The following publication Y. Jiang and F. H. F. Leung, "Using Double Regularization to Improve the Effectiveness and Robustness of Fisher Discriminant Analysis as A Projection Technique," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018 is available at https://doi.org/10.1109/IJCNN.2018.8489508.

Using Double Regularization to Improve the Effectiveness and Robustness of Fisher Discriminant Analysis as A Projection Technique

Yuechi Jiang
Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China
yuechi.jiang@connect.polyu.hk

Frank H. F. Leung
Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China
frank-h-f.leung@polyu.edu.hk

***Abstract*—Fisher Linear Discriminant Analysis (LDA) is a widely-used projection technique. Its application includes face recognition and speaker recognition. The kernel version of LDA (KDA) has also been developed, which generalizes LDA by introducing a kernel. LDA and KDA consists of a within-class scatter matrix and a between-class scatter matrix. The original formulations of LDA and KDA involve the inversion of the within-class scatter matrix, which may have singularity problem. A simple way to prevent singularity is adding a regularization term to the within-class scatter matrix. The resulting LDA and KDA are called Regularized LDA (RLDA) and Regularized KDA (RKDA). In this paper, we experimentally investigate how this regularization term will influence the performance of LDA and KDA. In addition, we introduce an extra regularization term to the between-class scatter matrix, and the resulting LDA and KDA are then called Doubly Regularized LDA (D-RLDA) and Doubly Regularized KDA (D-RKDA). We then apply LDA, KDA, RLDA, RKDA, D-RLDA and D-RKDA as a feature projection technique to two audio signal classification tasks. Gaussian Supervector (GSV) is used as the feature vector and linear Support Vector Machine (SVM) is used as the classifier. Experimental results show that, RLDA, D-RLDA, RKDA and D-RKDA are more effective than the conventional LDA and KDA. Besides, D-RLDA and D-RKDA are more robust than RLDA and RKDA.**

Keywords*—*Fisher linear discriminant analysis, kernel Fisher discriminant analysis, double regularization, audio signal classification

I. Introduction

Fisher Linear Discriminant Analysis (LDA) is a powerful feature projection technique. The kernel version of Fisher Linear Discriminant Analysis (KDA) has also been developed, which generalizes LDA by introducing a kernel [1][2]. The applications of LDA and KDA include face recognition [3]-[6], music classification [7] and speaker recognition [8]. In this paper, we apply LDA and KDA as a feature projection technique to deal with two audio signal classification tasks. One is microphone classification, aiming to recognize the recording microphone based on the recorded speech. The other is mobile phone classification, aiming to recognize the recording mobile phone according to the recorded speech.

It is known that Gaussian Supervector (GSV) is a good feature vector suitable for different audio classification tasks, such as speaker recognition [9], microphone identification [10][11], telephone handset identification [10], mobile phone identification [12] and verification [13]. The merit of GSV lies in that it can map audio signals of different lengths to a fixed-length feature vector. Therefore, in this paper, we use GSV as the feature vector for microphone and mobile phone classification. GSV is calculated based on a Universal Background Model (UBM) [9].

Support Vector Machine (SVM) is one of the most widely used classifiers. SVM has been applied to various audio signal classification tasks, such as speaker recognition [9], microphone identification [10][11][14][15], mobile phone identification [12][16]-[18], and telephone handset identification [10]. Therefore, in this paper, we use linear SVM as the classifier for microphone and mobile phone classification.

We apply LDA and KDA as the feature projection technique to improve the performance of the raw feature (i.e. GSV). The conventional LDA and KDA involve an operation of inverting the within-class scatter matrix, which may have singularity problem. This singularity problem can be solved by using Principal Component Analysis (PCA) to reduce the dimensionality of the raw feature [3]. However, dimensionality reduction may result in loss of information. In [1], this problem was solved by adding a small regularization term to the within-class scatter matrix. After adding this regularization term, the resulting LDA and KDA are then called Regularized LDA (RLDA) and Regularized KDA (RKDA) [19]. Although the authors in [1] pointed out that this regularization term could be useful, they did not further explore the usage of it. In this paper, we experimentally investigate how this regularization term will influence the performance of RLDA and RKDA. Besides of adding a regularization term to the within-class scatter matrix, we also add a regularization term to the between-class scatter matrix. The resulting LDA and KDA are then called Doubly Regularized LDA (D-RLDA) and Doubly Regularized KDA

The work described in this paper was substantially supported by a grant from The Hong Kong Polytechnic University (Project Account Code: RUG7).

(D-RKDA), as there are now two regularization terms. We then compare the performance of LDA, KDA, RLDA, RKDA, D-RLDA and D-RKDA as the feature projection technique in audio signal classification tasks. Experimental results show that RLDA and RKDA can improve the effectiveness over the conventional LDA and KDA, while D-RLDA and D-RKDA can improve both the effectiveness and robustness.

