCCA with bayesian optimization. Features extracted from EMD, VMD, and EWT are fused using dMCCA. To archive a
satisfactory detection result, architecture parameters of dMCCA is optimized with bayesian optimization. Preliminary experiments show feasibility of the proposed method.

Acknowledgment
This research is supported by Young Innovative Talents Project of 2018 Guangdong University’s key scientific research platform and scientific research project, project number: 2018GkQNCX087.

References
[1] Thurman, David J and Beghi, Ettore and Begley, Charles E and Berg, Anne T and Buchhalter, Jeffrey R and Ding, Ding and Hesdorffer, Dale C and Hauser, W Allen and Kazis, Lewis and Kobau, Rosemarie and others. (2011) Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia, 52:2-26.
[2] Srinivasan, V. and Eswaran, C. and Sriraam, And N. (2005) Artificial Neural Network Based Epileptic Detection Using Time-Domain and Frequency-Domain Features. Journal of Medical Systems, 29(6):647-660.
[3] Alexandros, T, Tzallas, Markos, G, Tsipouras, Dimitrios. (2009) Epileptic seizure detection in EEGs using time-frequency analysis. IEEE Transactions on Information Technology in Biomedicine, 13:703–710.
[4] Robinson N, Vinod A, Ang K K, Tee K P and Guan C. (2013) EEG-Based Classification of Fast and Slow Hand Movements Using Wavelet-CSP Algorithm. IEEE Transactions on Biomedical Engineering, 60:2123–2132.
[5] Oweis R J and Abdulhay E. (2011) Seizure classification in EEG signals utilizing Hilbert-Huang transform. Biomedical Engineering Online, 10:38–38.
[6] Bhattacharyya A, Sharma M, Pachori R B, Sircar P and Acharya U R. (2018) A novel approach for automated detection of focal EEG signals using empirical wavelet transform. Neural Computing and Applications, 29:47–57.
[7] Taran S and Bajaj V. (2018) Clustering variational mode decomposition for identification of focal EEG signals. IEEE sensors letters, 2:1–4.
[8] Boonyakitanont P, Lek-Uthai A, Chomtho K and Songsiri J. (2020) A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomedical Signal Processing and Control, 57:101702.
[9] Hotelling H. (1992) Breakthroughs in statistics. Springer.
[10] Akaho, Shotaro. (2006) A kernel method for canonical correlation analysis. https://arxiv.org/abs/cs/0609071
[11] Galen Andrew, Raman Arora, Jeff Bilmes, Karen Livescu. (2013) Deep Canonical Correlation Analysis. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta. pp.1247-1255.
[12] Kettenring, Jon R. (1971) Canonical analysis of several sets of variables. Biometrika. 58(3):433-451.
[13] Somandepalli, Krishna and Kumar, Naveen and Travadi, Ruchir and Narayanan, Shrikanth. (2019) Multimodal representation learning using deep multiset canonical correlation. https://arxiv.org/abs/1904.01775
[14] Andrzejak, Ralph G and Lehnertz, Klaus and Mormann, Florian and Riecke, Christoph and David, Peter and Elger, Christian E. (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Physical Review E, 64(6):061907