LOCAL METRIC LEARNING FOR EEG-BASED PERSONAL IDENTIFICATION

Dongqi Cai, Kai Liu, Fei Su

School of Information and Communication Engineering
Beijing University of Posts and Telecommunications

ABSTRACT

There has been an increasing attention on Electroencephalograph (EEG) based personal identification over the last decade. Most existing methods address this problem by Euclidean metric based Nearest Neighbor (NN) search. However, under various recording conditions, simple Euclidean distance cannot model the similarity relations between EEG signals precisely. To overcome this drawback, a local metric learning based on Large Margin Nearest Neighbor (L-LMNN) for EEG based personal identification is proposed in this paper. For each EEG sample, a separate local metric is learned, making the distance between intra-class EEG samples minimized and simultaneously those of inter-class EEG samples maximized. To balance the locality and computational efficiency, the local metrics are approximated by weighted linear combinations of a small set of anchor samples. Experimental results demonstrate that the proposed approach obtains competitive performance compared with state-of-the-art methods. It improves the identification accuracy overall, especially at shorter EEG durations, which is important for improving the practicability of EEG-based personal identification system.

Index Terms— EEG, metric learning, LMNN, KNN, person identification

1. INTRODUCTION

With the development of integrated circuit, bioelectric sensors and signal processing tools, Electroencephalograph (EEG) acquisition instruments have become less expensive and more portable and reliable, making the brain wave signal gathering process maneuverable and the signal quality in representing the status of the individuals improved. Therefore, EEG has played a role in many practical systems [1, 2].

Compared with other biometric features, using human brain activities as a new modality has several advantages [3]. It is confidential, very difficult to mimic, and almost impossible to steal, and furthermore easy to change the password on purpose according to the users' mental tasks. However, EEG signals could be influenced by many factors, such as diet, circadian, etc., and it is difficult to fully control all these factors during the recording process. This brings extreme challenges to the practical applications of EEG based personal identification.

Most existing studies [4, 5, 6] have tried to address this problem by seeking a distinctive and stable feature representation of EEG signals and making nearest neighbor (NN) search for identity classification of EEG signals. With the feature set combining auto-regression (AR) model parameter [7] and power spectrum density (PSD)[8, 9], the identification system achieved an average accuracy of 97.3% on a dataset of 40 subjects at the recording duration of 3 minutes [4]. This promising performance proved the statistical commonality existed in one's EEG signals. Further, the effects of person's diet, circadian and recording durations on the EEG signals have also been evaluated quantitatively by the well-designed personal identification experiments [5]. It suggested that including more diverse EEG samples (different diet, times of day, etc.) in modeling and extracting more robust invariant features should be two ways to obtain better identification performance.

Although the current methods for EEG based personal identification have achieved encouraging performance, they still have some drawbacks. In the existing EEG-based identification system, the precision drops down significantly with the decrease of EEG recording duration, which reduces the practicability of the system. Furthermore, EEG based personal identification is essentially an NN search problem between different EEG signals. With finite learning instances, its performance strongly depends on the use of an appropriate distance measure. All the existing methods simply choose a standard Euclidean distance measure. However, the Euclidean distance metric is not very well adapted to the pattern classification problems in EEG based personal identification. Firstly, applying a standard distance measure is treating all the feature dimensions equally. However, for an artificially designed EEG feature representation, it is extremely hard if not impossible that all the feature dimensions make the same contributions in determining the class label of the test instance. Moreover, severe changes of the recording condition can cause significant variations of the intra-class EEG samples. It is rare

This work is supported by Chinese National Natural Science Foundation (61101212, 61372169), National High Technology R&D Program of China (863 Program) (No.2012AA012505), and the Fundamental Research Funds for the Central Universities.

that a single configuration of the importance and correlations of the feature dimensions can adapt to all EEG samples' representations. Under such circumstances, a single global metric cannot provide accurate distance measures for different EEG samples.

To address the above mentioned drawbacks in this paper, metric learning based on LMNN is introduced into EEG based personal identification, aiming to learn the optimal Mahalanobis distance metric that can maximize matching accuracy regardless of the choice of representation. Specifically, Local Large Margin Nearest Neighbor (L-LMNN) local metric learning is designed for EEG based personal identification. For each EEG single representation, a local metric is learned by LMNN to determine the relative importance and correlations of each feature dimension, making the distance between two EEG signals minimized if they belong to the same person, and maximized if they belong to different persons.

