SFEMCCA: SUPERVISED FRACTIONAL-ORDER EMBEDDING MULTIVIEW CANONICAL CORRELATION ANALYSIS FOR VIDEO PREFERENCE ESTIMATION

Yoshiki Ito, Takahiro Ogawa and Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {ito, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

In this paper, we present supervised fractional-order embedding multiview canonical correlation analysis (SFEMCCA). SFEMCCA is a CCA method realizing the following three points: (1) learning noisy data with small number of samples and large number of dimensions, (2) multiview learning that can integrate three or more kinds of features, and (3) supervised learning using labels corresponding to the samples. In real data, it is necessary to deal with high dimensional noisy data with limited number of samples, and there are many cases where three or more kinds of multimodal and supervised data are treated in order to calculate more accurate projections. Therefore, SFEMCCA, which takes the above advantages (1)–(3) into account, is effective for data obtained from real environments. From experimental results, it was confirmed that accuracy improvements using SFEMCCA were statistically significant compared to the several conventional methods of supervised multiview CCA.

Index Terms— canonical correlation analysis, fractional-order technique, feature extraction, multiview approach.

1. INTRODUCTION

In research fields such as computer vision, pattern recognition and data visualization, many researchers have studied techniques to improve their accuracy by integrating multiple features of different characteristics. As one of the most famous methods, canonical correlation analysis (CCA) [1] is widely used for the integration of two kinds of features. CCA realizes effective feature integration by projecting each feature into a lower-dimensional canonical space where a pair correlation is maximized. In order to integrate features more accurately, several approaches such as locality preserving CCA (LPCCA) [2] and discriminative CCA (DCCA) [3] have been studied. For instance, LPCCA is a method extending CCA by locality preserving projection (LPP) [4], which takes the neighborhood structure into account. Moreover, supervised LPCCA (SLPCCA) [5], which preserves the neighborhood structure in the same class, has been proposed. On the other hand, DCCA is the method extending CCA by Fisher discriminant analysis (FDA) [6], which minimizes intra-class variance and maximizes inter-class variance. In addition, by integrating LPCCA and DCCA, they have been extended to discriminative locality preserving CCA (DLPCCA) [7] and its kernelized version (KDLPCCA) [8].

On the other hand, multiset CCA (MCCA) [9, 10], which maximizes the sum of three or more kinds of pair correlations, has been proposed. MCCA has been applied to higher level of research fields since the method can integrate not only two kinds of features but also three or more kinds of features. Moreover, by extending MCCA, supervised multi-view CCA (sMVCCA) [11] that can take class information into account has been proposed. Note that "multi(-)view" is the same as "multiset". By using sMVCCA, we have also conducted several researches to estimate videos that a target user likes but has not watched yet [12, 13]. These methods integrate the following three kinds of features: features obtained from videos (video features), sensing data obtained from target user's behavior while watching the videos (viewing behavior features) and evaluation scores for the videos by the user (label features). Then the optimal projections for the videos are calculated based on sMVCCA. By using "canonical video features", which are obtained by projecting the original video features, successful video recommendation becomes feasible. By using not only video features but also viewing behavior features and label features, it is possible to estimate the personalized preference of the videos more accurately than using only the original video features. In order to realize the estimation of personalized video preference more accurately, it is required to use a method which can integrate these features with higher accuracy.

Many CCA methods such as sMVCCA calculate the covariance (including autocovariance) of all features and estimate the optimal projections to transform each feature into the same lower-dimensional space. These methods assume the estimation under ideal conditions that the number of samples is sufficiently larger than the number of dimensions. Moreover, estimation of the correlation is performed on the assumption that the noise contained in the data has small influence. However, such ideal data is different from real data, and noise influences cannot be ignored. For example, in the eigenvalue decomposition (EVD) of the covariance matrix calculated from the data with noise, small number of samples and large number of dimensions, it is statistically proven that the eigenvalues become larger than those calculated from ideal data [14–16].

