35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A Robust Principal Component Analysis Algorithm for EEG-Based Vigilance Estimation

Li-Chen Shi, Ruo-Nan Duan and Bao-Liang Lu* *Senior Member, IEEE*

Abstract— **Feature dimensionality reduction methods with robustness have a great significance for making better use of EEG data, since EEG features are usually high-dimensional and contain a lot of noise. In this paper, a robust principal component analysis (PCA) algorithm is introduced to reduce the dimension of EEG features for vigilance estimation. The performance is compared with that of standard PCA, L1-norm PCA, sparse PCA, and robust PCA in feature dimension reduction on an EEG data set of twenty-three subjects. To evaluate the performance of these algorithms, smoothed differential entropy features are used as the vigilance related EEG features. Experimental results demonstrate that the robustness and performance of robust PCA are better than other algorithms for both off-line and on-line vigilance estimation. The average RMSE (root mean square errors) of vigilance estimation was 0.158 when robust PCA was applied to reduce the dimensionality of features, while the average RMSE was 0.172 when standard PCA was used in the same task.**

I. INTRODUCTION

Vigilance estimation is desirable in various human-machine interaction systems to monitor the vigilance of the operators in order to guarantee safety. Vigilance is defined as the ability to maintain focus of attention and to remain alert to stimuli for prolonged periods of time [1]. In our daily life, the operators of many human-machine interaction systems, especially vehicle drivers, should retain their vigilance at a relatively high level. Therefore, research on vigilance estimation has a practical significance for traffic safety. Electroencephalogram (EEG) has proven a signal closely related to vigilance, and various EEG-based estimating methods have been proposed [2], [3], [4], [5], [6].

The dimensionality of EEG features is usually quite high, increasing the complexity and reducing the operational efficiency of calculations based on the signals. Therefore, it is necessary to reduce the dimensionality of the features before we use them. Principal Component Analysis (PCA) is a approach widely used for dimensionality reduction. The classical PCA algorithm performs poorly in the presence of noise, skewing the extracted principal components, and therefore resulting in an inaccurate vigilance estimation. To improve the results, outliers can be manually removed, but this is a costly process for a large dataset. A noise insensitive algorithm is therefore required for EEG feature dimensionality reduction for vigilance estimation. In order to solve this problem, a robust PCA algorithm is introduced.

L. C. Shi , R. N. Duan, and B. L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University and MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.
* Corresponding author (bllu@sjtu.edu.cn)

Robust PCA is often used in the field of computer vision and image analysis. The robust PCA algorithm proposed by Campbell can effectively reduce the influence of data points with a lot of noise [7]. However, outliers in EEG data often occur in some particular electrodes and some particular frequency bands rather than all the features of the data points. Applying Campbell's approach to noisy EEG signals results in some dimensions being abandoned directly, therefore resulting in an undesirable amount of data loss. As a result, a more accurate robust PCA is required to process EEG features.

Experiments were carried out to collect EEG signals of twenty-three subjects whilst they were doing a monotonous visual task. Frequency-domain features were adopted to present vigilance-related EEG characteristics, and the feature sequence was smoothed before the dimensionality of features was reduced using robust PCA. Support vector machine (SVM) regression was applied to vigilance estimation. A comparison between standard PCA, L1-norm PCA, sparse PCA, and robust PCA proved that robust PCA was a suitable method in the dimensionality reduction of EEG data.

II. DATA

A. Experiments

The experimental data used was collected from a monotonous visual task. In the experiment, subjects were asked to sit in front of a screen which displayed traffic signs and was controlled by the NeuroScan Stim2 System. The traffic signs were divided into 4 groups based on the color, and each group contained 160 signs. A traffic sign was displayed for 500 milliseconds, and there was a 5-7 seconds interval with black screen between two traffic signs. The subjects were informed to recognize the color of the sign and press the button of this color.

Twenty-three students were chosen as subjects, and each of them participated in the experiment twice. The experiment was conducted after lunch, between 13:00-15:00 and lasted for 40-60 minutes.

