Seizure Detection Using Deep Discriminative Multi-set Canonical Correlation Analysis

Xuefeng Bai(✉), Lijun Yan, and Yang Li

School of Computer, Shenzhen Institute of Information Technology, Shenzhen, China
xuefeng.bai@outlook.com, yanlj@sziit.edu.cn, yangzai1529@163.com

Abstract. Due to the nonlinear and nonstationary properties in EEG signals, some seizure detection methods tried to decompose EEG signal into nonlinear and nonstationary components and use them for feature extraction. Seizure detection results showed a certain degree of improvement in these approaches. Based on this idea, more signal decomposition methods have been explored. Signal decomposition methods are designed according to different principles, which show different properties of signals. So, it can be more effective using features extracted from different signal decomposition methods. Based on this consideration, a novel method for seizure detection based on feature combination exploiting deep neural network is proposed in this paper. We introduced a discriminative extension of Deep Multi-set Canonical Correlation Analysis (DMCCA) for seizure detection. Features extracted from different decomposed signals are combined by a joint optimization target of discriminative loss and multi-set canonical correlation loss, which is both discriminative and canonical correlated. Preliminary experiments show the proposed method improves seizure detection results in terms of accuracy and AUC.

Keywords: Seizure detection · Deep linear discriminative analysis · Deep Multi-set Canonical Correlation Analysis

1 Introduction

Epilepsy is one of the most common neurological diseases in the world. To recognize a seizure, it is typically necessary for physicians to observe EEG signal of the patient carefully, which is a time consuming process. What's more, it is not realistic for long time duration EEG signals [1]. Thus, there is an urgent need for automatic detection of seizure.

Thanks to the development of signal processing and machine learning, various seizure detection methods have been proposed [2,3]. In the early, Fourier

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.

M. Guan and Z. Na (Eds.): MLICOM 2020, LNICST 342, pp. 129–136, 2021.
https://doi.org/10.1007/978-3-030-66785-6_15

transform based spectral features are introduced to classify seizures and positive results are obtained [4]. As there is not any time-domain characteristic can be obtained using Fourier transform, time-frequency domain features based on short time Fourier transform (STFT) have been exploited [5]. What's more, in order to analyze multi-resolution time-frequency characteristics, wavelet analysis is used to extract features from EEG signal [6]. Wavelet analysis is suitable for stationary signal. However, some researchers have pointed out that frequency of EEG signals may change over a period of time, which indicates it is nonstationary [7]. Thus, new signal processing tool is desired for EEG based seizure detection.

Recently, Empirical Mode Decomposition (EMD) [8] is proposed to decompose EEG signal into intrinsic mode functions (IMFs) and then extracted feature for seizure detection [9]. EMD is an adaptive signal decomposition method that is able to decompose signal into IMFs, which can be handled by Hilbert Transforms [8]. EMD based feature extraction has been applied in lots of seizure detection methods [9–13].

Another adaptive signal decomposition method is Empirical Wavelet Transform (EWT) [14]. In EWT, signal will be decomposed into a fix number of predetermined modes. Compared with EMD, IMFs decomposed by EWT are more consistent. Some researchers have proposed to use EWT for EEG based seizure detection [15,16].

More recently, a new type of adaptive signal decomposition method called Variational Mode Decomposition (VMD) is proposed [17]. In VMD, a signal is decomposed into an ensemble of band-limited intrinsic mode functions. Thanks to the Wiener filtering in the decomposition process, it is more robust to noise. There are also some work using VMD to detect seizure [18,19].

After decomposed by signal process tools mentioned above, various features can be extracted. Although these adaptive decomposition methods are different in principle, the decomposed IMFs are all compact around specific center frequencies and have well-behaved Hilbert transforms [20]. Thus, features related to amplitude, bandwidth modulation, as well as instantaneous phase and amplitude can be calculated.

In addition to exploring new signal process techniques or feature extract methods, making full usage of existing methods may improve classification results. As different signal decomposition methods reflect different aspects of the signal, it can be more effective using features extracted from different signal decomposition methods. According to this consideration, a deep neural network based feature fusion method is proposed in this paper.

The rest of this paper is organized as follows. In Sect. 2, details of the proposed method is presented. In Sect. 3, experimental results are shown and discussed. We conclude this paper at last.