In order to solve this problem, a technique called "fractional-order technique" has been reported in several studies [17–19]. For example, fractional order singular value decomposition representation (FSVDR) [17] performs a correction that makes the covariance matrix calculated from real data similar to the one calculated from the ideal data. This method suppresses the increase of singular values by raising a fractional-order parameter (between 0 to 1) for singular values obtained by singular value decomposition (SVD) of the covariance matrix. In [18], fractional-order embedding canonical correlation analysis (FECCA) has been proposed. This method introduces the fractional-order technique into CCA, and performs a correction that makes the projections obtained from CCA of the real data similar to the ones obtained from the CCA of the ideal data. Herewith, in the experiments, the identification accuracy of images

This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744, JP15K12023.

using the corrected projection is improved. Furthermore, FECCA has been extended to fractional-order embedding multiset canonical correlations (FEMCCs) [19], which can be applied to three or more kinds of features.

In this paper, we propose supervised fractional-order embedding multiview CCA (SFEMCCA), which newly introduces the fractional-order technique into sMVCCA. sMVCCA does not take into account the bad influences of noise disturbance, small number of samples and large number of dimensions. The projections based on such data differ from those based on the ideal data with the sufficient number of samples, and the deviation has negative effect on the estimation. On the other hand, our method (i.e., SFEMCCA) solves this problem by suppressing the increase in eigenvalues of covariance matrices. In real data, it is necessary to deal with high dimensional noisy data with limited number of samples. In addition, there are many cases where three or more kinds of multimodal and supervised data are treated in order to calculate more accurate projections. Therefore, SFEMCCA is effective for data obtained from real environments. Since the data used in our previous researches also include the above-mentioned points, it will be effective to apply the method to them.

2. RELATED WORKS

2.1. Canonical Correlation Analysis and Its Reguralization

First, we explain canonical correlation analysis (CCA) [1]. Given a pair of matrices $\boldsymbol{X} \in \mathbb{R}^{D_x \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{D_y \times N}$, covariance matrices $\boldsymbol{C}_{mn}$ ($m, n \in \{x, y\}$) are calculated as follows:

$$\boldsymbol{C}_{mn} = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{m}_i - \overline{\boldsymbol{m}}) (\boldsymbol{n}_i - \overline{\boldsymbol{n}})^{\mathrm{T}}, \tag{1}$$

where N is the number of samples, $\boldsymbol{m}_i$ and $\boldsymbol{n}_i$ are data obtained from ith sample of a feature, $\overline{\boldsymbol{m}} = (1/N) \sum_{i=1}^{N} \boldsymbol{m}_i$, and D_x and D_y are the dimensions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. In CCA, two optimal projection vectors $\hat{\boldsymbol{w}}_x \in \mathbb{R}^{D_x}$ and $\hat{\boldsymbol{w}}_y \in \mathbb{R}^{D_y}$ are calculated by the following function:

$$\begin{aligned} \left\{\hat{\boldsymbol{w}}_x, \hat{\boldsymbol{w}}_y\right\} &= \arg\max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \ \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_{xy} \boldsymbol{w}_y \\ \text{s.t.} \quad & \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_{xx} \boldsymbol{w}_x = 1, \ \ \boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{C}_{yy} \boldsymbol{w}_y = 1. \end{aligned} \tag{2}$$

Then we can calculate the projection vectors in such a way that the correlation of canonical variates $\hat{\boldsymbol{w}}_x^{\mathrm{T}} \boldsymbol{X}$ and $\hat{\boldsymbol{w}}_y^{\mathrm{T}} \boldsymbol{Y}$ is maximized. For the constraints of Eq. (2), a method including regularization terms such as $\boldsymbol{C}_{mm} \rightarrow \boldsymbol{C}_{mm} + \epsilon \boldsymbol{I}_{D_m}$ is called regularized canonical correlation analysis (RCCA) [20], where ϵ is a small parameter and $\boldsymbol{I}_{D_m} \in \mathbb{R}^{D_m \times D_m}$ is the identity matrix. The regularization terms can suppress over-fitting.