The main contributions of this work are two-folds. (1) We formulate EEG based personal identification as a metric learning problem, the optimal Mahalanobis distance metric is learned using LMNN for the similarity measurement between EEG signal representations. (2) In order to overcome the influence of different recording conditions on EEG signals, a local metric is learned for each sample using LMNN methods, and the local metrics are approximated by a weighted linear combination of the basis metrics. Experimental results show the effectiveness of the proposed L-LMNN method. It outperforms state-of-the-art methods and improves the identification accuracy especially at shorter EEG durations, which is important for improving the practicability of EEG-based personal identification system.

The rest of the paper is organized as follows. Section 2 introduces the proposed algorithm in detail. The experimental results are shown in section 3, which is followed by the conclusions and future work in section 4.

2. DETAILS OF THE PROPOSED METHOD

In this section, we will first introduce local metric learning and the guaranteed error bound of the approximation, then discuss how to learn the anchor points and local linear weights for each sample, and finally give a learning algorithm for basis metrics.

2.1. Local metric learning

Suppose a training set of N samples and the corresponding class labels to be denoted by $\{x_i, y_i\}_{i=1}^N$. As mentioned before, in local metric learning, a separate projection matrix L_i is learned for each sample x_i and the distance between x_i and x_j is calculated as

$$d(x_i, x_j) = \|L_i x_i - L_j x_j\|_2^2. \tag{1}$$

Note that the projection matrixes L_i is defined for the i^{th} sample, which means that in order to obtain good locality, each projection matrix in our approach only corresponds to a single sample. We call $\{L_i\}_{i=1}^N$ the local metric or local projection matrix of sample $\{x_i\}_{i=1}^N$.

However, learning local metrics for each sample needs extensive calculations. To strike a balance between locality and computational efficiency, the projection matrix is learned using a method inspired by Local Coordinate Coding (LCC) [10]. Specifically, as shown in Eq.2, the local projection matrix $\{L_i\}_{i=1}^N$ is approximated by a weighted linear combination of a small set of basis metrics $\{L_k\}_{k=1}^K$, which are associated with a set of anchor sample points $\{u_k\}_{k=1}^K$.

$$\begin{aligned} &L_i = \textstyle\sum_{k=1}^K w_{ik} L_k, \quad L_k \sim u_k \\ &w_{ik} \geq 0, \quad \textstyle\sum_{k=1}^K w_{ik} = 1, \quad i = 1, ..., N \end{aligned} \tag{2}$$

where $\{w_{ki}\}_{k=1}^K$ are the combination weights.

The feasibility and error bound of the above parameterizations can be verified using the following definition and Lemma. Let vector-valued function $f(x_i)$ denote the vector form of L_{Xi}. According to the following definition, $f(x_i)$ is a Lipschitz smooth function [10].

Definition 1 *A vector-valued function $f(x)$ on R^d is (α, β, p) -Lipschitz smooth with respect to a vector norm $\|\cdot\|$ if $\|f(x)-f(x')\| \leq \alpha\|x-x'\|$ and $\|f(x)-f(x')-\bigtriangledown f(x')^T(x-x')\| \leq \beta\|x-x'\|^{1+p}$, where $\bigtriangledown f(x')$ is the derivative of f function at x'. We assume $\alpha, \beta > 0$ and $p \in (0, 1]$.*

Any Lipschitz smooth real function can be approximated by a linear combination of function values on a set of anchor points, and the following lemma gives the respective error bound for this approximation.

Lemma 1 *Let $\{W, \mathbf{\Theta}\}$ be nonnegative weightings and their corresponding set of anchor points $\mathbf{\Theta}$ in R^d. Let f be an (α, β, p)-Lipschitz smooth vector function. We have for all $x \in R^d$:*

$$\begin{aligned} &\left\|f(x) - \textstyle\sum_{\theta\in\Theta} w_\theta(x) f(\theta)\right\| \leq \alpha \left\|x - \textstyle\sum_{\theta\in\Theta} w_\theta(x)\theta\right\| \\ &+\beta \textstyle\sum_{\theta\in\Theta} w_\theta(x) \left\|\theta - \textstyle\sum_{\theta\in\Theta} w_\theta(x)\theta\right\|^{1+p}. \end{aligned} \tag{3}$$

Local metric learning can be divided into two parts: learning of the anchor points and linear weights and learning of the basis metrics.