This paper is organized as follows. In Section II, we describe the formulations of LDA, RLDA and D-RLDA. In Section III, we describe the formulations of KDA, RKDA and D-RKDA. In Section IV, we briefly describe the audio datasets used in this paper. In Section V, we show and discuss the experimental results. A conclusion will be drawn in Section VI.

II. Fisher Linear Discriminant Analysis

A. Conventional LDA

Given N training vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_N\}$ belonging to K different classes, let C_k denote class k and N_k denote the number of vectors in class k. LDA aims to find a projection space, where vectors coming from the same class are moved closer while vectors coming from different classes are moved farther. The projection space is represented by a projection matrix $\boldsymbol{W}$ whose i-th column vector $\boldsymbol{w}_i$ is namely a projection direction or basis in the projection space. The relationship between the original vector $\boldsymbol{x}_n$ and the projected vector $\boldsymbol{y}_n$ is expressed as,

$$y_n = W^T x_n \tag{1}$$

LDA finds the projection matrix $\boldsymbol{W}$ by maximizing the objective function $J(\boldsymbol{W})$ defined in (2) below, where $\boldsymbol{S}_B$ is the between-class scatter matrix given by (3), $\boldsymbol{S}_W$ is the within-class scatter matrix given by (4), $\boldsymbol{m}_k$ is the mean value of those vectors in class k as given by (5), and $\boldsymbol{m}$ is the mean value of all the vectors as given by (6) [20].

$$J(W) = Trace\{(W^T S_W W)^{-1}(W^T S_B W)\} \tag{2}$$

where

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T \tag{3}$$

$$S_W = \sum_{k=1}^{K} \sum_{x_n \in C_k} (x_n - m_k)(x_n - m_k)^T \tag{4}$$

$$m_k = \frac{1}{N_k} \sum_{x_n \in C_k} x_n \tag{5}$$

$$m = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{6}$$

The solutions of maximizing $J(\boldsymbol{W})$ are the eigenvectors of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$, namely $\boldsymbol{w}_i$ is an eigenvector [20]. The number of projection directions is then the number of eigenvectors.

B. Regularized LDA (RLDA)

As $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$ involves the operation of matrix inversion, in order to prevent $\boldsymbol{S}_W$ to be singular, we can add a matrix $\varepsilon_W \boldsymbol{I}$ as a regularization term to $\boldsymbol{S}_W$, where ε_W is a nonnegative value and $\boldsymbol{I}$ is an identity matrix with the same dimensionality as $\boldsymbol{S}_W$. The resulting LDA is namely RLDA, and the solutions of RLDA are then the eigenvectors of $(\boldsymbol{S}_W+\varepsilon_W\boldsymbol{I})^{-1}\boldsymbol{S}_B$.

C. Doubly Regularized LDA (D-RLDA)

As can be seen from (3) and (4), the rank of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$ is at most K-1, because the rank of $\boldsymbol{S}_B$ is at most K-1 [20]. This means there are at most K-1 orthogonal eigenvectors of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$, i.e. K-1 orthogonal projection directions. Thus, we also add a regularization term $\varepsilon_B\boldsymbol{I}$ to $\boldsymbol{S}_B$, where ε_B is a nonnegative value and $\boldsymbol{I}$ is an identity matrix. Adding $\varepsilon_B\boldsymbol{I}$ to $\boldsymbol{S}_B$ can modify the rank of $\boldsymbol{S}_B$, and therefore can potentially increase the number of orthogonal projection directions. The resulting LDA is namely D-RLDA, and the solutions of D-RLDA will be the eigenvectors of $(\boldsymbol{S}_W+\varepsilon_W\boldsymbol{I})^{-1}(\boldsymbol{S}_B+\varepsilon_B\boldsymbol{I})$. If ε_B=0, D-RLDA becomes RLDA; if ε_W=0 and ε_B=0, D-RLDA becomes the conventional LDA.

In summary, adding $\varepsilon_W\boldsymbol{I}$ to $\boldsymbol{S}_W$ can prevent singularity, and adding $\varepsilon_B\boldsymbol{I}$ to $\boldsymbol{S}_B$ can increase the number of orthogonal projection directions. Singularity and low rank are two basic limitations of LDA; but fortunately, these two limitations can be partly overcome by adding two regularization terms to the original formulation of LDA.

III. Kernel Fisher Discriminant Analysis

A. Conventional KDA

According to (4), $\boldsymbol{S}_W\boldsymbol{w}_i$ can be expanded as follows,

$$\begin{aligned}
S_W w_i &= \sum_{k=1}^{K} \sum_{x_n \in C_k} (x_n - m_k)(x_n - m_k)^T w_i \\
&= \sum_{k=1}^{K} \sum_{x_n \in C_k} x_n (x_n - m_k)^T w_i \\
&\quad - \sum_{k=1}^{K} \sum_{x_n \in C_k} m_k (x_n - m_k)^T w_i \\
&= \sum_{k=1}^{K} \sum_{x_n \in C_k} x_n (x_n - m_k)^T w_i \\
&\quad - \sum_{k=1}^{K} \sum_{x_n \in C_k} \left(\frac{1}{N_k} \sum_{x_j \in C_k} x_j \right) (x_n - m_k)^T w_i \\
&= \sum_{n=1}^{N} x_n \sum_{k=1, x_n \in C_k}^{K} (x_n - m_k)^T w_i \\
&\quad - \sum_{n=1}^{N} x_n \sum_{k=1, x_n \in C_k}^{K} \sum_{x_j \in C_k} \frac{(x_j - m_k)^T w_i}{N_k}
\end{aligned} \tag{7}$$