2.2. Multiset Canonical Correlation Analysis

Next, we explain multiset CCA (MCCA) [9, 10]. Given matrices $\boldsymbol{X}_m \in \mathbb{R}^{D_m \times N}$ ($m \in \{m_1, m_2, \cdots, m_M\}$) from M kinds of features for integration, covariance matrices are calculated in the same manner as Eq. (1). In MCCA, M kinds of optimal projection vectors $\hat{\boldsymbol{w}}_m \in \mathbb{R}^{D_m}$ are calculated by the following optimization:

$$\begin{aligned} \left\{\hat{\boldsymbol{w}}_{m_1}, \hat{\boldsymbol{w}}_{m_2}, \cdots, \hat{\boldsymbol{w}}_{m_M}\right\} &= \arg\max_{\forall \boldsymbol{w}_m} \left(\sum_m \sum_{n \neq m} \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{C}_{mn} \boldsymbol{w}_n \right) \\ \text{s.t.} \quad & \sum_m \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{C}_{mm} \boldsymbol{w}_m = 1. \end{aligned} \tag{3}$$

In this way, by maximizing the sum of all pair correlations, CCA is extended for integration of three or more kinds of features.

3. SUPERVISED FRACTIONAL-ORDER EMBEDDING MULTIVIEW CANONICAL CORRELATION ANALYSIS

This section presents a novel canonical correlation analysis method, supervised fractional-order embedding multiview canonical correlation analysis (SFEMCCA). SFEMCCA is a method realizing the following three points: (1) learning noisy data with small number of samples and large number of dimensions, (2) multiview learning that can integrate three or more kinds of features, and (3) supervised learning using labels corresponding to the samples.

First, given matrices $\boldsymbol{X}_m \in \mathbb{R}^{D_m \times N}$ ($m \in \{m_1, m_2, \cdots, m_M, L\}$), autocovariance matrices $\boldsymbol{C}_{mm}$ and covariance matrices $\boldsymbol{C}_{mn}$ ($m \neq n$) are calculated in the same manner as Eq. (1). Since SFEMCCA is one of the supervised CCA methods, we use label features "$\boldsymbol{X}_L$" besides M kinds of features. $\boldsymbol{X}_L$ is a matrix obtained from D_L-dimensional one-hot vectors corresponding to each sample, where D_L is the number of classes. Next, the SVD is performed to the autocovariance matrices and covariance matrices as follows:

$$\boldsymbol{C}_{mn} = \boldsymbol{P}_{mn} \boldsymbol{\Lambda}_{mn} \boldsymbol{Q}_{mn}^{\mathrm{T}}, \tag{4}$$

$$\boldsymbol{\Lambda}_{mn} = \mathrm{diag}(\lambda_{mn,1}, \lambda_{mn,2}, \cdots, \lambda_{mn,D_{mn}}). \tag{5}$$

In the above equations, $\boldsymbol{P}_{mn} = \{\boldsymbol{p}_{mn,1}, \boldsymbol{p}_{mn,2}, \cdots, \boldsymbol{p}_{mn,D_{mn}}\}$ and $\boldsymbol{Q}_{mn} = \{\boldsymbol{q}_{mn,1}, \boldsymbol{q}_{mn,2}, \cdots, \boldsymbol{q}_{mn,D_{mn}}\}$ are the left singular matrix and the right singular matrix corresponding to $\boldsymbol{\Lambda}_{mn}$, respectively, where $D_{mn} = \mathrm{rank}(\boldsymbol{C}_{mn})$ and $\lambda_{mn,1} \geq \lambda_{mn,2} \geq \cdots \geq \lambda_{mn,D_{mn}}$. Note that the SVD is equal to the EVD if $m = n$. In order to suppress the increase of eigenvalues calculated by noisy data with small number of samples and large number of dimensions, autocovariance matrices and covariance matrices are reconstructed by using two kinds of fractional-order parameters ξ and ζ ($0 < \xi < 1$ and $0 < \zeta < 1$) as follows:

$$\tilde{\boldsymbol{\Lambda}}_{mn} = \begin{cases} \mathrm{diag}(\lambda_{mn,1}^{\xi}, \lambda_{mn,2}^{\xi}, \cdots, \lambda_{mn,D_{mn}}^{\xi}) & \text{if } m = n \\ \mathrm{diag}(\lambda_{mn,1}^{\zeta}, \lambda_{mn,2}^{\zeta}, \cdots, \lambda_{mn,D_{mn}}^{\zeta}) & \text{if } m \neq n, \end{cases} \tag{6}$$

$$\tilde{\boldsymbol{C}}_{mn} = \boldsymbol{P}_{mn} \tilde{\boldsymbol{\Lambda}}_{mn} \boldsymbol{Q}_{mn}^{\mathrm{T}}. \tag{7}$$

