IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 20, 2019, accepted October 7, 2019, date of publication October 11, 2019, date of current version October 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946912

Interest Level Estimation Based on Tensor Completion via Feature Integration for Partially Paired User's Behavior and Videos

TETSUYA KUSHIMA[1], (Student Member, IEEE), SHO TAKAHASHI[2], (Member, IEEE), TAKAHIRO OGAWA[3], (Senior Member, IEEE), AND MIKI HASEYAMA[3], (Senior Member, IEEE)

[1]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
[2]Faculty of Engineering, Hokkaido University, Sapporo 060-8628, Japan
[3]Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

Corresponding author: Tetsuya Kushima (kushima@lmd.ist.hokudai.ac.jp)

This work was partly supported by JSPS KAKENHI Grant Number JP17H01744 and the MIC/SCOPE #181601001.

ABSTRACT A novel method for interest level estimation based on tensor completion via feature integration for partially paired users' behavior and videos is presented in this paper. The proposed method defines a novel canonical correlation analysis (CCA) framework that is suitable for interest level estimation, which is a hybrid version of semi-supervised CCA (SemiCCA) and supervised locality preserving CCA (SLPCCA) called semi-supervised locality preserving CCA (S2LPCCA). For partially paired users' behavior and videos in actual shops and on the Internet, new integrated features that maximize the correlation between partially paired samples by the principal component analysis (PCA)-mixed CCA framework are calculated. Then videos that users have not watched can be used for the estimation of users' interest levels. Furthermore, local structures of partially paired samples in the same class are preserved for accurate estimation of interest levels. Tensor completion, which can be applied to three contexts, videos, users and "canonical features and interest levels," is used for estimation of interest levels. Consequently, the proposed method realizes accurate estimation of users' interest levels based on S2LPCCA and the tensor completion from partially paired training features of users' behavior and videos. Experimental results obtained by applying the proposed method to actual data show the effectiveness of the proposed method.

INDEX TERMS Feature integration, S2LPCCA, user behavior, interest level estimation, tensor completion.

I. INTRODUCTION

Interest level estimation is a technique that is useful in marketing for customer-oriented companies and users of their contents [1]–[7]. Interest level estimation means predictions of users' evaluations of various contents such as items in shops and videos on the Internet. The use of interest level estimation would enable companies to analyze their users' shopping behavior and establish an effective strategy for selling their contents because they would know their users' interests in their contents [1]. Furthermore, the user can discover contents that the user prefers and notice new preferences by using interest level estimation for content recommendation [2], [3]. Therefore, studies on estimation of users' interest levels have been extensively carried out [8]–[13].

Studies on interest level estimation have had different aims [8]–[11]. First, several methods for interest level estimation in actual shops have been proposed [8], [9]. Liu *et al.* focused on data for users' behavior obtained from surveillance cameras for estimating users' interests in actual shops [8]. They assumed that specific movements such as viewing and picking up items in the shop indicate a high level of interest, and such movements were extracted from the behavior data. Wang *et al.* proposed a system that can analyze users' shopping behavior by using shopping carts that sense users' movements in actual shops [9]. From their

The associate editor coordinating the review of this manuscript and approving it for publication was Lu An.

studies [8], [9], it was confirmed that data for users' behavior obtained by sensors are effective for interest level estimation. Secondly, several methods for investigating users' behavior that can effectively express users' interests have been proposed [10], [11]. Ma *et al*. focused on eye movements when watching a video and estimated video frames that users showed interest in by using users' eye movements [10]. That study was based on the knowledge that eye movement information is useful for revealing observers' interests [14]. Ding *et al*. extracted electroencephalography (EEG) features and tagged users' emotions when users were watching a video. EEG is one of the most popular and accessible neural signal measurement techniques [11]. From their studies [10], [11], it was confirmed that users are likely to express their interests with behavior when watching a video as contents. The above-described methods [8]–[11] have shown that there is a strong relationship between users' behavior and users' interests, and valuable results of experiments on interest level estimation were obtained in those studies. However, the problem of a limited number of samples, which is often a problem in the real world, was not considered in those studies. Therefore, we consider a method that can solve this problem.

For accurate interest level estimation in the real world, consideration must be given to the number of samples, *e.g.*, information on users' behavior and information on videos. The number of videos that one user has watched would generally be much smaller than the number of videos in shops and the number of videos on the Internet. Since data for the users' behavior can only be obtained if the users watch videos, most of the videos cannot be paired with users' behavior. In a method that does not take into account the limited number of samples, overfitting to few samples, *i.e.*, videos paired with the users' behavior, may occur. In order to solve this problem, construction of a method that can effectively use videos that the users have not watched is required. In the research fields of domain adaptation and multivariate analysis, methods that can handle incomplete samples have been proposed [15]–[24]. The incomplete samples are usually called "semi-paired" or "partially paired" samples in these methods. Mehrkanoon *et al*. proposed a method for canonical correlation analysis (CCA) called regularized semi-paired kernel CCA (RSP-KCCA) that can be applied to semi-paired samples consisting of different domains [15]. CCA is an efficient method for feature integration in two different variables [25]. Guo *et al*. proposed joint intermodal and intramodal semi-paired CCA (I^2SCCA) that can preserve within-view similarity and cross-view correlation without class information [16]. Kimura *et al*. proposed a semi-supervised CCA (SemiCCA) that can handle semi-paired samples while maintaining a simple structure of CCA by introducing a framework of principal component analysis (PCA) into CCA framework [17]. These methods achieve effective feature integration based on partially paired samples. Especially, since the final application in our study is the estimation of users' interests, construction of a framework that can effectively handle class information such as interest levels is required. Yang *et al*. proposed supervised locality preserving CCA (SLPCCA) [18] as a method can increase the separation performance between classes by preserving local structures in the same class. By incorporating SLPCCA with SemiCCA, which has a formula structure in which other elements can be easily incorporated, we consider that a method that is closest to the purpose of our study is achieved. Accurate estimation of users' interest levels based on partially paired users' behavior data and videos can be expected by using feature integration combining SemiCCA and SLPCCA.