2 Methodology

2.1 Deep Linear Discriminant Analysis (Deep LDA)

As shown in Eq. 1, LDA tries to find a projection matrix **A** that maximizes the ratio of between class scatter Σ_{X_b} and within class scatter Σ_{X_w}.

$$\arg\max_{A} \frac{\left|\mathbf{A}\boldsymbol{\Sigma}_{X_b}\mathbf{A}^T\right|}{\left|\mathbf{A}\boldsymbol{\Sigma}_{X_w}\mathbf{A}^T\right|} \tag{1}$$

A is determined by solving a eigenvalue problem $\boldsymbol{\Sigma}_{X_b}\mathbf{e} = \mathbf{v}\boldsymbol{\Sigma}_{X_w}\mathbf{e}$, where $\mathbf{v}$ are eigenvalues. What's more, $\mathbf{v}$ quantifies the interval in direction of eigenvectors and the projection matrix A is the corresponding eigenvector $\mathbf{e}$ of this group.

A drawback of LDA is lack of ability to handle non-linear projection. Although non-linear method such as kernel LDA [21] is proposed, it is still difficult to design a well-adaptive kernel.

DeepLDA is a non-linear extension of LDA using deep neural network [22]. As shown in Fig. 1(a), a LDA loss function is put on top of a deep neural network in DeepLDA, which is able to learn latent representations. With the constrain of LDA loss function, DeepLDA maximize the separation between classes. Optimization target of DeepLDA is set as maximizing k smallest eigenvalues $\{v_1, \ldots, v_k\}$:

$$\arg\max_{\Theta} \frac{1}{k}\sum_{i=1}^{k} v_i \tag{2}$$

With objective function set as Eq. 2, deep neural network in DeepLDA is optimized to transform features into more discriminative form.

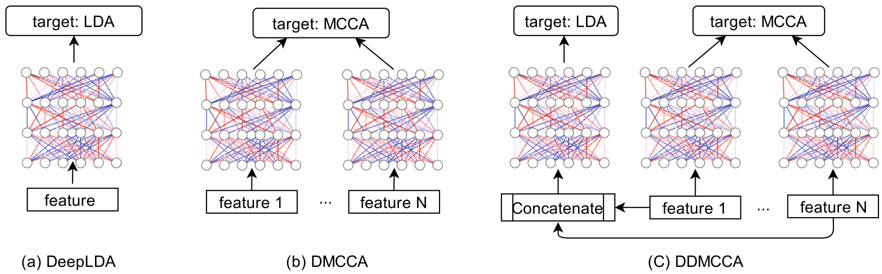

Fig. 1. Schematic sketches of nonlinear extension methods based on deep neural network

2.2 Deep Multi-set Canonical Correlation Analysis (DMCCA)

The goal of MCCA is trying to maximize the correlation between multiple data sets, as shown in Eq. 3.

$$\arg\max_{\mathbf{v}_d} \frac{\mathbf{v}_d^\top \mathbf{R}_B \mathbf{v}_d}{\mathbf{v}_d^\top \mathbf{R}_W \mathbf{v}_d} \tag{3}$$

where $\mathbf{R}_B$ and $\mathbf{R}_w$ is between-set and within-set covariance matrices. MCCA finds projection vectors $\mathbf{v}_d$ by solving a generalized eigenvalue problem:

$$\mathbf{R}_B\mathbf{V} = \mathbf{R}_W\mathbf{V}\mathbf{\Lambda} \tag{4}$$

Generally speaking, nonlinear information in the data sets is not handled. In order to handle these nonlinear information, Deep MCCA is proposed. Deep MCCA is an extension of MCCA [23], which is able to learn nonlinear relationship from features.

As shown in Fig. 1(b), main idea of DMCCA is set multi-set CCA as target function on the top of several deep neural networks, each network deal with a specific modality. Target of DMCCA optimization is shown in Eq. 6

$$\arg\max_{\Theta} \frac{1}{D}\Sigma_{d=1}^{D}\rho_d \tag{5}$$

where ρ_d is the inter-set correlation of N modalities.

With the constrain of this target function, DMCCA handles nonlinear transformations from different modalities when maximizing the ratio of between-modality and within-modality covariance of the input data.

2.3 Deep Discriminative Canonical Correlation Analysis (DDMCCA)

As described above, target of Deep LDA is designed for maximizing the ratio of between class scatter and within class scatter, and discriminative information is involved in the training process. While target of Deep MCCA is designed for maximizing the ratio of between modality covariance and within modality covariance, and modality information is involved in the training process.

Both discriminative information and modality information may improve the classification ability of transformed features. Thus, we proposed to use both discriminative information and modality information when training deep neural network, which is named as Deep Discriminative Multi-set CCA (DDMCCA). Schematic of DDMCCA is shown in Fig. 1(c). Main idea of the present work is to use both target function of Deep LDA and Deep MCCA. We optimize the deep neural network with a joint target function that both the discriminative power and correlation between different modalities are involved:

$$\arg\max_{\Theta} \left\{ \lambda_{LDA}\frac{1}{k}\sum_{i=1}^{k} v_i + \lambda_{MCCA}\frac{1}{D}\Sigma_{d=1}^{D}\rho_d \right\} \tag{6}$$

λ_{LDA} and λ_{MCCA} are weight factors of two target function respectively.