2.2. Local Linear Weighting Learning

To obtain anchor points and the weights for all samples, a straightforward manner is to formulate it as an optimization problem by minimizing the reconstruction error bound in Eq.3 as that in LCC [10]. However, solving such an optimization problem is computationally expensive. In this paper, a much more efficient and practical method is employed. For anchor points, K-means clustering is applied on the sample space $\{x_i\}_{i=1}^N$ and the samples corresponding to the cluster centers are chosen as the anchor points $\{u_k\}_{k=1}^K$. The k-means method could guarantee the independence of the anchor points to a certain degree. For linear weightings, we simply compute the corresponding weights using inverse Euclidian distance [11] based weighting for nearest neighbors. Empirical results show that this does not lead to a decrease of the performance compared with minimizing codings [10]. Specifically, the weights are calculated as follow

$$W_{ik} = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{1}{2}\frac{d(x_i, u_k)}{\sigma^2}), \qquad (4)$$

where the scale parameter σ determines the smoothness of weightings through the size of the kernel.

2.3. Basis Metric Learning

According to Eq.1 and Eq.2, after expressing the local projection matrixes as a liner combination of basis metrics, the distance between x_i and x_j under projection matrixes L_i and L_j is given by

$$\begin{aligned} d(x_i, x_j) &= \left\| \sum_{k=1}^{K} w_{ik} L_k x_i - \sum_{k=1}^{K} w_{jk} L_k x_j \right\|_2^2 \\ &= \left\| \sum_{k=1}^{K} L_k (w_{ik} x_i - w_{jk} x_j) \right\|_2^2 \end{aligned} \qquad (5)$$

there are K local metrics $L_k{}_{k=1}^{K}$ needed to be learned. To simplify the learning process, we denote $L = [L_1, L-2, ..., L_K]$ as $Z_{ij} = [w_{i1}x_i^T - wj1x_j^T, ..., w_{iK}x_i^T - wjKx_j^T]^T$. So the distance between x_i and x_j can be rewritten as follow

$$d(x_i, x_j) = \|LZ_{ij}\|_2^2 \qquad (6)$$

Learning of L can resort to many kinds of traditional metric learning methods, LMNN metric learning is adopted in this paper.

For sample x_i, suppose $\eta_{ij} \in \{0, 1\}$ indicate whether x_j is a k nearest neighbor of x_i with the same label, and $y_{ij} \in \{0, 1\}$ indicate whether x_i and x_j belong to the same class or not. In LMNN, minimizing ε_{pull} (Eq.7) could minimize the distance between each training point and its K nearest neighbors with the same labels, and minimizing ε_{push} (Eq.8) could maximize the distance between all points with different labels which are closer than the aforementioned distance of the K nearest neighbors plus a constant margin.

$$\varepsilon_{pull} = \Sigma_{i,j}^{N} \eta_{ij} d(x_i, x_j) \qquad (7)$$

$$\varepsilon_{push} = \sum_{i,j}^{N} \sum_{l=1}^{N} \eta_{ij}(1 - y_{il})[1 + d(x_i, x_j) - d(x_i, x_l)]_+ \qquad (8)$$

The affine combination of ε_{pull} and ε_{push} could define the overall cost. Specifically, the cost function is given by:

$$\varepsilon_{LMNN} = (1 - \mu)\varepsilon_{pull} + \mu\varepsilon_{push} \qquad (9)$$

where μ is a tuning parameter, and the expression $[z]_+ = max(z, 0)$ denotes the standard loss. The cost function consists of two terms, the first term penalizes large distances between each training sample and its target neighbors, while the second term penalizes small distance between each training sample and the impostors (i.e., all the other training samples with different labels that closer than the target neighbors).

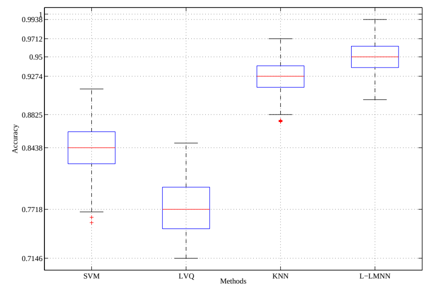


Fig. 1. Boxplots of SVM, LVQ, 1NN and L-LMNN method on Lab09 dataset at the recording duration of 180 seconds.

3. EXPERIMENTS

3.1. Datasets

This paper uses three datasets Lab09, Lab10 and Lab11, which contain EEG signals originally recorded by an unobtrusive, fast and easy-to-place equipment HXD-1. Lab09 consists of 480 EEG signals taken from 40 subjects with 12 segments per subject. Lab10 consists of 528 EEG signals taken from 33 subjects with 16 segments per subject. Lab11 consists of 176 EEG signals taken from 11 subjects with 16 segments per subject. The subjects were required to stay calm during recording. The data collection scheme is the same as that in [4, 5, 6]. Each original EEG segment lasts for around 5 minutes. We randomly cut out S sub-segments to simulate the data with a specific recording duration. For each subject in the dataset, we randomly select 70 percent segments for training and the remaining 30 percent for testing. 100 splits of the training and testing sets are randomly generated each time, and all the results in the following experiments are the average of 100 splits. The number of anchor points used in the proposed L-LMNN is denoted by C.