If defining a new coefficient $\beta_n^{(W)}$ as in (8), $\boldsymbol{S}_W\boldsymbol{w}_i$ can then be expressed in terms of $\beta_n^{(W)}$ as given in (9).

$$\beta_n^{(W)} = \sum_{k=1,x_n\in C_k}^{K} \left(x_n - m_k\right)^T w_i - \sum_{k=1,x_n\in C_k}^{K} \sum_{j=1,x_j\in C_k}^{N} \frac{\left(x_j - m_k\right)^T w_i}{N_k} \tag{8}$$

$$S_W w_i = \sum_{n=1}^{N} x_n \beta_n^{(W)} \tag{9}$$

Eq. (9) supports the claim in [1] and [11] that, if $\boldsymbol{w}_i$ is a solution of maximizing $J(\boldsymbol{W})$, $\boldsymbol{w}_i$ must be a linear combination of all the vectors in the training set, as expressed in (10), where $\boldsymbol{v}_i$ is an N-dimension column vector and $(\boldsymbol{v}_i)_n$ denotes its n-th element.

$$w_i = \sum_{n=1}^{N} \left(v_i\right)_n x_n \tag{10}$$

If we map the original vector $\boldsymbol{x}_n$ to another dimensional space using a mapping function φ and then apply LDA in the mapped feature space, we then have (11), where $\boldsymbol{w}_i^{(\varphi)}$ denotes the projection direction in the mapped space, $\boldsymbol{v}_i^{(\varphi)}$ denotes the N-dimension coefficient vector in the mapped space, and $\varphi(\boldsymbol{x}_n)$ is the mapped feature vector, similar to (10). The superscript φ denotes the mapped feature space.

$$w_i^{(\varphi)} = \sum_{n=1}^{N} \left(v_i^{(\varphi)}\right)_n \varphi(x_n) \tag{11}$$

In the original feature space, we have the between-class scatter matrix $\boldsymbol{S}_B$ and the within-class scatter matrix $\boldsymbol{S}_W$. Correspondingly, in the mapped feature space, we have the between-class scatter matrix $\boldsymbol{S}_B^{(\varphi)}$ and the within-class scatter matrix $\boldsymbol{S}_W^{(\varphi)}$. It has been shown in [11] that, $\boldsymbol{w}_i^{(\varphi)T}\boldsymbol{S}_B^{(\varphi)}\boldsymbol{w}_i^{(\varphi)}$ and $\boldsymbol{w}_i^{(\varphi)T}\boldsymbol{S}_W^{(\varphi)}\boldsymbol{w}_i^{(\varphi)}$ can be expressed in terms of $\boldsymbol{v}_i^{(\varphi)}$ and two new matrices $\boldsymbol{U}_B$ and $\boldsymbol{U}_W$, as given by (12) and (13) below.

$$w_i^{(\varphi)^T} S_B^{(\varphi)} w_i^{(\varphi)} = v_i^{(\varphi)^T} U_B v_i^{(\varphi)} \tag{12}$$

$$w_i^{(\varphi)^T} S_W^{(\varphi)} w_i^{(\varphi)} = v_i^{(\varphi)^T} U_W v_i^{(\varphi)} \tag{13}$$

$\boldsymbol{U}_B$ and $\boldsymbol{U}_W$ are given by (14) and (15) below, where $\boldsymbol{l}_k$, $\boldsymbol{l}$ and $\boldsymbol{q}_n$ are N-dimension column vectors, whose j-th elements are given by (16), (17) and (18) respectively. The detailed derivation can be found in [11].

$$U_B = \sum_{k=1}^{K} N_k (l_k - l)(l_k - l)^T \tag{14}$$

$$U_W = \sum_{k=1}^{K} \sum_{x_n\in C_k} (q_n - l_k)(q_n - l_k)^T \tag{15}$$

where

$$\left(l_k\right)_j = \frac{1}{N_k} \sum_{x_n\in C_k} \varphi(x_j)^T \varphi(x_n) \tag{16}$$

$$\left(l\right)_j = \frac{1}{N} \sum_{n=1}^{N} \varphi(x_j)^T \varphi(x_n) \tag{17}$$

$$\left(q_n\right)_j = \varphi(x_j)^T \varphi(x_n) \tag{18}$$

By defining a kernel function $k(\boldsymbol{x}_j,\boldsymbol{x}_n)=\varphi(\boldsymbol{x}_j)^T\varphi(\boldsymbol{x}_n)$, we then have the KDA. Compared to LDA, KDA includes an implicit feature mapping before projection, and this mapping can be useful. In this paper, we use the Gaussian kernel defined in (19), where h is the kernel parameter.