During the experiment, responses of the subject were recorded by NeuroScan Stim2 System, and a 62-channel electrode cap was employed to record the EEG signals with the sample rate 500Hz synchronously. The arrangement of the electrodes was based on extended 10/20 system, and the signals were filtered to 0.1-100Hz.

B. Vigilance Measurement

An index of vigilance level was needed to evaluate the performance of the EEG-based vigilance estimation algo-

978-1-4577-0216-7/13/$26.00

rithm. Since the decrease of vigilance usually led to wrong response, the local error rate of response was usually adopted as the measurement of vigilance level [8]. Related research had proven that the fluctuations of vigilance were with the cycle lengths longer than 4 minutes [8], so the local error rate was calculated using the average error rate within a 2-minute window at 2-second step.

III. METHODS

A. Standard PCA Algorithm

PCA is a statistical algorithm which is widely used for dimensionality reduction as well as many other applications. PCA can be defined as a linear projection transforming the original multivariate data set to a new data set of linearly independent variables. The first principal component has the maximum variance; and each succeeding component in turn is orthogonal to the existing components and with the maximum variance.

Equivalently, standard PCA can be defined as the linear projection which minimizes the reconstruction error instead of maximizing the variance. According to this definition, the objective function of PCA can be expressed as follows,

$$J(U,V) = \min_{U,V} ||X - UV||^2 = \sum_{i=1}^{n} (x_i - Uv_i)^2, \quad (1)$$

where $U = (u_1, ..., u_k)$ is the first k projection vectors, and $V = (v_1, ..., v_n)$ is the data set after projection with constraints, $U^T U = I_k$ and $V = U^T X$.

B. L1-norm PCA Algorithm

Compared with the standard PCA algorithm, the basic idea of L1-norm PCA algorithm is to replace the L2-norm reconstruction error in Eq. (1) with L1-norm reconstruction error. L2-norm is more sensitive to noise than L1-norm, so the influence of noise decreased with the use of L1-norm. Considering the characteristics of EEG features, an extended L1-norm PCA algorithm firstly proposed by Kwak [9], is applied.

C. Sparse PCA Algorithm

The principle of sparse PCA is to use the important features of data when computing the principal components, instead of reducing the effect of outliers on the principal components projection matrix as L1-norm PCA does. Most elements in the projection vectors of standard PCA and L1-norm PCA are not zero, whereas the projection vectors of sparse PCA contain many zero elements. Because of the instability of EEG features and the differences between the experimental environments and between the subjects, the projection from the original EEG features to the principal components varied in different experiments. As the PCA algorithm can be seen as a minimization of the reconstruction error, PCA is essentially a regression, and shrinkage methods can be applied to increase the stability of the model.

Sparse PCA transforms the standard PCA to a regression about the reconstruction error, and makes the projection matrix sparse using loss functions. Zou *et al.* proposed a sparse PCA algorithm with a relatively high efficiency [10], and it is employed to reduce the dimension of EEG features. This kind of sparse PCA converts the PCA to a regression which is solved by using the elastic net algorithm [11].

D. Robust PCA Algorithm

Robust PCA can effectively remove the influence of outliers on the projection matrix. The main idea of robust PCA is to solve PCA problem using a weight method. Taking into account the issues of computational complexity and the characteristics of EEG features, the robust PCA algorithm proposed by Torre *et al.* [12] is adopted to reduce the dimensions of the data.

The algorithm extends the objective function of robust PCA, that the reconstruction error of weighted data points given by Xu *et al.* [13] to the reconstruction error of weighted features of each data point,

$$E(U,V,\mu,L) = \sum_{i=1}^{n} \sum_{p=1}^{d} (L_{pi}(e_{pi}^2/\sigma_p^2) + P(L_{pi})), \quad (2)$$

where $e_{pi} = x_{pi} - \mu_p - \sum_{j=1}^{k} u_{pj} v_{ji}$ is the reconstruction error of each feature, $X = (x_1, ..., x_n)$ is a d dimensional data set, $U = (u_1, ..., u_k)$ is the projection matrix of the first k principal components of the data set X, $V = (v_1, ..., v_n)$ is the data set after projection, $\mu = (\mu_1, ..., \mu_d)$ is the mean value of each feature, $0 \leq L_{pi} \leq 1$ is the weight of each feature, $P(L_{pi})$ is the penalty function, and $\sigma = (\sigma_1, ..., \sigma_d)$ is the scale factor for calculating the reconstruction error of each feature. In this algorithm, all the parameters need to be estimated except σ.