The above mainly describes features obtained from users' behavior and videos. Below, we introduce a suitable estimation method. Recently, methods for estimating unknown values with other known data have been proposed [26]–[29]. Unknown values and known data indicate targeted users' interest levels and features obtained from the users' behavior and videos, respectively, in the case of interest level estimation. Song *et al*. used tensor completion that can complete unknown entries in data constructed with a tensor [26]. Specifically, users' interests can be estimated by construction of a tensor that consists of interest levels and various contexts. Liao *et al*. constructed tensors by using highway traffic data and predicted the traffic by applying tensor completion to the tensors based on traffic data [27]. Since a tensor can consist of various contexts, users' behavior-based data are suitable for tensor completion. Therefore, by using tensor completion and a hybrid version of SemiCCA and SLPCCA with partially paired users' behavior data and videos, we can realize interest level estimation that can accurately estimate users' interests and effectively use users' behavior-based data.

In this paper, we present a novel method for interest level estimation based on tensor completion via feature integration for partially paired users' behavior and videos. A CCA framework that is suitable for interest level estimation using data for users' behavior, called semi-supervised locality preserving CCA (S2LPCCA), is used in the proposed method. S2LPCCA is a hybrid version of the factors of SemiCCA and SLPCCA. Thus, S2LPCCA is suitable for both partially paired samples and interest level estimation. The flow of our interest level estimation using S2LPCCA is described below. First, we extract features from users' behavior when watching videos obtained by several sensors and these videos, and S2LPCCA is applied to the extracted features, which are partially paired samples as shown Fig. 1. For partially paired users' behavior and videos, new integrated features that maximize the correlation between such partially paired samples by a PCA-mixed CCA framework are calculated as canonical features in S2LPCCA. Then we can use the videos that users have not watched for estimation of users' interest levels. Furthermore, the separation performance between classes of these calculated canonical features is increased by preserving local structures in the same class for accurate estimation of interest levels obtained from users watching videos in S2LPCCA. The classes indicate users' interest levels in the proposed method. Secondly, the proposed method

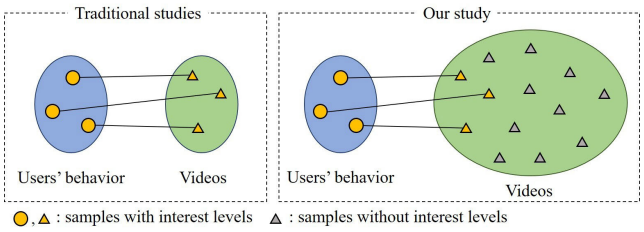

FIGURE 1. Samples used in traditional studies and our study. Triangles that are not connected with circles indicate videos that do not pair with users' behavior data. These paired samples are labeled by users' interest levels.

applies tensor completion to the canonical features and the interest levels. Specifically, three-dimensional tensors for which modes represent videos, users and "canonical features and interest levels" are constructed. By applying the tensor completion to the above tensors, the unknown values are recovered in the tensors, *i.e.*, the unknown interest levels can be estimated for videos that users have not watched. Consequently, the proposed method realizes accurate estimation of users' interest levels by using S2LPCCA and the tensor completion with partially paired users' behavior and videos.

The remainder of this paper is organized as follows. In section II, some related works are described. In section III, we explain features used in the proposed method. In section IV, the proposed method consisting of S2LPCCA and the tensor completion is explained. In section V, experimental results are shown for verifying the effectiveness of the proposed method. Finally, conclusions are given in section VI.

II. RELATED WORKS

In this section, we describe mathematical formulas in some related works, including SemiCCA and SLPCCA in II-A and the tensor completion in II-B.

A. SEMICCA AND SLPCCA

First, we explain CCA, which is a base method of SemiCCA and SLPCCA. CCA calculates projection vectors for maximizing the correlation between paired variables. For instance, paired variables $\boldsymbol{X}_P$ and $\boldsymbol{Y}_P$ are defined as follows:

$$\boldsymbol{X}_P = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{N_{\text{pair}}}] \in \mathbb{R}^{d_x \times N_{\text{pair}}}, \tag{1}$$

$$\boldsymbol{Y}_P = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{N_{\text{pair}}}] \in \mathbb{R}^{d_y \times N_{\text{pair}}}, \tag{2}$$

where N_{pair} is the number of paired variables, and d_x and d_y are the numbers of dimensions of $\boldsymbol{X}_P$ and $\boldsymbol{Y}_P$, respectively. To calculate the projection vectors $\hat{\boldsymbol{w}}_x \in \mathbb{R}^{d_x}$ and $\hat{\boldsymbol{w}}_y \in \mathbb{R}^{d_y}$, the following generalized eigenvalue problem is solved:

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}_P\boldsymbol{Y}_P^{\mathrm{T}} \\ \boldsymbol{Y}_P\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{X}_P\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_P\boldsymbol{Y}_P^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix}. \tag{3}$$

Next, we explain SemiCCA. SemiCCA can be applied to partially paired variables and calculates projection vectors for maximizing the correlation between partially paired variables. For instance, partially paired variables are $\boldsymbol{X} = [\boldsymbol{X}_P, \boldsymbol{X}_U]$ and $\boldsymbol{Y} = [\boldsymbol{Y}_P, \boldsymbol{Y}_U]$, and $\boldsymbol{X}_U$ and $\boldsymbol{Y}_U$ are defined as follows:

$$\boldsymbol{X}_U = [\boldsymbol{x}_{U,1}, \boldsymbol{x}_{U,2}, \cdots, \boldsymbol{x}_{U,N_x}] \in \mathbb{R}^{d_x \times N_x}, \tag{4}$$

$$\boldsymbol{Y}_U = [\boldsymbol{y}_{U,1}, \boldsymbol{y}_{U,2}, \cdots, \boldsymbol{y}_{U,N_y}] \in \mathbb{R}^{d_y \times N_y}, \tag{5}$$

where N_x and N_y are the numbers of variables of $\boldsymbol{X}_U$ and $\boldsymbol{Y}_U$, respectively. To calculate the projection vectors $\hat{\boldsymbol{w}}_x \in \mathbb{R}^{d_x}$ and $\hat{\boldsymbol{w}}_y \in \mathbb{R}^{d_y}$, the following generalized eigenvalue problem is solved:

$$\begin{aligned} &\left(\beta \begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}_P\boldsymbol{Y}_P^{\mathrm{T}} \\ \boldsymbol{Y}_P\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{S}_{xx} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{yy} \end{bmatrix}\right) \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix} \\ &= \lambda \left(\beta \begin{bmatrix} \boldsymbol{X}_P\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_P\boldsymbol{Y}_P^{\mathrm{T}} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{I}_{d_x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_y} \end{bmatrix}\right) \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix}, \end{aligned} \tag{6}$$

where $\boldsymbol{S}_{xx} = \boldsymbol{X}_P\boldsymbol{X}_P^{\mathrm{T}} + \boldsymbol{X}_U\boldsymbol{X}_U^{\mathrm{T}}$, $\boldsymbol{S}_{yy} = \boldsymbol{Y}_P\boldsymbol{Y}_P^{\mathrm{T}} + \boldsymbol{Y}_U\boldsymbol{Y}_U^{\mathrm{T}}$, and β is a parameter that controls the trade-off between CCA and PCA. SemiCCA can utilize additional unpaired samples by smoothly bridging CCA for only the paired samples and PCA, one of the major tools to capture the global structure of samples in an unsupervised manner [17].

Finally, we explain SLPCCA. SLPCCA can increase the separation performance between classes by preserving local structures in the same class and it calculates projection vectors for maximizing the correlation between paired variables with preservation of local structures. For instance, the above $\boldsymbol{X}_P$ and $\boldsymbol{Y}_P$ are given as paired variables. To calculate the projection vectors $\hat{\boldsymbol{w}}_x \in \mathbb{R}^{d_x}$ and $\hat{\boldsymbol{w}}_y \in \mathbb{R}^{d_y}$, the following generalized eigenvalue problem is solved:

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}_P\boldsymbol{L}_{xy}\boldsymbol{Y}_P^{\mathrm{T}} \\ \boldsymbol{Y}_P\boldsymbol{L}_{xy}\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix} \\ &\qquad = \lambda \begin{bmatrix} \boldsymbol{X}_P\boldsymbol{L}_{xx}\boldsymbol{X}_P^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_P\boldsymbol{L}_{yy}\boldsymbol{Y}_P^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{w}}_x \\ \hat{\boldsymbol{w}}_y \end{bmatrix}, \end{aligned} \tag{7}$$

where $\boldsymbol{L}_{xy}$, $\boldsymbol{L}_{xx}$ and $\boldsymbol{L}_{yy}$ are calculated based on the similarity between variables [18]. SLPCCA aims at keeping the local discriminative information of samples that belong to the same class.

B. TENSOR COMPLETION

First, we explain matrix completion [30], which is a base method of the tensor completion. Matrix completion via rank minimization is a well-known method that can complete unknown entries with other known ones. In the matrix completion, unknown entries are completed by solving the following optimization problem:

$$\begin{aligned} &\arg\min_{\boldsymbol{X}_{MC}} \ \operatorname{rank}(\boldsymbol{X}_{MC}), \\ &\text{s.t. } X_{MC}(p_1, p_2) = M_{MC}(p_1, p_2), \quad (p_1, p_2) \in \Psi, \end{aligned} \tag{8}$$

where $\boldsymbol{X}_{MC}$ and $\boldsymbol{M}_{MC}$ are a completed matrix and an incomplete matrix, respectively. Moreover, $X_{MC}(p_1, p_2)$ and $M_{MC}(p_1, p_2)$ represent (p_1, p_2)-th entries of $\boldsymbol{X}_{MC}$ and $\boldsymbol{M}_{MC}$, respectively, and Ψ is the set of locations corresponding to known entries. This optimization problem can be solved by

TABLE 1. User behavior features used in the proposed method.