Next, we calculate the optimal projection vectors maximizing the sum of all pair correlations by using $\tilde{\boldsymbol{C}}_{mm}$ and $\tilde{\boldsymbol{C}}_{mn}$ ($m \neq n$), which were calculated by Eq. (7), as follows:

$$\begin{aligned} \left\{\hat{\boldsymbol{w}}_{m_1}, \hat{\boldsymbol{w}}_{m_2}, \cdots, \hat{\boldsymbol{w}}_{m_M}, \hat{\boldsymbol{w}}_L\right\} &= \arg\max_{\forall \boldsymbol{w}_m} \left(\sum_m \sum_n \boldsymbol{w}_m^{\mathrm{T}} \tilde{\boldsymbol{C}}_{mn} \boldsymbol{w}_n \right) \\ \text{s.t.} \quad & \sum_m \boldsymbol{w}_m^{\mathrm{T}} (\tilde{\boldsymbol{C}}_{mm} + \epsilon \boldsymbol{I}_{D_m}) \boldsymbol{w}_m = 1. \end{aligned} \tag{8}$$

Then we apply the Lagrange multiplier method to Eq. (8), and the following Lagrange function is obtained:

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{w}, \eta) = & \sum_m \sum_n \boldsymbol{w}_m^{\mathrm{T}} \tilde{\boldsymbol{C}}_{mn} \boldsymbol{w}_n \\ & - \eta \left\{ 1 - \sum_m \boldsymbol{w}_m^{\mathrm{T}} (\tilde{\boldsymbol{C}}_{mm} + \epsilon \boldsymbol{I}_{D_m}) \boldsymbol{w}_m \right\}, \end{aligned} \tag{9}$$

where η is a Lanrange multiplier. The Lagrange function is transformed by partial differentiations by all kinds of projection vectors as follows:

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}_m} = 2\left[\sum_n \tilde{\boldsymbol{C}}_{mn}\boldsymbol{w}_n - \eta\left\{(\tilde{\boldsymbol{C}}_{mm} + \epsilon \boldsymbol{I}_{D_m})\boldsymbol{w}_m\right\}\right]. \tag{10}$$

The extremum values of partial differentiations of Eq. (9) is calculated from Eq. (10). Next, the optimization problem in Eq. (8) is transformed into the following generalized eigenvalue problem:

$$\begin{bmatrix} \boldsymbol{0} & \tilde{\boldsymbol{C}}_{m_1m_2} & \cdots & \tilde{\boldsymbol{C}}_{m_1m_M} & \tilde{\boldsymbol{C}}_{m_1L} \\ \tilde{\boldsymbol{C}}_{m_2m_1} & \boldsymbol{0} & & \tilde{\boldsymbol{C}}_{m_2m_M} & \tilde{\boldsymbol{C}}_{m_2L} \\ \vdots & & \ddots & & \vdots \\ \tilde{\boldsymbol{C}}_{m_Mm_1} & \tilde{\boldsymbol{C}}_{m_Mm_2} & & \boldsymbol{0} & \tilde{\boldsymbol{C}}_{m_ML} \\ \tilde{\boldsymbol{C}}_{Lm_1} & \tilde{\boldsymbol{C}}_{Lm_2} & \cdots & \tilde{\boldsymbol{C}}_{Lm_M} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{m_1} \\ \boldsymbol{w}_{m_2} \\ \vdots \\ \boldsymbol{w}_{m_M} \\ \boldsymbol{w}_L \end{bmatrix}$$
$$= \eta_d \left(\begin{bmatrix} \tilde{\boldsymbol{C}}_{m_1m_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{C}}_{m_2m_2} & & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & & \tilde{\boldsymbol{C}}_{m_Mm_M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \tilde{\boldsymbol{C}}_{LL} \end{bmatrix} + \epsilon \boldsymbol{I}_{D_{sum}} \right) \begin{bmatrix} \boldsymbol{w}_{m_1} \\ \boldsymbol{w}_{m_2} \\ \vdots \\ \boldsymbol{w}_{m_M} \\ \boldsymbol{w}_L \end{bmatrix}, \tag{11}$$

where $D_{sum} = \sum_{\forall m} D_m$, $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_d \geq \cdots > \eta_D$, and $D \leq \min\{D_m\}$. Note that the η_d obtained by solving the above generalized eigenvalue problem are eigenvalues. Moreover, we can obtain the optimal projection matrix for each feature as follows:

$$\hat{\boldsymbol{W}}_m = [\hat{\boldsymbol{w}}_{m,1}, \hat{\boldsymbol{w}}_{m,2}, \cdots, \hat{\boldsymbol{w}}_{m,D}] \in \mathbb{R}^{D_m \times D}. \tag{12}$$