2.4 Feature Extraction and Feature Selection

In order to perform seizure detection using deep neural network based method mentioned above, features should be extracted from IMFs. We extract features from both spectral domain and time domain.

The first extracted feature is Spectral Energy (SE), given by Eq. 7

$$SE = \frac{1}{N} \sum_{f=0}^{\frac{f_s}{2}} P_{XX}(f) \tag{7}$$

where N is the total number of spectral coefficients, P_{XX} is the PSD estimated by Welch's method.

Then Spectral Entropy(SEP) is calculated as in 8

$$SEP = -\sum_{f=0}^{\frac{f_S}{2}} \bar{P}_{XX}(f) \log \left[\bar{P}_{XX}(f) \right] \tag{8}$$

where $\bar{P}_{XX}$ is the normalized PSD.

Main frequency is an important characteristic of signal. Thus, Spectral Peak (SP) of PSD is exploited. Beside this, Spectral Centroid (SC) is also extracted as in Eq. 9

$$SC = \frac{\sum_{f=0}^{\frac{f_S}{2}} \omega(f) M(f)}{\sum_{f=0}^{\frac{f_S}{2}} M(f)} \tag{9}$$

Bandwidth of AM and FM are also extracted as in Eq. 10, where A is the amplitude of the analytic signal, E is the Energy.

$$\begin{aligned} B_{AM}^2 &= \tfrac{1}{E} \int \left(\tfrac{dA(t)}{dt} \right)^2 dt \\ \langle \omega \rangle &= \tfrac{1}{E} \int \tfrac{d\phi(t)}{dt} A^2(t) dt \\ B_{FM}^2 &= \tfrac{1}{E} \int \left(\tfrac{d\phi(t)}{dt} - \langle \omega \rangle \right)^2 A^2(t) dt \end{aligned} \tag{10}$$

After that, several time-domain features are extracted such as Hjorth parameters and statistical moments, which are defined as:

$$Mob(x) = \sqrt{\frac{Var\left(\frac{dx(t)}{dt}\right)}{Var(x(t))}} \tag{11}$$

$$Comp(x) = \frac{Mob\left(\frac{dx(t)}{dt}\right)}{Mob(x(t))} \tag{12}$$

$$SK(x) = E\left[\left(\frac{x(t) - \mu}{\sigma} \right)^3 \right] \tag{13}$$

$$Std(x) = E\left[\left(\frac{x(t) - \mu}{\sigma} \right)^3 \right] \tag{14}$$

3 Experimental Results

We use a publicly available database offered by the University of Bonn [24] 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 category.

Samples of each category is decomposed by EMD, EWT and VMD into 6, 6 and 5 modes respectively, which is the best parameters according to [20]. Then, Deep Multi-set CCA, Deep LDA and Deep Discriminative Multi-set CCA are used to obtain the fused features.

Table 1. Classification AUC of different classification methods.

Methods	KNN	Linear SVM	RBF SVM	GP	NN
EEG	98.05	97.92	99.16	98.97	99.10
EMD	96.60	99.20	99.07	99.20	99.03
VMD	97.93	99.50	99.28	95.72	99.50
EWT	96.80	99.26	99.04	99.33	99.61
DMCCA	99.09	99.84	99.79	99.71	99.80
DeepLDA	97.55	98.00	98.67	95.97	98.53
DDMCCA	99.65	99.95	99.97	99.94	99.94

To evaluate performance of the obtained features, AUC and accuracy are computed using K nearest neighbors (KNN), Linear and RBF SVM, Gaussian Process classification (GP) and neural network classifier (NN). Classification results of AUC are shown in Table 1. With the help of the discriminative loss function, the highest AUC value is obtained by Deep Discriminative MCCA, with slight superior values for other methods.

Accuracy value of these methods are show in Table 2. All of the classification methods provide an accuracy above 90%. In these method, Deep Discriminative MCCA got the best performance.

Table 2. Accuracy of different classification methods.

Methods	KNN	Linear SVM	RBF SVM	GP	NN
EEG	94.67	90.33	92.67	96.00	92.67
EMD	90.00	95.00	94.00	95.33	94.67
VMD	94.00	95.67	96.33	83.00	95.33
EWT	89.00	94.67	93.67	94.67	94.67
DMCCA	95.67	97.33	97.00	96.33	96.00
DeepLDA	92.00	91.00	91.67	34.33	93.67
DDMCCA	97.67	97.67	98.67	98.33	98.00

4 Conclusion

We presented a discriminative extension of Deep MCCA, which is named as Deep Discriminative MCCA (DDMCCA). In this work, we maximize multi-set canonical correlation with a discriminative loss. A public available seizure dataset is used to verify the feasibility of DDMCCA. Features extracted from IMFs of EMD, EWT and VMD are used as the training features of the deep neural networks. Preliminary experiments indicate this method has the potential to improve seizure classification performance.