Based on robust statistics [14], if robust function Geman-McClure is used to express the reconstruction error of features, minimizing Eq. (2) is equivalent to minimizing the following formula,

$$E(U,V,\mu) = \sum_{i=1}^{n} \sum_{p=1}^{d} \rho(e_{pi}, \sigma_p), \quad (3)$$

where $\rho(e_{pi}, \sigma_p) = \frac{e_{pi}^2}{e_{pi}^2 + \sigma_p^2}$.

According to the theory of robust statistics, Eq. (3) can be converted into an objective function of linearly weighted errors of all features as follows,

$$\begin{aligned} &E(U,V,\mu,W) \\ &= \textstyle\sum_{i=1}^{n} (x_i - \mu - Uv_i)^T W_i (x_i - \mu - Uv_i) \\ &= \textstyle\sum_{p=1}^{d} (x^p - \mu_p 1_n - V^T u^p)^T W^p (x^p - \mu_p 1_n - V^T u^p), \end{aligned}$$

where $W_i = diag(w_{1i}, ..., w_{di})$ is a diagonal matrix constituted of the weights of the i_{th} sample's features, and $W^p = diag(w_{p1}, ..., w_{pn})$ is a diagonal matrix constituted of the weights of the p_{th} features in all the data.

Torre *et al.* designed a fast iteration algorithm based on the method of gradient descent. σ in this algorithm is the standard deviation of the p_{th} feature in the data set, which can be robust estimated using median absolute deviation.

This algorithm can effectively reduce the influence of the outliers of EEG features when the projection matrix of principal components is calculated. Also it makes use of EEG features as much as possible instead of abandoning EEG samples which contain some abnormal feature values, at the same time maintaining a relatively high efficiency for the algorithm.

This robust PCA cannot guarantee the orthogonality of the projection vectors during the solving process, so the projection matrix should be orthogonalized by the Gramm-Schmidt algorithm before it is used; standard PCA can also be applied to decompose the dataset reconstructed by a method based on robust PCA, and linearly transform the data using the projection matrix.

E. Data Analysis

The preprocessing of the EEG signals includes the noise reduction and artifact removal, which reduces the sampling rate to 100Hz. Since the EEG signals sampled around occipital lobe closely relate to the vigilance, data sampled from 9 electrodes (P1, Pz, P2, PO3, POz, PO4, CP1, CPz, CP2) around occipital lobe are adopted. Features are extracted every 2 seconds, and one experiment lasts for 40–60 minutes, so 1200–1800 samples from each experiment are obtained. The vigilance-related feature used is the logarithmic form of the EEG spectral feature, which is given the name differential entropy feature [15]. The frequency band of the features distributed from 2Hz to 44Hz, and the frequency is divided into 2Hz segments. The feature sequence is smoothed using a moving average filter with the window length 2 minutes for off-line method. For on-line vigilance estimation, a moving average filter with the window length 1 minute is applied, as well as a linear dynamic system (LDS) to smooth the feature sequence [5]. The dimensionality of the feature vector is 189, and reduced to 10 by PCA algorithm. The estimation method employed was SVM regression with the kernel function RBF, which generates the final vigilance estimation results of our system. When evaluating the algorithm, the training data was the EEG signal and the response collected from the first experiment of one subject, and the test data was from the experiment of the same subject the second time.

IV. RESULTS AND DISCUSSION

A. Results

The performance of different PCA algorithms are shown in Figs. 1 and 2, and Table I and II. As we can see from Figs. 1 and 2, in both off-line and on-line vigilance estimation, the root mean square errors (RMSEs) of different PCA algorithms are close to each other on most of the data, while the RMSE of robust PCA is obviously lower than that of other PCA on data containing much noise caused by the face and neck movement, such as the data of subject 2 and subject 23. This result infers that the four kinds of PCA algorithms have a similar performance, whereas the outstanding robustness of robust PCA algorithm makes it more suitable to process EEG data with much noise.