$$k(x_j, x_n) = e^{-\left\|x_j - x_n\right\|^2 / h} \tag{19}$$

Then instead of finding the eigenvectors of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$ for $\boldsymbol{w}_i$, we now need to find the eigenvectors of $\boldsymbol{U}_W^{-1}\boldsymbol{U}_B$ for $\boldsymbol{v}_i^{(\varphi)}$. After finding $\boldsymbol{v}_i^{(\varphi)}$, for any given vector $\boldsymbol{x}_t$, its projected version $\boldsymbol{y}_t$ can be computed using (20), where $(\boldsymbol{y}_t)_i$ is the i-th element of $\boldsymbol{y}_t$.

$$\left(y_t\right)_i = w_i^{(\varphi)^T} \varphi(x_t) = \sum_{n=1}^{N} \left(v_i^{(\varphi)}\right)_n \varphi(x_n)^T \varphi(x_t) = \sum_{n=1}^{N} \left(v_i^{(\varphi)}\right)_n k(x_n, x_t) \tag{20}$$

B. Regularized KDA (RKDA)

Like LDA, an operation of matrix inversion is required in finding the eigenvectors of $\boldsymbol{U}_W^{-1}\boldsymbol{U}_B$. Thus, in order to prevent singularity (as well as improve the performance), we can add a regularization term $\varepsilon_W\boldsymbol{I}$ to $\boldsymbol{U}_W$, where ε_W is a nonnegative value and $\boldsymbol{I}$ is an identity matrix with the same dimensionality as $\boldsymbol{U}_W$. The resulting KDA is namely RKDA, and the solutions of RKDA are then the eigenvectors of $(\boldsymbol{U}_W+\varepsilon_W\boldsymbol{I})^{-1}\boldsymbol{U}_B$.

C. Doubly Regularized KDA (D-RKDA)

From (14) and (15), the rank of $\boldsymbol{U}_W^{-1}\boldsymbol{U}_B$ is still at most K-1, because the rank of $\boldsymbol{U}_B$ is at most K-1. This means there are at most K-1 orthogonal eigenvectors of $\boldsymbol{U}_W^{-1}\boldsymbol{U}_B$, i.e. K-1 orthogonal projection directions. Thus, we also add a regularization term $\varepsilon_B\boldsymbol{I}$ to $\boldsymbol{U}_B$, where ε_B is a nonnegative value and $\boldsymbol{I}$ is an identity matrix. Adding $\varepsilon_B\boldsymbol{I}$ to $\boldsymbol{U}_B$ can modify the rank of $\boldsymbol{U}_B$, and therefore can potentially increase the number of orthogonal projection directions. The resulting KDA is namely D-RKDA, and the solutions of D-RKDA will be the eigenvectors of $(\boldsymbol{U}_W+\varepsilon_W\boldsymbol{I})^{-1}(\boldsymbol{U}_B+\varepsilon_B\boldsymbol{I})$. If ε_B=0, D-RKDA

TABLE I. MICROPHONE SPEECH DATASET

Notation	Microphone Model	Number of Utterances		Duration
		Training	***Testing***	
M1	AKG C410B Head Mounted	240	260	2s ~ 5s
M2	AKH D80S Desktop	240	260	
M3	SONY ECM 66B Lapel	240	260	
M4	TARGET Lapel	240	260	
UBM	All the models	599		10s ~ 100s

TABLE II. MOBILE PHONE SPEECH DATASET

Notation	Mobile Phone Model	Number of Utterances		Duration
		Training	***Testing***	
MB1	Apple iPhone5	60	60	1s ~ 6s
MB2	HTC Desire C	60	60	
MB3	HTC Sensation XE	60	60	
MB4	LG GS290	60	60	
MB5	LG L3	60	60	
MB6	LG Optimus L5	60	60	
MB7	LG Optimus L9	60	60	
MB8	Nokia 5530	60	60	
MB9	Nokia C5	60	60	
MB10	Nokia N70	60	60	
MB11	Samsung E1230	60	60	
MB12	Samsung E2121B	60	60	
MB13	Samsung E2600	60	60	
MB14	Samsung Galaxy GT-I9100 S2	60	60	
MB15	Samsung Galaxy Nexus S	60	60	
MB16	Samsung GT-I8190 Mini	60	60	
MB17	Samsung GT-N7100 (Galaxy Note2)	60	60	
MB18	Samsung S5830i	60	60	
MB19	Sony Ericsson C510i	60	60	
MB20	Sony Ericsson C902	60	60	
MB21	Vodafone Joy 845	60	60	
UBM	All the models	2520		

becomes RKDA; if ε_W=0 and ε_B=0, D-RKDA becomes the conventional KDA.