References

1. Mohseni, H., Maghsoudi, A., Shamsollahi, M.B.: Seizure detection in EEG signals: a comparison of different approaches. In: International Conference of the IEEE Engineering in Medicine and Biology Society, New York, USA, August 2006, pp. 6724–6727 (2006)
2. Sharmila, A., Geethanjali, P.: A review on the pattern detection methods for epilepsy seizure detection from EEG signals. Biomed. Eng. **64**(5), 507–517 (2019)
3. Sharmila, A.: Epilepsy detection from EEG signals: a review. J. Med. Eng. Technol. **42**(5), 368–380 (2018)
4. Srinivasan, V., Eswaran, C., Sriraam, N.: Artificial neural network based epileptic detection using time-domain and frequency-domain features. J. Med. Syst. **29**(6), 647–660 (2005). https://doi.org/10.1007/s10916-005-6133-1
5. Tzallas, A.T., Tsipouras, M.G., Fotiadis, D.I.: Epileptic seizure detection in EEGs using time-frequency analysis. IEEE Trans. Inf. Technol. Biomed. **13**(5), 703–710 (2009)
6. Robinson, N., Vinod, A.P., Ang, K.K., Tee, K.P., Guan, C.: EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm. IEEE Trans. Biomed. Eng. **60**(8), 2123–2132 (2013)
7. Pachori, R.B., Sircar, P.: EEG signal analysis using FB expansion and second-order linear TVAR process. Sig. Process. **88**(2), 415–420 (2008)
8. Long, S.R., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. **454**(1971), 903–995 (1998)

9. Oweis, R.J., Abdulhay, E.: Seizure classification in EEG signals utilizing Hilbert-Huang transform. Biomed. Eng. Online **10**(1), 38 (2011)
10. Pachori, R.B.: Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. Res. Lett. Signal Process. **2008**, 14 (2008)
11. Bajaj, V., Pachori, R.B.: Classification of seizure and nonseizure EEG signals using empirical mode decomposition. IEEE Trans. Inf Technol. Biomed. **16**(6), 1135–1142 (2012)
12. Alam, S.M.S., Bhuiyan, M.I.H.: Detection of seizure and epilepsy using higher order statistics in the EMD domain. IEEE J. Biomed. Health Inform. **17**(2), 312–318 (2013)
13. Alickovic, E., Kevric, J., Subasi, A.: Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. Biomed. Signal Process. Control **39**, 94–102 (2018)
14. Gilles, J.: Empirical wavelet transform. IEEE Trans. Signal Process. **61**(16), 3999–4010 (2013)
15. Bhattacharyya, A., Sharma, M., Pachori, R.B., Sircar, P., Acharya, U.R.: A novel approach for automated detection of focal EEG signals using empirical wavelet transform. Neural Comput. Appl. **29**(8), 47–57 (2016). https://doi.org/10.1007/s00521-016-2646-4
16. Saxena, S., Hemanth, C., Sangeetha, R.G.: Classification of normal, seizure and seizure-free EEG signals using EMD and EWT. In: 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software, pp. 360–366, March 2017
17. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. IEEE Trans. Signal Process. **62**(3), 531–544 (2014)
18. Taran, S., Bajaj, V.: Clustering variational mode decomposition for identification of focal EEG signals. IEEE Sens. Lett. **2**(4), 1–4 (2018)
19. Ravi Kumar, M., Srinivasa Rao, Y.: Epileptic seizures classification in EEG signal based on semantic features and variational mode decomposition. Cluster Comput. **22**(6), 13521–13531 (2019)
20. Carvalho, V.R., Moraes, M.F.D., Braga, A.P., Mendes, E.M.A.M.: Evaluating three different adaptive decomposition methods for EEG signal seizure detection and classification. BioRxiv, p. 691055 (2019)
21. Roth, V., Steinhage, V.: Nonlinear discriminant analysis using kernel functions. In: Advances in Neural Information Processing Systems 12, pp. 568–574. MIT Press (2000)
22. Dorfer, M., Kelz, R., Widmer, G.: Deep linear discriminant analysis. In: 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016 (2016)
23. Somandepalli, K., Kumar, N., Travadi, R., Narayanan, S.: Multimodal representation learning using deep multiset canonical correlation (2019)
24. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**(6), 061907 (2001)