	Features	Dimensions
OpenPose	Means and variances over movements of the nose, neck and center of the hip positions	12
	Means and variances over movements of both eyes, ears, shoulders, elbows, wrists and hips	48
Eye Tracker	Means and variances over movements of gaze positions	4
Total		**64**

minimizing the truncated nuclear norm of $\boldsymbol{X}_{MC}$ [31]. Therefore, this optimization problem is rewritten as the following formula [31]:

$$\begin{aligned} &\arg\min_{\boldsymbol{X}_{MC}} \ ||\boldsymbol{X}_{MC}||_r, \\ &\text{s.t. } X_{MC}(p_1, p_2) = M_{MC}(p_1, p_2), \quad (p_1, p_2) \in \Psi, \end{aligned} \tag{9}$$

where $||\boldsymbol{X}_{MC}||_r = \sum_{sv=r+1}^{\min(MC_{\text{row}}, MC_{\text{column}})} \sigma_{sv}(\boldsymbol{X}_{MC})$ (MC_{row} and MC_{column} being the numbers of rows and columns of $\boldsymbol{X}_{MC}$, respectively), $\sigma_{sv}(\boldsymbol{X}_{MC})$ is the sv-th largest singular value of $\boldsymbol{X}_{MC}$, and r is a parameter related to the number of singular values. The matrix completion is a method limited to representation by two modes.

We next explain the tensor completion, which extends modes of the matrix completion. The tensor completion can complete unknown entries of a tensor for which represent N_{TC} kinds of contexts, where N_{TC} is the number of the tensor's modes. In the tensor completion via the above rank minimization, unknown entries are completed by solving the following optimization problem [32], [33]:

$$\begin{aligned} &\arg\min_{\boldsymbol{\mathcal{X}}_{TC}} \ \text{rank}(\boldsymbol{\mathcal{X}}_{TC}), \\ &\text{s.t. } \mathcal{X}_{TC}(p_1, p_2, \cdots, p_{N_{TC}}) = \mathcal{T}_{TC}(p_1, p_2, \cdots, p_{N_{TC}}), \\ &\quad (p_1, p_2, \cdots, p_{N_{TC}}) \in \Psi, \end{aligned} \tag{10}$$

where $\boldsymbol{\mathcal{X}}_{TC}$ and $\boldsymbol{\mathcal{T}}_{TC}$ are a completed tensor and an incomplete tensor, respectively. This optimization problem can be solved by applying the matrix completion to matrices unfolded along each mode [26]. Therefore, this optimization problem is rewritten as the following formula [33]:

$$\begin{aligned} &\arg\min_{\boldsymbol{\mathcal{X}}_{TC}} \ \sum_{n=1}^{N_{TC}} ||\boldsymbol{\mathcal{X}}_{TC,(n)}||_r, \\ &\text{s.t. } \mathcal{X}_{TC}(p_1, p_2, \cdots, p_{N_{TC}}) = \mathcal{T}_{TC}(p_1, p_2, \cdots, p_{N_{TC}}), \\ &\quad (p_1, p_2, \cdots, p_{N_{TC}}) \in \Psi, \end{aligned} \tag{11}$$

where $\boldsymbol{\mathcal{X}}_{TC,(n)}$ is an unfolded matrix along the n-th mode of $\boldsymbol{\mathcal{X}}_{TC}$. The tensor completion can be applied to data including three or more kinds of contexts unlike the matrix completion.

III. FEATURE EXTRACTION

In this section, we explain two kinds of features used in the proposed method. We show the extraction of user behavior features in III-A and extraction of content features in III-B.

A. USER BEHAVIOR FEATURES

Recently, methods that use data for user behavior obtained from sensors have been proposed [34]–[42]. OpenPose is the latest method for detecting body skeleton positions as 2D data and can accurately detect them by using deep neural networks based on an affinity for body parts [34]. Furthermore, Tobii Eye Tracker 4C[1] is a device that can detect users' eye-gaze positions, which is related to users' interests [35], [36], as 2D data. The convenience of acquisition of data for user behavior, *i.e.*, ease of use in the real world, was considered in those studies. OpenPose is a method that can be achieved with only cameras that can take color images [34], and Tobii Eye Tracker 4C is a low-cost consumer-level remote eye tracker [37]–[39]. Therefore, since these methods and devices are suitable for the interest level estimation in the real world, we use them in this study.

In order to extract user behavior features, we use data for the positions obtained by OpenPose and Tobii Eye Tracker 4C. We then calculate means and variances over movements of those positions as user behavior features and obtain user behavior feature vector $\boldsymbol{f}_{i_P,j} \in \mathbb{R}^{d_f}$ for each i_P-th ($i_P = 1, 2, \cdots, I_P$; I_P being the number of videos that users have watched) video of each j-th ($j = 1, 2, \cdots, J$; J being the number of users) user, where $d_f = 64$ as shown in Table 1.