IV. AUDIO DATASETS

In this paper, we use Ahumada-25 [21] for microphone classification, and MOBIPHONE [22] for mobile phone classification. Ahumada-25 consists of utterances recorded using 4 different microphones. For each microphone, 240 utterances are used to form the training set and 260 utterances are used to form the testing set. Consequently, 960 utterances are used for training, 1040 utterances are used for testing. Another 599 utterances are used for UBM, which includes all the models of microphones [11]. MOBIPHONE consists of utterances recorded from 21 different mobile phones. In our experiments, we use the modified MOBIPHONE described in [18]. For each mobile phone, 60 utterances are used to form the training set and 60 utterances are used to form the testing set. Consequently, 1260 utterances are used for training, 1260 utterances are used for testing. Another 2520 utterances are used for UBM, which includes all the models of mobile phones [18]. Details are also shown in Table I and Table II.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

GSV calculated from a 32-mixture UBM with a relevance factor of 5 is used as the feature vector. The UBM is constructed using the mixture splitting technique [23] and the Expectation Maximization (EM) algorithm [24]. More details of GSV can be found in [9]-[11]. The frame-level feature is the 24-dimension Mel-frequency Cepstral Coefficient (MFCC) obtained using Hamming window with 50ms frame length and 10ms frame shift. More details of MFCC can be found in [25]. The linear SVM is implemented using LIBSVM [26]. In the following, the performances of the raw feature and the raw feature assisted by LDA and KDA in microphone classification and mobile phone classification are compared. On using different versions of LDA (i.e. conventional LDA, RLDA, and D-RLDA), different regularization parameters ε_W and ε_B are evaluated. On using different versions of KDA (i.e. conventional KDA, RKDA, and D-RKDA), different regularization parameters as well as different kernel parameters are evaluated.

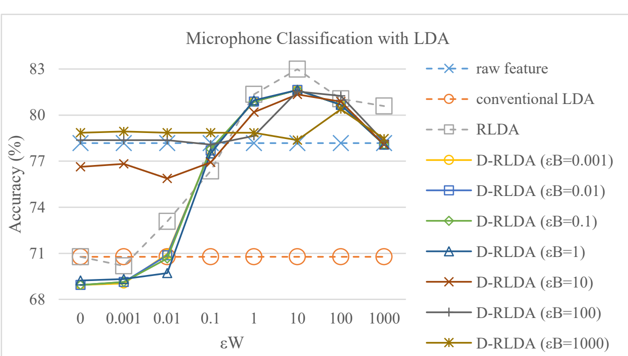

Fig. 1. Microphone classification with LDA.

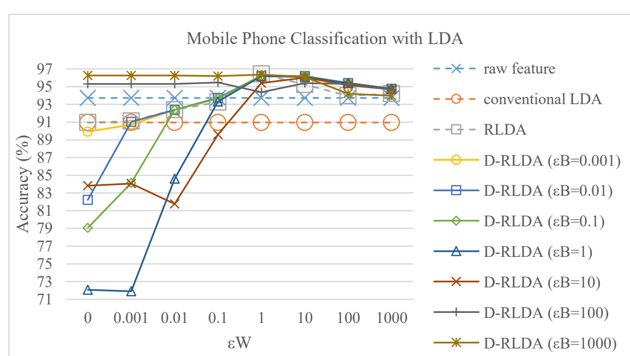

Fig. 2. Mobile phone classification with LDA.

A. *Classification Assisted by LDA*

In this part, we compare the performances of the conventional LDA, Regularized LDA (RLDA) and Doubly Regularized LDA (D-RLDA), in microphone classification and mobile phone classification. The classification results are illustrated in Fig. 1 and Fig. 2. In the figures, the polyline marked with "raw feature" represents the results of using GSV only, whereas other polylines represent the results of applying different versions of LDA to GSV. In fact, RLDA is merely a special case of D-RLDA with ε_B=0. Since "raw feature" and "conventional LDA" are unrelated to ε_W, the results are merely horizontal lines.

From Fig. 1 and Fig. 2, it can be seen that, the conventional LDA may not improve the performance of the raw feature in some cases; however, by choosing suitable regularization parameters, both RLDA and D-RLDA can give performance improvement. This indicates that RLDA and D-RLDA can be more effective than the conventional LDA, and therefore able to improve the quality of the raw feature. On using RLDA, when ε_W is small, the performance tends to be improved with the increase of the value of ε_W, however, when ε_W is too large, the performance tends to drop. On using D-RLDA, with a fixed value of ε_B, similar to RLDA, the performance is first improved and then degraded, with the increase of the value of ε_W (e.g. considering D-RLDA with ε_B=1 in Fig. 1 and Fig. 2). Another observation is that, the performances of D-RLDA with different values of ε_B can be quite different when a small value of ε_W is used, but tend to converge when the value of ε_W is large (e.g. comparing D-RLDA with ε_W=0.001 and ε_W=1000 in Fig. 1 and Fig. 2). This indicates that the two regularization terms are probably interacting with each other.