By using the above projection matrix, all features are projected to a lower-dimensional canonical space to obtain the following canonical features:

$$\hat{\boldsymbol{X}}_m = \hat{\boldsymbol{W}}_m^{\mathrm{T}} \boldsymbol{X}_m \left(\boldsymbol{I}_N - \frac{1}{N} \boldsymbol{1}_N \boldsymbol{1}_N^{\mathrm{T}} \right) \in \mathbb{R}^{D \times N}, \tag{13}$$

where $\boldsymbol{1}_N = [1, \cdots, 1]^{\mathrm{T}}$ is the N-dimensional vector that all elements are 1. In this way, the fractional-order technique can be newly introduced into sMVCCA. As we mentioned earlier, the projections based on noisy data with small number of samples and large number of dimensions differ from those based on the ideal data in the point of the increase in eigenvalues of covariance matrices. On the other hand, SFEMCCA suppresses them in Eq. (6) by using two kinds of fractional-order parameters in order to solve the problem. Therefore, SFEMCCA is effective for data obtained from real environments.

4. EXPERIMENTAL RESULTS

This section shows experimental results. In this experiment, we verify the performance of our method by applying it to estimation of videos that an user likes [12, 13]. In this study, three kinds of features (video features, viewing behavior features and label features) are integrated, and a classifier is trained by using "canonical video features". Therefore, it is suitable to apply our method to this study. First, the overview of the dataset used in this experiment is shown in **4.1**. Next, we explain experimental conditions in **4.2**. Finally, we show experimental results and their discussion in **4.3**.

4.1. Dataset

In this experiment, three keywords, "movie", "news" and "sports", were given as queries to YouTube. Five video clips were obtained for each keyword; that is, 15 video clips (65 seconds for each video) in total were prepared for the experiment. The subjects were eight men and two women of about 22 years of age. The subjects watched all of the video clips in a sitting position. A 15-inch display was set at a distance of one meter from the subjects, and a Kinect sensor to extract their viewing behavior was set on the display. Then the subjects evaluated all of the video clips by five ordinal grades, i.e., 5 (high preference), 4 (preference), 3 (undecided), 2 (low preference) and 1 (very low preference), by a console input using a keyboard. Note that three features (video, viewing behavior, label) used in our method were extracted at 1 fps, and we did not extract features for five seconds immediately after watching each video in order to secure the time to adjust the users' posture. In this way, a dataset including the three features could be obtained. The overview of the three features are shown below.

Video features (4241 dimensions):
We used an architecture of AlexNet [21] to extract the deep convolutional neural network (DCNN)-based visual features. Based on [22], features generated from 6th hidden layer of AlexNet, which was learned by ImageNet [23], (DeCAF_6) were used in our method. We then obtained 4096-dimensional visual features by using Caffe [24] as open source software of deep learning. Moreover, we adopted 145-dimensional audio features obtained by using MIRToolbox [25], which consist of Dynamics, Spectral, Timbre, Tonal and Rhythm. From the above, 4241-dimensional video feature vectors were obtained for each sample.

Viewing behavior features (22 dimensions):
Viewing behavior features including facial features and body movement features were obtained from each subject by using a Kinect sensor[1]. We extracted 14-dimensional facial features by using facial expression descriptor based on Action Units [26] and 8-dimensional body features from a subject's region and coordinates of the subject's skeleton. From the above, we obtained 22-dimensional viewing behavior feature vectors for each sample. The details are shown in [12].

Label features (5 dimensions):
Each target subject evaluated all of the video clips in five ordinal grades while watching them. We obtained an evaluation score of a sample of a video from the subject. This score was converted into 5-dimensional one-hot label feature vectors based on [11] for each sample.