B. CONTENT FEATURES

Traditionally, methods that extract hand-crafted features and deep neural network (DNN)-based features from videos have been proposed [43]–[48]. As hand-crafted features, HSV color histogram (HSVCH) [43] and Bag of features [44] based on Speeded-Up Robust Features (SURF-Bof) [45] are often used. DNN-based features can be mainly obtained by using a DNN model trained by open datasets, called transfer learning [46]–[48]. Recently, since representation performances of DNN-based features are high, DNN-based features have been used in various studies on video classification, captioning and retrieval [49]–[56]. The Inception-v3 model [46] is well known in the above research field. Therefore, since DNN-based features obtained with the Inception-v3 model are suitable for accurate interest level estimation, we use them in this study.

In order to extract content features, we use videos as contents. We obtain 2,048-dimensional output vectors in the third pooling layer of Inception-v3 [46] by using TensorFlow[2] from each frame of these videos. Then we calculate mean vectors of the output vectors as content feature vectors $\boldsymbol{v}_{i_P,j} \in \mathbb{R}^{d_v}$ for each i_P-th video of each j-th user and $\tilde{\boldsymbol{v}}_{i_U,j} \in \mathbb{R}^{d_v}$ for each i_U-th ($i_U = 1, 2, \cdots, I_U$;I_U being the number of videos

[1] https://tobiigaming.com/eye-tracker-4c

[2] https://www.tensorflow.org

that users have not watched) video, where $d_v = 2,048$. Note that $\boldsymbol{v}_{i_P,j}$ is paired with $\boldsymbol{f}_{i_P,j}$, and $\tilde{\boldsymbol{v}}_{i_U,j}$ does not have paired user behavior features.

IV. INTEREST LEVEL ESTIMATION BASED ON TENSOR COMPLETION VIA S2LPCCA

In this section, our interest level estimation method based on tensor completion via S2LPCCA is presented. In IV-A, we explain feature integration based on S2LPCCA. In IV-B, estimation of unknown interest levels via the tensor completion is presented.

A. FEATURE INTEGRATION BASED ON S2LPCCA

In this subsection, we explain our feature integration based on S2LPCCA. In the proposed method, canonical features are calculated by applying feature integration based on S2LPCCA to the user behavior features and content features obtained in III-A and III-B, respectively.

First, we define matrices $\boldsymbol{F}_j$, $\boldsymbol{V}_j$ and $\tilde{\boldsymbol{V}}_j$ as follows:

$$\begin{aligned}\boldsymbol{F}_j = [&\boldsymbol{f}^{(1)}_{1,j}, \boldsymbol{f}^{(1)}_{2,j}, \cdots, \boldsymbol{f}^{(1)}_{n_1,j},\\ &\cdots, \boldsymbol{f}^{(C)}_{1,j}, \boldsymbol{f}^{(C)}_{2,j}, \cdots, \boldsymbol{f}^{(C)}_{n_C,j}],\end{aligned} \tag{12}$$

$$\begin{aligned}\boldsymbol{V}_j = [&\boldsymbol{v}^{(1)}_{1,j}, \boldsymbol{v}^{(1)}_{2,j}, \cdots \boldsymbol{v}^{(1)}_{n_1,j},\\ &\cdots, \boldsymbol{v}^{(C)}_{1,j}, \boldsymbol{v}^{(C)}_{2,j}, \cdots, \boldsymbol{v}^{(C)}_{n_C,j}],\end{aligned} \tag{13}$$

$$\tilde{\boldsymbol{V}}_j = [\tilde{\boldsymbol{v}}_{1,j}, \tilde{\boldsymbol{v}}_{2,j}, \cdots, \tilde{\boldsymbol{v}}_{I_U,j}], \tag{14}$$

where n_c is the number of samples that users have watched in the c-th class ($c = 1, 2, \cdots, C$; C being the number of classes corresponding to interest levels). It is assumed that these matrices are centered. To arrange these features in the class information, we respectively redefine $\boldsymbol{f}^{(c)}_{g,j} \in \mathbb{R}^{d_f}$ and $\boldsymbol{v}^{(c)}_{g,j} \in \mathbb{R}^{d_v}$ as the g-th sample of the c-th class obtained from the j-th user in Eqs. (12) and (13). Moreover, we define matrices $\boldsymbol{F}_P$, $\boldsymbol{V}_P$ and $\boldsymbol{V}_U$ by using $\boldsymbol{F}_j$, $\boldsymbol{V}_j$ and $\tilde{\boldsymbol{V}}_j$ as follows:

$$\boldsymbol{F}_P = [\boldsymbol{F}_1, \boldsymbol{F}_2, \cdots, \boldsymbol{F}_j, \cdots, \boldsymbol{F}_J], \tag{15}$$

$$\boldsymbol{V}_P = [\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_j, \cdots, \boldsymbol{V}_J], \tag{16}$$

$$\boldsymbol{V}_U = [\tilde{\boldsymbol{V}}_1, \tilde{\boldsymbol{V}}_2, \cdots, \tilde{\boldsymbol{V}}_j, \cdots, \tilde{\boldsymbol{V}}_J]. \tag{17}$$