It can also be observed that, although ε_W is probably the dominant regularization parameter controlling the effectiveness of RLDA and D-RLDA because the polylines change obviously with the change of ε_W, the other regularization parameter ε_B can control the robustness of D-RLDA, making D-RLDA less dependent on the choice of ε_W (e.g. considering D-RLDA with ε_B=1000 in Fig. 1 and Fig. 2).

B. *Classification Assisted by KDA*

In this part, we compare the performances of the conventional KDA, Regularized KDA (RKDA) and Doubly Regularized KDA (D-RKDA), in microphone classification and mobile phone classification. Results of microphone classification are illustrated in Fig. 3, while results of mobile phone classification are illustrated in Fig. 4. In the figures, the polyline marked with "raw feature" represents the results of using GSV only, whereas other polylines represent the results of applying different versions of KDA to GSV. In fact, RKDA is merely a special case of D-RKDA with ε_B=0. Since "raw feature" and "conventional KDA" are unrelated to ε_W, the results are merely horizontal lines. On using different versions of KDA, different values of Gaussian kernel parameter h are evaluated (i.e. h=200, 500, 1000, 2000).

From Fig. 3 and Fig. 4, it can be seen that, the choice of the kernel parameter influences the effectiveness of the conventional KDA, and if the kernel parameter is not properly chosen, the conventional KDA cannot improve the performance of the raw feature (e.g. considering the plots in Fig. 4). Yet, with suitable regularization parameters, RKDA and D-RKDA can be more effective and thus can improve the performance of the raw feature even if the conventional KDA

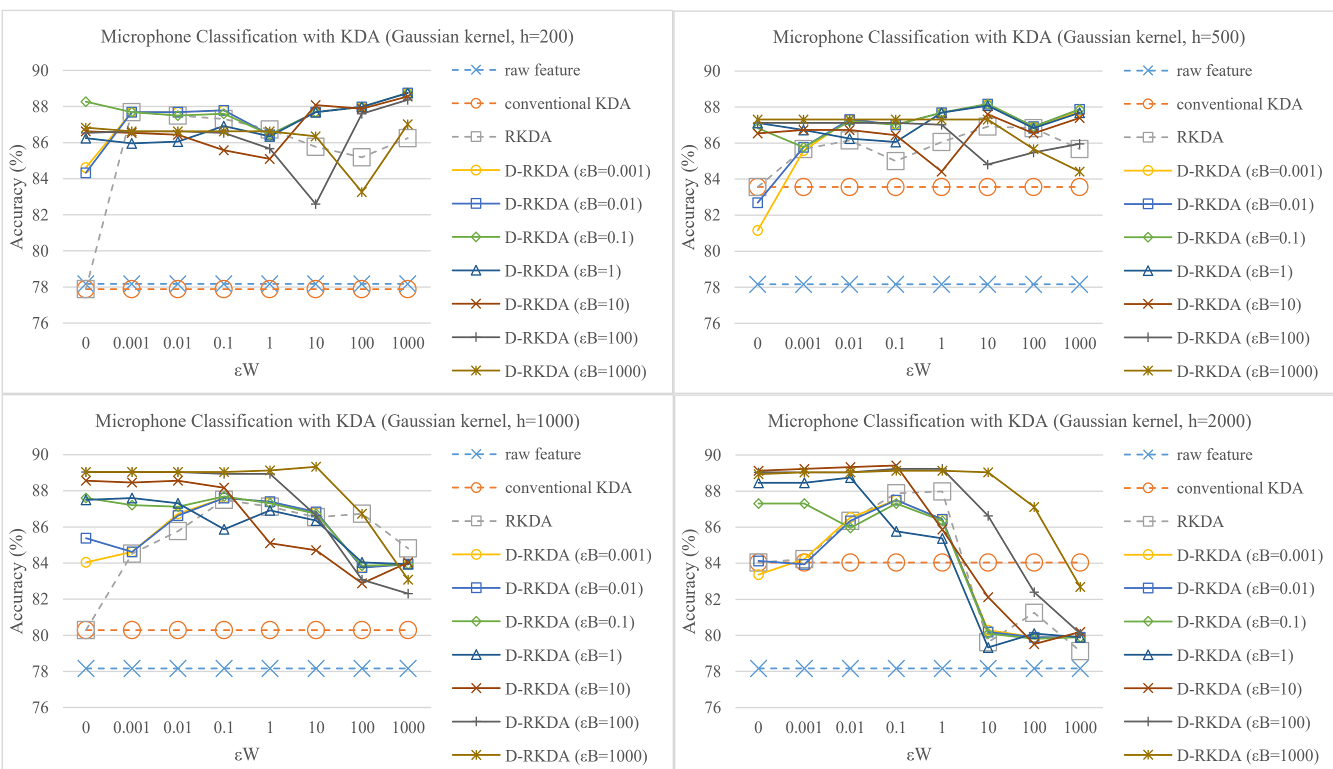

Fig. 3. Microphone classification with KDA.