The data used in this application includes the features with large number of dimensions such as video features (4241 dimensions). Moreover, it can be expected that some noise is mixed in viewing behavior features since the features are obtained by the sensor. Therefore, it is effective to apply SFEMCCA to these data.

4.2. Experimental Conditions

For the three kinds of features, we applied our method (i.e., SFEMCCA) to obtain the canonical video features as shown in Eq. (13). In our method, we empirically set $\epsilon = 0.01$, $D = 4 < \min\{4241, 22, 5\}$, $\xi \in \{0.6, 0.7, \cdots, 1.0\}$ and $\zeta \in \{0.6, 0.7, \cdots, 1.0\}$ for all subjects. An optimal set of the two parameters was determined for each subject. We compared our proposed SFEMCCA with the following three comparative methods: sMVCCA, DLPCCA and SLPCCA, which are one of the state-of-the-art methods of supervised multiview CCA as described earlier.

[1] http://www.microsoft.com/en-us/kinectforwindows/

Table 1. A comparison between MAE and MZE values of SFEMCCA and those of the conventional supervised multiview CCA methods.

	SFEMCCA		sMVCCA [11]		DLPCCA [7]		SLPCCA [5]	
Subjects	MAE	MZE	MAE	MZE	MAE	MZE	MAE	MZE
1	**0.429**	**0.341**	0.457	0.359	0.974	0.658	1.100	0.850
2	**0.401**	**0.357**	0.443	0.382	0.864	0.667	0.899	0.763
3	**0.567**	**0.443**	0.597	0.460	1.062	0.717	1.102	0.738
4	**0.571**	**0.422**	0.594	0.449	1.232	0.777	1.231	0.776
5	**0.422**	**0.322**	0.471	0.338	0.672	0.572	0.709	0.584
6	**0.452**	**0.415**	0.494	0.456	0.690	0.613	0.696	0.596
7	**0.341**	**0.296**	0.363	0.309	0.676	0.598	0.676	0.598
8	**0.433**	**0.378**	0.481	0.413	1.022	0.739	0.973	0.723
9	**0.358**	**0.313**	0.373	0.314	0.690	0.548	0.683	0.545
10	**0.451**	**0.398**	0.505	0.433	1.009	0.704	1.007	0.704
Average	**0.443**	**0.369**	0.478	0.391	0.889	0.659	0.908	0.688

In this experiment, support vector ordinal regression with implicit constraints (SVORIM) [27] was learned by canonical video features obtained by our method and the comparative methods to estimate unknown labels. In [28], it has been shown that SVORIM is one of the most effective ordinal regression methods in terms of Mean Absolute Error (MAE) and Mean Zero-one Error (MZE) as follows:

$$\mathrm{MAE} = \frac{1}{N_t}\sum_{i=1}^{N_t}\left|l_i^{\mathrm{Pr}} - l_i^{\mathrm{GT}}\right|, \quad \mathrm{MZE} = \frac{1}{N_t}\sum_{i=1}^{N_t}[\![l_i^{\mathrm{Pr}} \neq l_i^{\mathrm{GT}}]\!], \tag{14}$$

where N_t is the number of test samples, l_i^{Pr} is the predicted evaluation score of the ith sample, and l_i^{GT} is the ground truth (real score evaluated by the user) of the ith sample. Moreover, $[\![\cdot]\!]$ is a Boolean expression that outputs one if the inner condition $l_i^{\mathrm{Pr}} \neq l_i^{\mathrm{GT}}$ is true, otherwise zero. The range of MAE values is from 0 to 4, and that of MZE values is from 0 to 1. We can compare an average error by using MAE and accuracy without considering the order by using MZE. The lower their values are, the higher the performance is. Note that we adopted the Gaussian kernel in SVORIM, and an optimal set of the kernel parameter and the constant parameter corresponding to the best MAE was decided by a grid search [29]. In this experiment, we conducted 15-fold cross-validation and compared the performance of our method with the performances of comparative methods by using MAE and MZE values.