For integrating these features with class information, we first define similarity matrices $\boldsymbol{S}_f = \{S^f_{x,y}\}^N_{x,y=1}$ ($N = I_P \times J$) and $\boldsymbol{S}_v = \{S^v_{x,y}\}^N_{x,y=1}$ by using only $\boldsymbol{F}_P$ and $\boldsymbol{V}_P$. $S^f_{x,y}$ and $S^v_{x,y}$ are defined as follows:

$$S^f_{x,y} = \begin{cases} e^{-\|\boldsymbol{f}_x - \boldsymbol{f}_y\|^2/t_f} & \boldsymbol{f}_y \in \Omega_{\boldsymbol{f}_x} \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

$$S^v_{x,y} = \begin{cases} e^{-\|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2/t_v} & \boldsymbol{v}_y \in \Omega_{\boldsymbol{v}_x} \\ 0 & \text{otherwise}, \end{cases} \tag{19}$$

where $\Omega_{\boldsymbol{f}_x}$ and $\Omega_{\boldsymbol{v}_x}$ are the sets of k nearest neighbor of $\boldsymbol{f}_x$ and $\boldsymbol{v}_x$, respectively. Also, t_f and t_v are defined as follows:

$$t_f = \frac{1}{N(N-1)} \sum_{x=1}^{N} \sum_{y=1}^{N} \|\boldsymbol{f}_x - \boldsymbol{f}_y\|^2, \tag{20}$$

$$t_v = \frac{1}{N(N-1)} \sum_{x=1}^{N} \sum_{y=1}^{N} \|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2. \tag{21}$$

By using these similarity matrices, the proposed method preserves local structures of $\boldsymbol{F}_P$ and $\boldsymbol{V}_P$. Furthermore, the proposed method redefines similarity matrices $\tilde{\boldsymbol{S}}_f$ and $\tilde{\boldsymbol{S}}_v$ in order to use class information as follows:

$$\tilde{\boldsymbol{S}}_f = \boldsymbol{S}_f \circ \boldsymbol{A}, \tag{22}$$

$$\tilde{\boldsymbol{S}}_v = \boldsymbol{S}_v \circ \boldsymbol{A}, \tag{23}$$

where "$\circ$" denotes the Hadamard product, and $\boldsymbol{A}$ is a blocked diagonal matrix as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{1}_{n_1 \times n_1} & & & 0 \\ & \ddots & & \\ & & \boldsymbol{1}_{n_x \times n_x} & \\ & & & \ddots \\ 0 & & & \boldsymbol{1}_{n_C \times n_C} \end{bmatrix} \in \mathbb{R}^{N \times N}. \tag{24}$$

Next, we calculate projection vectors $\hat{\boldsymbol{w}}_f$ and $\hat{\boldsymbol{w}}_v$ that maximize the correlation between variables $\boldsymbol{F}_P$ and $\boldsymbol{V}_P$ with $\boldsymbol{V}_U$ in order to integrate partially paired features. Specifically, we calculate $\hat{\boldsymbol{w}}_f$ and $\hat{\boldsymbol{w}}_v$ by solving the following generalized eigenvalue problem with the similarity matrices $\tilde{\boldsymbol{S}}_f$ and $\tilde{\boldsymbol{S}}_v$:

$$\begin{aligned}&\left(\beta \begin{bmatrix} \boldsymbol{0} & \boldsymbol{F}_P \boldsymbol{L}_{fv} \boldsymbol{V}_P^{\mathrm{T}} \\ \boldsymbol{V}_P \boldsymbol{L}_{fv} \boldsymbol{F}_P^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{F}_P \boldsymbol{F}_P^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{vv} \end{bmatrix}\right) \begin{bmatrix} \hat{\boldsymbol{w}}_f \\ \hat{\boldsymbol{w}}_v \end{bmatrix} \\ &= \lambda \left(\beta \begin{bmatrix} \boldsymbol{F}_P \boldsymbol{L}_{ff} \boldsymbol{F}_P^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_P \boldsymbol{L}_{vv} \boldsymbol{V}_P^{\mathrm{T}} \end{bmatrix} \right. \\ &\left. + (1-\beta) \begin{bmatrix} \boldsymbol{I}_{d_f} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_v} \end{bmatrix}\right) \begin{bmatrix} \hat{\boldsymbol{w}}_f \\ \hat{\boldsymbol{w}}_v \end{bmatrix},\end{aligned} \tag{25}$$

where $\boldsymbol{L}_{fv} = \boldsymbol{D}_{fv} - \tilde{\boldsymbol{S}}_f \circ \tilde{\boldsymbol{S}}_v$, $\boldsymbol{L}_{ff} = \boldsymbol{D}_{ff} - \tilde{\boldsymbol{S}}_f \circ \tilde{\boldsymbol{S}}_f$, and $\boldsymbol{L}_{vv} = \boldsymbol{D}_{vv} - \tilde{\boldsymbol{S}}_v \circ \tilde{\boldsymbol{S}}_v$. In addition, $\boldsymbol{D}_{fv}$, $\boldsymbol{D}_{ff}$ and $\boldsymbol{D}_{vv}$ are diagonal matrices, and (x, x)-th elements in these matrices are the sum of the x-th row elements of matrices $\tilde{\boldsymbol{S}}_f \circ \tilde{\boldsymbol{S}}_v$, $\tilde{\boldsymbol{S}}_f \circ \tilde{\boldsymbol{S}}_f$ and $\tilde{\boldsymbol{S}}_v \circ \tilde{\boldsymbol{S}}_v$, respectively. Furthermore, $\boldsymbol{S}_{vv} = \boldsymbol{V}_P \boldsymbol{V}_P^{\mathrm{T}} + \boldsymbol{V}_U \boldsymbol{V}_U^{\mathrm{T}}$, and β is a parameter that controls the trade-off between SLPCCA and PCA. We then calculate projection matrix $\boldsymbol{W}_v = [\hat{\boldsymbol{w}}^1_v, \hat{\boldsymbol{w}}^2_v, \cdots, \hat{\boldsymbol{w}}^d_v]$ and obtain canonical features $\hat{\boldsymbol{V}}_j$ by using this matrix as follows:

$$\hat{\boldsymbol{V}}_j = \boldsymbol{W}_v^{\mathrm{T}} \boldsymbol{V}'_j, \tag{26}$$

where d is the dimension of the canonical features and satisfies $d \leq \min(d_f, d_v)$, and $\boldsymbol{V}'_j = [\boldsymbol{V}_j, \tilde{\boldsymbol{V}}_j]$. Consequently, the proposed method can calculate the features that are effective for interest level estimation from partially paired samples with class information by using S2LPCCA. Furthermore, by restraining the overfitting to such a small number of samples with S2LPCCA, accurate estimation of interest levels can be expected.

B. ESTIMATION OF UNKNOWN INTEREST LEVELS VIA TENSOR COMPLETION

In this subsection, we explain the estimation of unknown interest levels via the tensor completion. First, the proposed method constructs the following three-dimensional tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ by using I $(= I_P + I_U)$ videos, J users, and K $(= d + 1)$ elements including the canonical features and their corresponding interest levels:

$$\begin{cases} \mathcal{T}_{i,j,k} = l_{i,j} & (\text{if } k = 1) \\ \mathcal{T}_{i,j,k} = \hat{V}_j(k-1, i) & (\text{if } k = 2, \cdots, K), \end{cases} \tag{27}$$

where $l_{i,j}$ is the j-th user's interest levels for the i-th $(i = 1, 2, \cdots, I)$ video, and $l_{i,j}$ corresponding to the videos that the j-th user has not watched are unknown. In Eq. (27), $\hat{V}_j(k-1, i)$ represents $(k-1, i)$-th elements of $\hat{\boldsymbol{V}}_j$.

Next, the proposed method estimates unknown interest levels by applying the tensor completion to the tensor constructed in Eq. (27). In the tensor completion, the following optimization problem is solved:

$$\begin{aligned} &\arg\min_{\boldsymbol{\mathcal{X}}} \ \text{rank}(\boldsymbol{\mathcal{X}}), \\ &\text{s.t. } \mathcal{X}(p_1, p_2, p_3) = \mathcal{T}(p_1, p_2, p_3), \\ &\quad (p_1, p_2, p_3) \in \Psi, \end{aligned} \tag{28}$$

where $\boldsymbol{\mathcal{X}}$ is a completed tensor. Moreover, $\mathcal{X}(p_1, p_2, p_3)$ and $\mathcal{T}(p_1, p_2, p_3)$ represent (p_1, p_2, p_3)-th entries of $\boldsymbol{\mathcal{X}}$ and $\mathcal{T}$, respectively, and Ψ is the set of locations corresponding to known entries. This optimization problem can be solved by applying the matrix completion to matrices unfolded along each mode [26]. Therefore, this optimization problem is rewritten as the following formula [33]:

$$\begin{aligned} &\arg\min_{\boldsymbol{\mathcal{X}}} \ \sum_{n=1}^{3} ||\boldsymbol{\mathcal{X}}_{(n)}||_r, \\ &\text{s.t. } \mathcal{X}(p_1, p_2, p_3) = \mathcal{T}(p_1, p_2, p_3), \\ &\quad (p_1, p_2, p_3) \in \Psi, \end{aligned} \tag{29}$$

where $\boldsymbol{\mathcal{X}}_{(n)}$ is an unfolded matrix along the n-th mode of $\boldsymbol{\mathcal{X}}$, $||\boldsymbol{\mathcal{X}}_{(n)}||_r = \sum_{sv=r+1}^{\min(\mathcal{X}_{\text{row}}, \mathcal{X}_{\text{column}})} \sigma_{sv}(\boldsymbol{\mathcal{X}}_{(n)})$ ($\mathcal{X}_{\text{row}}$ and $\mathcal{X}_{\text{column}}$ being the numbers of rows and columns of $\boldsymbol{\mathcal{X}}_{(n)}$, respectively), $\sigma_{sv}(\boldsymbol{\mathcal{X}}_{(n)})$ is the sv-th largest singular value of $\boldsymbol{\mathcal{X}}_{(n)}$, and r is a parameter related to the number of singular values. Consequently, the proposed method estimates unknown interest levels by the tensor completion. As described above, the tensor completion can estimate users' interest levels for videos regardless of whether they have been watched or not. Furthermore, the proposed method can estimate interest levels by using the tensor completion that can be applied to three contexts, videos, users and "canonical features and interest levels."

V. EXPERIMENTAL RESULTS

In this section, we show experimental results to verify the effectiveness of the proposed method. In V-A, the experimental conditions are explained. Results of the experiment that verify the effectiveness of using S2LPCCA and the tensor completion for interest level estimation are presented in V-B.