fails. In addition, as can be seen from Fig. 4, the performances of RKDA and D-RKDA tend to degrade with the increase of the value of ε_W, meaning that both RKDA and D-RKDA are highly dependent on the choice of ε_W. However, on using D-RKDA, increasing the value of ε_B tends to make the performance less dependent on the choice of ε_W (e.g. comparing D-RKDA with ε_W=10, 100, 1000 in the bottom two plots in Fig. 3 and Fig. 4). From Fig. 3, increasing the value of ε_B may even improve the effectiveness of D-RKDA (e.g. considering the bottom two plots in Fig. 3). These observations indicate that RKDA and D-RKDA are less dependent on the choice of the kernel parameter than the conventional KDA. Compared to RKDA, the extra regularization term involved in D-RKDA makes it more robust and effective than RKDA.

VI. Conclusion

In this paper, we experimentally investigate how the regularization term will influence the effectiveness of Regularized Fisher Linear Discriminant Analysis (RLDA) as well as Regularized kernel Fisher Discriminant Analysis (RKDA). By choosing suitable regularization parameters, RLDA and RKDA can outperform the conventional LDA and KDA as the feature projection technique for signal classification. Besides, to make RLDA and RKDA more robust to different choices of regularization parameters, we propose the Doubly Regularized LDA (D-RLDA) and Doubly Regularized KDA (D-RKDA), which include two regularization terms. The regularization terms aim to 1) solve the singularity problem in matrix inversion, and 2) increase the number of projection directions that can be obtained. Experimental results on two audio signal classification tasks demonstrate that, D-RLDA and D-RKDA can be more robust and even more effective than RLDA and RKDA.

In particular, on using KDA as the feature projection technique, the kernel parameter plays an important role. If the kernel parameter is not wisely chosen, applying KDA may even degrade the performance of the raw feature. Fortunately, even the conventional KDA fails to improve the performance of the raw feature in some cases, with suitably chosen regularization parameters, RKDA and D-RKDA can still improve the performance of the raw feature.

References

[1] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller, “Fisher discriminant analysis with kernels,” in *Proc. IEEE Neural Networks for Signal Processing Workshop*, 1999, pp. 41-48.

[2] G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach,” *Neural Computation*, vol. 12, no. 10, pp. 2385-2404, 2000.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[4] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “Face recognition using LDA-based algorithms,” *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 195-200, 2003.

[5] M. H. Yang, “Kernel Eigenfaces vs. kernel Fisherfaces: face recognition using kernel methods,” in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2002.

[6] J. Yang, A. F. Frangi, J. Y. Yang, D. Zhang, and Z. Jin, “KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230-244, 2005.