Table I and II show that both the average RMSE and the standard deviation of robust PCA algorithm is the lowest. It can be indicated from this result that the performance and robustness of robust PCA is both better than those of L1-norm PCA, sparse PCA, and standard PCA.

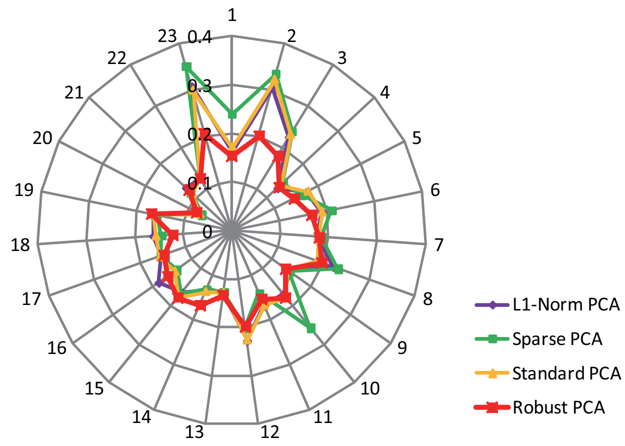

Fig. 1. RMSEs of different feature dimension reduction algorithms for off-line vigilance estimation

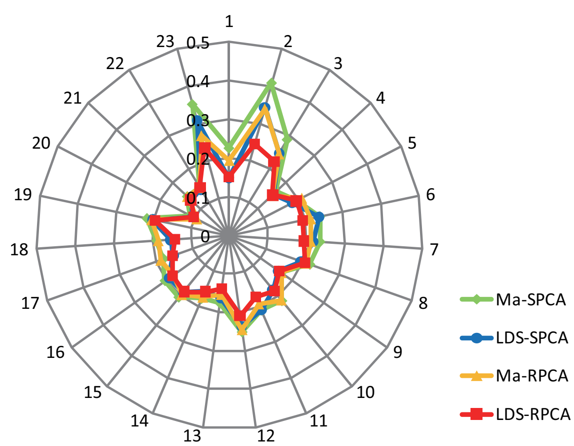

Fig. 2. RMSEs of different feature dimension reduction algorithms and different smoothing methods for on-line vigilance estimation

B. Discussion

The experimental results demonstrate the high accuracy and robustness of robust PCA when coping with EEG data with much noise no matter whether it is used in off-line conditions or on-line conditions, and what smoothing method is applied. Compared with standard PCA, robust PCA constructs the reconstruction error of principal components using the McClure robust function, and determines the range of noise by calculating the distribution variance of EEG features. Robust PCA can reduce the influence of specific noise features in one data point, which will not affect the other features. There is indeed some noise in the EEG features before dimension reduction, so applying the robust PCA will increase the performance when computing principal components.

TABLE I

RMSEs OF DIFFERENT FEATURE DIMENSION REDUCTION ALGORITHMS FOR OFF-LINE VIGILANCE ESTIMATION

Subject	PCA algorithms			
	Robust	L1-norm	Sparse	Std.
1	**0.155**	0.161	0.239	0.172
2	**0.203**	0.305	0.334	0.322
3	**0.18**	0.225	0.238	0.231
4	0.132	**0.128**	0.142	0.137
5	**0.145**	0.171	0.161	0.176
6	**0.168**	0.189	0.21	0.189
7	0.181	0.177	0.185	**0.177**
8	0.196	0.22	0.231	**0.186**
9	**0.136**	0.145	0.143	0.138
10	0.175	**0.168**	0.256	0.174
11	0.153	0.166	**0.139**	0.166
12	**0.198**	0.227	0.201	0.226
13	0.134	0.13	**0.127**	0.128
14	0.165	0.134	**0.131**	0.136
15	0.176	**0.163**	0.168	0.174
16	0.162	0.184	**0.139**	0.144
17	**0.148**	0.155	0.156	0.157
18	**0.122**	0.162	0.146	0.157
19	0.169	0.165	**0.158**	0.165
20	0.0824	0.0763	**0.0708**	0.0751
21	0.122	0.129	0.116	**0.112**
22	0.126	**0.125**	0.131	0.127
23	**0.208**	0.307	0.35	0.299
Average	**0.158**	0.175	0.181	0.172
±sd	±0.031	±0.054	±0.068	±0.055