4.3. Experimental Results and Discussion

Results presented in Table 1 show the effectiveness of SFEMCCA since we can see that MAE and MZE of SFEMCCA are lower than those of all comparative methods for all the subjects. Moreover, by comparing sMVCCA with DLPCCA and SLPCCA, we can confirm that using label features directly as one of the features is highly effective. Then we adopted Welch's t-test to determine whether the difference between "MAE and MZE values of SFEMCCA" and "those of comparative methods" was significant or not. As a result, it was confirmed that accuracy improvements of MAE and MZE using SFEMCCA were statistically significant under the significance level 0.01 compared to all of the conventional methods of supervised multiview CCA.

Furthermore, we conducted an experiment to confirm whether the fractional-order technique worked effectively in small number of samples. In Fig. 1, we show the error corrections of η_d (i.e., eigenvalues) in Eq. (11), where "GT" means the eigenvalues calculated by all samples extracted in these experiments (900 samples), "nonFrac" means those calculated by one third of them (300 samples), and "Frac" means those calculated by 300 samples but the

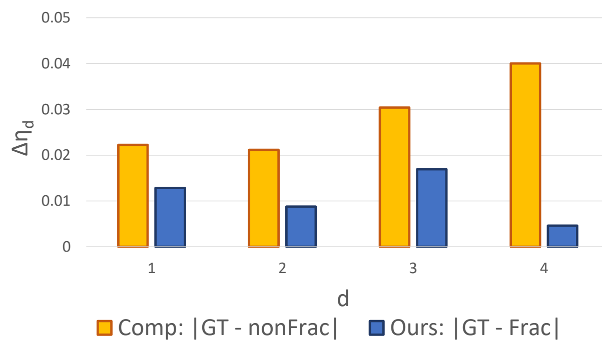

Fig. 1. The error corrections of η_d by using fractional-order technique of SFEMCCA.

covariances are corrected by two kinds of fractional-order parameters (ξ and ζ) in SFEMCCA. The setting to use "one third" follows our previous works [12, 13]. Specifically, the number of samples used to calculate projections is one third, and the number of samples used for training and test by SVORIM is two thirds (600 samples). If the differences of eigenvalues between "GT" and "Frac" (i.e., |GT − Frac|) are lower than those between "GT" and "nonFrac" (i.e., |GT − nonFrac|), it can be expected that fractional-order technique works effectively. As shown in Fig. 1, it can be confirmed that the technique worked effectively since |GT − Frac| were close to zero compared to |GT − nonFrac|. By suppressing eigenvalues of each covariance matrix, ones obtained by solving the generalized eigenvalue problem in Eq. (11) were corrected to approach "GT". These error corrections worked effectively for classification of evaluation scores for the videos. Thus, the results indicated that SFEMCCA could extract the canonical video features more accurately.

5. CONCLUSIONS

In this paper, we presented supervised fractional-order embedding multiview canonical correlation analysis (SFEMCCA). SFEMCCA is a method that newly introduces the fractional-order technique into sMVCCA. Our method (i.e., SFEMCCA) suppresses the increase in eigenvalues of covariance matrices calculated by data with noise, small number of samples and large number of dimensions. In the real world, it is necessary to deal with such the data, and there are many cases where three or more kinds of multimodal and supervised data is treated in order to calculate more accurate projections. Therefore, SFEMCCA is effective for data obtained from real environments. Since our previous researches (e.g., estimation of personalized preference for video) also includes the above-mentioned points, experimental results showed effectiveness of applying SFEMCCA to them.

6. REFERENCES

[1] H. Hotelling, “Relations between two sets of variates,” *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[2] T.-K. Sun and S.-C. Chen, “Locality preserving CCA with applications to data visualization and pose estimation,” *Image and Vision Computing*, vol. 25, no. 5, pp. 531–543, 2007.

[3] T.-K. Sun, S.-C. Chen, Z. Jin, and J.-Y. Yang, “Kernelized discriminative canonical correlation analysis,” in *Proc. the IEEE Int. Conf. Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2007, vol. 3, pp. 1283–1287.

[4] X. He and P. Niyogi, “Locality preserving projections,” in *Proc. the Advances in Neural Information Processing Systems (NIPS)*, 2004, pp. 153–160.

[5] J. Yang and X. Zhang, “Feature-level fusion of fingerprint and finger-vein for personal identification,” *Pattern Recognition Letters*, vol. 33, no. 5, pp. 623–628, 2012.