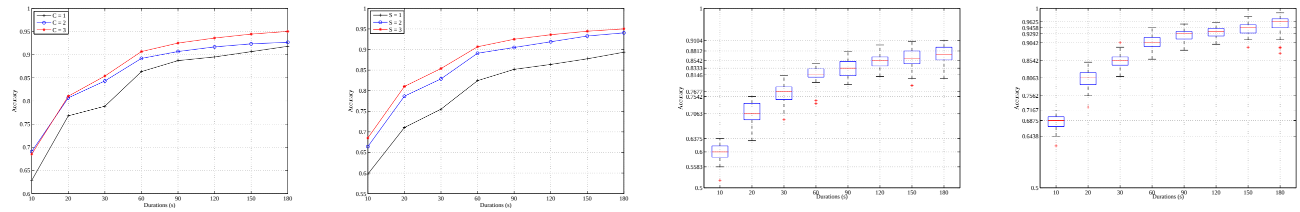
3.2. Feature sets

To capture the instinctive property of EEG signal for personal identification, A feature set combining the auto-regression (AR) model prediction coefficients [7], power spectrum density (PSD) [8, 9] and wavelet packet transform (WPT) based representation [12, 13] is used.

The performance of an AR model with an order ranging from 10 to 50 was tested, and the optimal order is found to be 19. Since the PSD below 4Hz is frequently contaminated by ocular artifacts, and the PSD above 33Hz is affected by a system specific notch filter, we use PSD at the particular frequency range of 4Hz to 32Hz with a frequency band being represented by 4 points as in [4]. Furthermore, the mean energy and variance of each subband decomposed by 6-level db4 WPT are extracted. The dimension of the final feature set is 255. Fisher's linear discriminant analysis (FDA) [14] is used for feature dimension reduction.

3.3. Performance Evaluation

The comparison of our L-LMNN method with three baseline methods on Lab09 dataset is shown in Fig.1. The baseline



(a) (b) (c) (d)

Fig. 2. (a) is the performance under different number of anchor points on Lab09. (b) is the performance using different number of sub-segments. (c) and (d) are performance comparisons based on different EEG recording durations on Lab09. (c) is using KNN classifier. (d) is using the proposed L-LMNN.

Table 1. Comparisons of identification accuracies on Lab09, Lab10 and Lab11 datasets (%)

Dataset	Method	10s	20s	30s	60s	90s	120s	150s	180s
Lab09	KNN	60.28	70.88	76.21	80.65	83.55	85.25	85.87	86.70
	L-LMNN	69.13	80.33	85.30	90.79	92.39	93.45	94.28	95.33
Lab10	KNN	50.99	62.12	67.09	71.84	73.47	75.13	74.87	77.01
	L-LMNN	58.47	72.53	78.24	84.77	84.69	87.01	87.56	90.06
Lab11	KNN	80.00	81.52	84.85	87.73	88.33	90.15	89.39	93.94
	L-LMNN	85.15	90.45	90.76	92.73	93.03	94.39	97.58	98.03

methods include multi-class SVM, Kohonen's LVQ Network and KNN classifier (k=1 in our experiments). The number of sub-segments used to simulate the data with 180 seconds and the number of anchor points in L-LMNN are both set to 1 in evaluating the basic performance. As can be seen from Fig.1, KNN outperforms SVM and LVQ respectively, which is consistent with the performance evaluation in [4]. And the proposed L-LMNN has a better performance than KNN, which achieves an average accuracy of 95% at the recording duration of 180 seconds. This verifies the effectiveness of learning local metrics for each EEG data. The performance of our method will be further evaluated by comparing with KNN in the following experiments.

Fig.2(a) and (b) show the identification results of L-LMNN under different number of anchor points and sub-segments on Lab09. Fig.2(a) is under $S = 3$ and Fig.2(b) is under $C = 3$. Theoretically, the number of anchor samples C depends on the complexity of the identification problem. More complex problems often require a larger number of anchor points to better model the data. Referring to the number of sub-segments, including more diverse EEG samples is a way to obtain better identification performance[5]. From Fig.2(a) and (b), we can conclude that longer recording duration achieves higher accuracy. However, too long recording duration renders the system impractical in real applications. It can also be concluded that the performance increases with the number of anchor points and sub-segments used in L-LMNN. Based on this experiment, we use $S = 3$ and $C = 3$ in the following settings.