TABLE II

RMSEs OF DIFFERENT FEATURE DIMENSION REDUCTION ALGORITHMS FOR ON-LINE VIGILANCE ESTIMATION

Method	LDS		Moving Average	
	R. PCA	Std. PCA	R. PCA	Std. PCA
Average	**0.177**	0.192	**0.198**	0.214
±sd	±0.035	±0.055	±0.049	±0.066

L1-norm PCA obtains principal components by minimizing the L1-norm reconstruction error, which can help to reduce the influence of noise features on principal components. However, it can not reduce the influence of specific noise features but only all features on principal components. In our experiment, the average RMSE of standard PCA is 0.172, which means that EEG features are relatively steady after feature smoothing. So using L1-norm to reduce the influence of all features may affect the normal features, and it does not necessarily increase the performance of PCA.

Sparse PCA is used to guarantee the reliability and stability of the solution. It chooses the important features from the original data to compose the principal components, which may not extract the principal components accurately. For the reason that most EEG features contain noise, the importance of one feature in most data points will make sparse PCA treat it as an important feature regardless of the abnormality of the feature value in other data points. As a result, the stability of sparse PCA decreases when computing the principal components.

In all of the experiments, subjects were allowed to move their face and eyes, which would introduce some noise data, also the thinking activity or the change of resistance of the electrodes would also produce noise features. In this condition, robust PCA performs better than other algorithms.

V. CONCLUSION

In this paper, a robust PCA algorithm for vigilance estimation is introduced and compared the performance of standard PCA, L1-norm PCA, sparse PCA, and robust PCA in dimensionality reduction on EEG data for both on-line and off-line vigilance estimation. Experimental results demonstrate that the performance and robustness of robust PCA are the best among these four methods, which implies that robust PCA is the most suitable method for reducing the feature dimension of EEG data for vigilance estimation.

VI. ACKNOWLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2009CB320901 and Grant No. 2013CB329401), and the European Union Seventh Framework Program (Grant No. 247619).

REFERENCES

[1] J. Warm, R. Parasuraman, and G. Matthews, "Vigilance requires hard mental work and is stressful," *Human Factors*, vol. 50, no. 3, pp. 433–441, 2008.

[2] T. Jung, S. Makeig, M. Stensmo, and T. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Trans. on Biomed. Eng.*, vol. 44, no. 1, pp. 60–69, 1997.

[3] P. Davidson, R. Jones, and M. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. on Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, 2007.

[4] C. Lin, L. Ko, I. Chung, T. Huang, Y. Chen, T. Jung, and S. Liang, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits and Systems-I*, vol. 53, no. 11, pp. 2469–2476, 2006.

[5] L. Shi and B. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Proc. 32nd Annu. Int. Conf. IEEE EMBS*. Buenos Aires, Argentina: IEEE, Aug. 31-Sep. 4 2010, pp. 6587–6590.

[6] ——, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135–143, 2013.

[7] N. Campbell, "Robust procedures in multivariate analysis I: Robust covariance estimation," *Applied Statistics*, vol. 29, pp. 231–237, 1980.

[8] S. Makeig and M. Inlow, "Lapse in alertness: coherence of fluctuations in performance and EEG spectrum," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, no. 1, pp. 23–35, 1993.

[9] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, 2008.

[10] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[11] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[12] F. D. L. Torre and M. Black, "A framework for robust subspace learning," *Int. J. Computer Vision*, vol. 54, pp. 117–142, 2003.

[13] L. Xu and A. Yuille, "Robust principal component analysis by self-organizing rules based on statistical physics approach," *IEEE Trans. Neural Netw.*, vol. 6, pp. 131–143, 1995.

[14] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Computer Vision*, vol. 19, no. 1, pp. 57–91, 1996.

[15] L. Shi, Y. Jiao, and B. Lu, "Differential entropy feature for EEG-based vigilance estimation," to appear in *35th Annu. Int. Conf. IEEE EMBS*.