[6] K. Fukunaga, “Introduction to statistical pattern recognition,” *Academic Press*, 1990.

[7] X. Zhang, N. Guan, Z. Luo, and L. Lan, “Discriminative locality preserving canonical correlation analysis,” in *Proc. Chinese Conf. Pattern Recognition (CCPR)*, 2012, pp. 341–349.

[8] R. Sawata, T. Ogawa, M. Haseyama, “Novel audio feature projection using KDLPCCA-based correlation with EEG features for favorite music classification,” *IEEE Trans. Affective Computing*, 2017 (Accepted for publication).

[9] J. R. Kettenring, “Canonical analysis of several sets of variables,” *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[10] A. A. Nielsen, “Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data,” *IEEE Trans. Image Processing*, vol. 1, pp. 293–305, 2002.

[11] G. Lee, A. Singanamalli, H. Wang, M. D. Feldman, S. R. Master, N. N. C. Shih, E. Spangler, T. Rebbeck, J. E. Tomaszewski, and A. Madabhushi, “Supervised multi-view canonical correlation analysis (sMVCCA): Integrating histologic and proteomic features for predicting recurrent prostate cancer,” *IEEE Trans. Medical Imaging*, vol. 34, no. 1, pp. 284–297, 2015.

[12] Y. Ito, T. Ogawa, and M. Haseyama, “Personalized video preference estimation based on early fusion using multiple users’ viewing behavior,” in *Proc. the IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 3006–3010.

[13] Y. Ito, T. Ogawa, and M. Haseyama, “Preference estimation for video recommendation using DCNN features and viewing behavior,” in *Proc. the IEEE Global Conf. Consumer Electronics (GCCE)*, 2017, pp. 295–296.

[14] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[15] A. Hendrikse, L. Spreeuwers, and R. Veldhuis, “A bootstrap approach to eigenvalue correction,” in *Proc. the IEEE Int. Conf. Data Mining (ICDM)*, 2009, pp. 818–823.

[16] K. Yata and M. Aoshima, “Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations,” *Journal of Multivariate Analysis*, vol. 105, no. 1, pp. 193–215, 2012.

[17] J. Liu, S.-C. Chen, and X. Tan, “Fractional order singular value decomposition representation for face recognition,” *Pattern Recognition*, vol. 41, no. 1, pp. 378–395, 2008.

[18] Y.-H. Yuan, Q.-S. Sun, and H.-W. Ge, “Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition,” *Pattern Recognition*, vol. 47, no. 3, pp. 1411–1424, 2014.

[19] Y.-H. Yuan and Q.-S. Sun, “Fractional-order embedding multiset canonical correlations with applications to multi-feature fusion and recognition,” *Neurocomputing*, vol. 122, pp. 229–238, 2013.

[20] D. R. Hardoon, S. Sandor, and S.-T. John, “Canonical correlations analysis: An overview with application to learning methods,” *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. the Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[22] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrel, “DeCAF: A deep convolutional activation feature for generic visual recognition,” in *Proc. the IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 647–655.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. the IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proc. the ACM Int. Conf. Multimedia (ACMMM)*, 2014, pp. 675–678.

[25] O. Lartillot and P. Toiviainen, “A Matlab toolbox for musical feature extraction from audio,” in *Proc. Int. Conf. Digital Audio Effects (DAFX)*, 2007, pp. 237–244.

[26] J. F. Cohn, Z. Ambadar, and P. Ekman, “Observer-based measurement of facial expression with the facial action coding system,” *The Handbook of Emotion Elicitation and Assessment*, 2007, pp. 203–221.

[27] W. Chu and S. S. Keerthi, “Support vector ordinal regression,” *Neural Computation*, vol. 19, no. 3, pp. 792–815, 2007.

[28] P. A. Gutierrez, M. P.-Ortiz, J. S.-Monedero, F. F.-Navarro, and C. H.-Martinez, “Ordinal regression methods: Survey and experimental study,” *IEEE Trans. Knowledge and Data Engineering*, vol. 28, no. 1, pp. 127–146, 2016.

[29] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A practical guide to support vector classification,” *Technical Report, Department of Computer Science, National Taiwan University*, 2003, pp. 1–16.