Fig.2(c) and Fig.2(d) show the performance of KNN and L-LMNN under different recoding durations on Lab09. In this figure, L-LMNN outperforms KNN by an average of 8.95% under all the recording durations. The improvement is especially significant when the duration is shorter. For example, the average performance gain at 60-second duration is 10.14% using L-LMNN. This is important for improving the practicability of the real EEG based personal identification system. This experiment also verifies the advantage of local metric when measuring the distance between complex EEG data pairs over traditional Euclidean distance.

Table 1 shows the average identification performance on three datasets, Lab09, Lab10 and Lab11. The experiments are conducted using the same settings, in which the number of anchor points for L-LMNN and the number of sub-segments are both 3. As shown, the proposed L-LMNN outperforms KNN under all the recording durations on all the three datasets. This demonstrates the merits of adapting metric to local sample variations in similarity measurement of EEG signals.

4. CONCLUSIONS

In this paper, a local metric learning based on Large Margin Nearest Neighbor (L-LMNN) for EEG based personal identification is proposed to overcome the drawback of Euclidean metric. For each EEG sample, a separate local metric is learned, making the distance between intra-class EEG signals minimized and simultaneously those of inter-class EEG samples maximized. To balance the locality and computation efficiency, the local metrics are approximated by a weighted linear combination of the basis metrics, which correspond to a small set of anchor samples. Experimental results demonstrate that the proposed approach obtains competitive performance compared with state-of-the-art methods. It improves the identification accuracy overall, especially at shorter EEG durations, which is important for improving the practicability of EEG-based personal identification system.

5. REFERENCES

[1] Ramaswamy Palaniappan and Danilo P Mandic, "Biometrics from brain electrical activity: a machine learning approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 738–742, 2007.

[2] Sebastien Marcel and José del R Millán, "Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 743–752, 2007.

[3] Sebastien Marcel and José del R Millán, "Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 743–752, 2007.

[4] Fei Su, Liwen Xia, Anni Cai, Yibing Wu, and Junshui Ma, "Eeg-based personal identification: from proof-of-concept to a practical system," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3728–3731.

[5] Fei Su, Liwen Xia, Anni Cai, and Junshui Ma, "Evaluation of recording factors in eeg-based personal identification: A vital step in real implementations," in *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3861–3866.

[6] Fei Su, Liwen Xia, Anni Cai, and Junshui Ma, "A dual-biometric-modality identification system based on fingerprint and eeg," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–6.

[7] R.B Paranjape, J Mahovsky, L Benedicenti, and Z Koles, "The electroencephalogram as a biometric," in *Electrical and Computer Engineering, 2001. Canadian Conference on*. IEEE, 2001, vol. 2, pp. 1363–1366.

[8] M Poulos, M Rangoussi, and N Alexandris, "Neural network based person identification using eeg features," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*. IEEE, 1999, vol. 2, pp. 1117–1120.

[9] C.R Hema, M.P Paulraj, and Harkirenjit Kaur, "Brain signatures: a modality for biometric authentication," in *Electronic Design, 2008. ICED 2008. International Conference on*. IEEE, 2008, pp. 1–4.

[10] Kai Yu, Tong Zhang, and Yihong Gong, "Nonlinear learning using local coordinate coding," in *Advances in Neural Information Processing Systems*, 2009, pp. 2223–2231.

[11] Jan C van Gemert, Jan-Mark Geusebroek, Cor J Veenman, and Arnold WM Smeulders, "Kernel codebooks for scene categorization," in *Computer Vision–ECCV 2008*, pp. 696–709. Springer, 2008.

[12] Sijuan Huang and Xiaoming Wu, "Feature extraction and classification of eeg for imagery movement based on mu/beta rhythms," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*. IEEE, 2010, vol. 2, pp. 891–894.

[13] Banghua Yang, Guozheng Yan, and Rongguo Yan, "The feature extraction in brain-computer interface based on best basis of wavelet packet," *JOURNAL-SHANGHAI JIAOTONG UNIVERSITY-CHINESE EDITION*, vol. 39, no. 11, pp. 1879, 2005.

[14] Max Welling, "Fisher linear discriminant analysis," *Department of Computer Science, University of Toronto*, vol. 3, 2005.