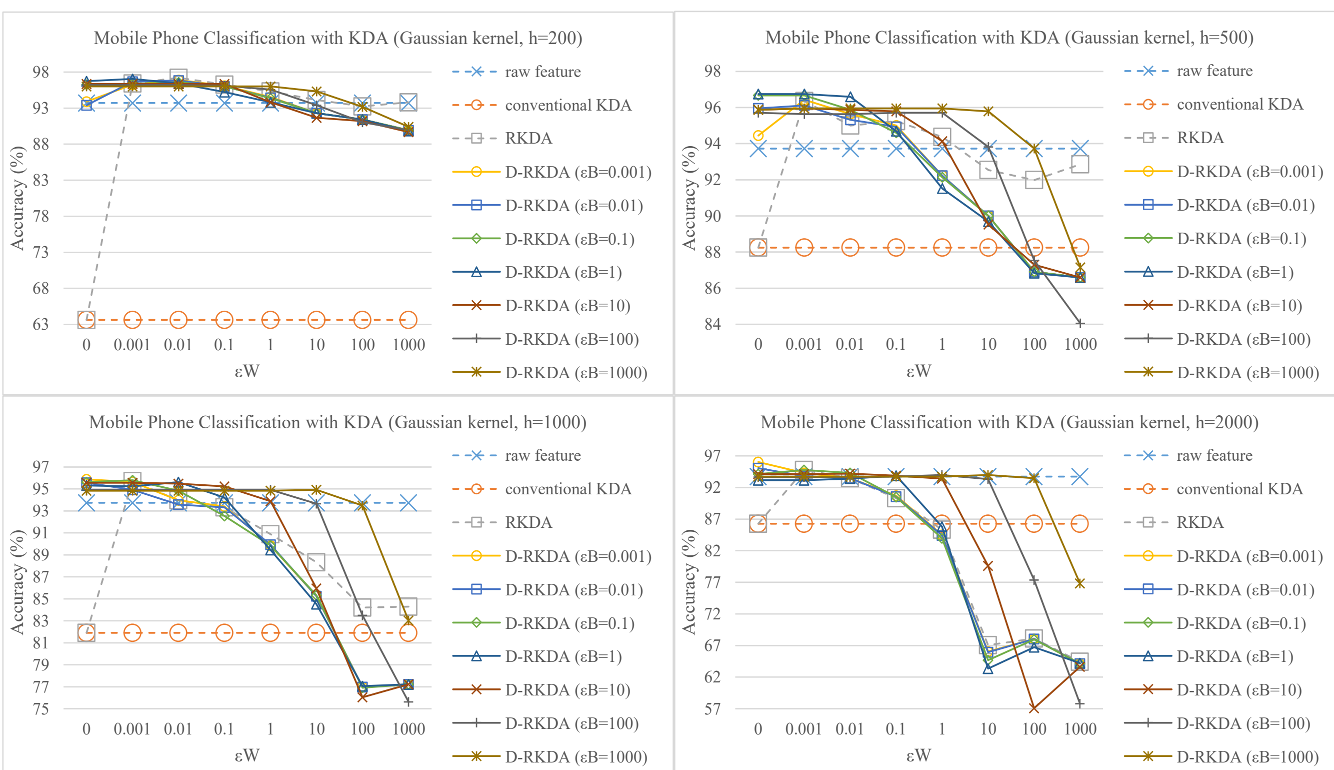

Fig. 4. Mobile phone classification with KDA.

[7] E. A. Cortizo, M. R. Zurera, and F. L. Ferreras, “Application of Fisher linear discriminant analysis to speech/music classification,” in *Proc. IEEE Int. Conf. on Computer as a Tool*, 2005, pp. 1666-1669.

[8] J. H. L. Hansen and T. Hasan, “Speaker recognition by machines and humans: a tutorial review,” *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74-99, 2015.

[9] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, “Support vector machines using GMM supervectors for speaker verification,” *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308-311, 2006.

[10] D. G. Romero and C. Y. E. Wilson, “Automatic acquisition device identification from speech recordings,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 1806-1809.

[11] Y. Jiang and F. H. F. Leung, “Using regularized Fisher discriminant analysis to improve the performance of Gaussian supervector in session and device identification,” in *Proc. IEEE Int. Joint Conf. on Neural Networks (IJCNN)*, 2017, pp. 705-712.

[12] C. L. Kotropoulis, “Source phone identification using sketches of features,” *IET Biometrics*, vol. 3, no. 2, pp. 75-83, 2014.

[13] L. Zou, Q. He, and J. Wu, “Source cell phone verification from speech recordings using sparse representation,” *Digital Signal Processing*, vol. 62, pp. 125-136, 2017.

[14] H. Q. Vu, S. Liu, X. Yang, Z. Li, and Y. Ren, “Identifying microphone from noisy recordings by using representative instance one class-classification approach,” *Journal of networks*, vol. 7, no. 6, pp. 908-917, 2012.

[15] F. Kurniawan, M. S. M. Rahim, M. S. Khalil, and M. K. Khan, “Statistical based audio forensic on identical microphones,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 5, pp. 2211-2218, 2016.

[16] C. Hanilci, F. Ertas, T. Ertas, and O. Eskidere, “Recognition of brand and models of cell-phones from recorded speech signals,” *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 2, pp. 625-634, 2012.

[17] C. Hanilci and F. Ertas, “Optimizing acoustic features for source cell-phone recognition using speech signals,” in *Proc. 1st ACM Workshop on Information Hiding and Multimedia Security*, 2013, pp. 141-148.

[18] Y. Jiang and F. H. F. Leung, “Mobile phone identification from speech recordings using weighted support vector machine,” in *Proc. IEEE 42nd Annual Conf. on Industrial Electronics (IECON)*, 2016, pp. 963-968.

[19] Z. Zhang, G. Dai, C. Xu, and M. I. Jordan, “Regularized discriminant analysis, ridge regression and beyond,” *Journal of Machine Learning Research*, vol. 11, pp. 2199-2228, 2010.

[20] C. M. Bishop, “Linear models for classification,” in *Pattern Recognition and Machine Learning*, Springer, 2006, ch. 4, pp. 179-224.

[21] J. O. Garcia, J. G. Rodriguez, and V. M. Aguiar, “AHUMADA: a large speech corpus in Spanish for speaker characterization and identification,” *Speech Communication*, vol. 31, no. 2, pp. 255-264, 2000.

[22] C. Kotropoulos and S. Samaras, “Mobile phone identification using recorded speech signals,” in *Proc. Int. Conf. on Digital Signal Processing (DSP)*, 2014, pp. 586-591.

[23] S. Young *et al.*, *The HTK book (v3.4)*, Cambridge: Cambridge University Press, 2006, pp. 156-157.

[24] D. Reynolds, “Gaussian mixture models,” in *Encyclopedia of Biometrics*, 2015, pp. 827-832.

[25] X. Huang, A. Acero, and H. W. Hon, “Speech Signal Representations,” in *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Upper Saddle River, NJ: Prentice Hall PTR, 2001, ch. 6, pp. 273-333.

[26] C. C. Chang and C. J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.