FIGURE 2. Environment of the experiment. A blue rectangle and a red rectangle indicate a Web camera for OpenPose and Tobii Eye Tracker 4C, respectively.

A. EXPERIMENTAL CONDITIONS

First, we explain the environment of the experiment. In this experiment, we obtained movie trailers of five genres ("action," "comedy," "music," "science" and "sports") from YouTube[3] as target contents based on [57]–[60], and the number of movie trailers in each genre was 10. Therefore, the total number of movie trailers was 50. Then we first gave the subjects 10 seconds as time to prepare, and we next showed these videos on a display to the subjects in the following order:

1) Watching one video for 30 seconds
2) Evaluating the video in four classes[4] for 5 seconds
3) Repeating 1) and 2) until the subjects had watched all of the videos

The environment of the experiment is shown in Fig. 2. The subjects were eight men and two women who were approximately 22 years old. We obtained data based on their behavior when they were watching each video by OpenPose and Tobii Eye Tracker 4C. Then we estimated the subjects' interest levels by the proposed method with the videos and the subjects' behavior obtained in this environment.

Next, we explain the verification method to confirm the effectiveness of the proposed method. In this experiment, we randomly selected $10, 20, \cdots, 80, 90\%$ of videos from all videos and defined them as videos that subjects had not watched. Experimental results with low percentages and high percentages of videos that subjects had not watched are shown for verifying the effectiveness of the proposed method for data including many videos that are paired with users' behavior and many videos that are not paired with users' behavior, respectively. By estimating these unknown interest

[3] https://www.youtube.com

[4] 4 (very interesting), 3 (a little interesting), 2 (not interesting), and 1 (not interesting at all), $C = 4$.

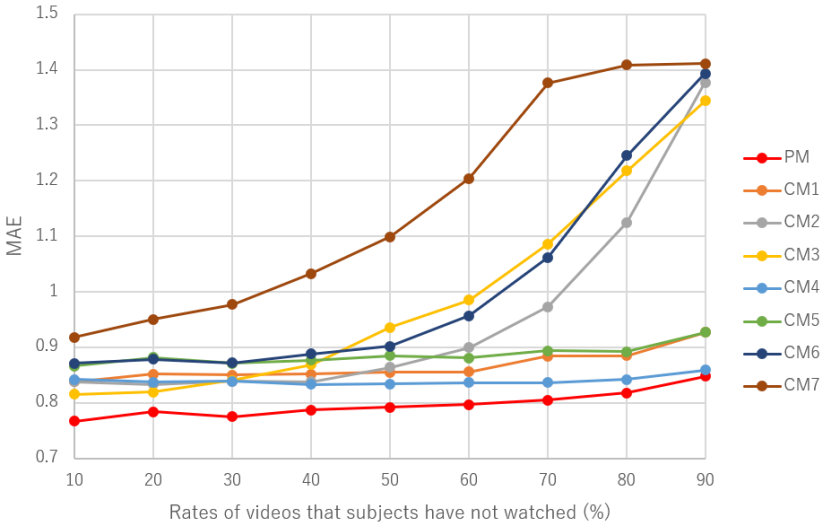

FIGURE 3. Results obtained by the PM and CMs1–7 for each rate of videos that subjects have not watched. The values are means of the subjects.

levels in various conditions, we verify the robustness of the proposed method.

We compared the proposed method (PM) with seven comparative methods (CMs1–7) for verifying the effectiveness of the proposed method. CM1, CM2 and CM3 estimate subjects' interest levels with features integrated by SLPCCA, SemiCCA and CCA, respectively. Specifically, CM1 does not use information of videos that the subjects have not watched, and we utilized CM1 for verifying the effectiveness of using information of videos that the subjects have not watched. CM2 does not use class information in the feature integration, and we utilized CM2 for verifying the effectiveness of using feature integration with class information. CM3 uses simple CCA, and we utilized CM3 for verifying the effectiveness of mixing SemiCCA and SLPCCA. In CMs1–3, unknown interest levels are estimated by the tensor completion. CM4 estimates unknown interest levels by the matrix completion with S2LPCCA. Specifically, the matrix completion can represent only two modes, and we utilized CM4 for verifying the effectiveness of the tensor completion that can represent three modes, *i.e.*, videos, users and ''canonical features and interest levels.'' Furthermore, CM5, CM6 and CM7 estimate unknown interest levels by the matrix completion with SLPCCA, SemiCCA and CCA, respectively. We utilized CMs5–7 for verifying whether the proposed method is effective regardless of schemes of the estimation.

To compare the performances of the PM and CMs1–7, we used mean absolute error (MAE) as follows:

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{miss}}} \sum_{s=1}^{N_{\mathrm{miss}}} |l_s^{\mathrm{PRE}} - l_s^{\mathrm{GT}}|, \tag{30}$$

where N_{miss} is the number of samples for which interest levels are missing, l_s^{PRE} is the predicted interest level of the s-th sample, and l_s^{GT} is its ground truth. In Eq. (30), the lower this measure is, the higher the performance of the method is. Furthermore, we also used standard deviation, and the number of trials of estimation was one hundred in each sample. Note that we set $\beta = 0.5$ and $r = 3$ providing the best performance in each method.

B. EXPERIMENTAL RESULTS OF INTEREST LEVEL ESTIMATION

The results of the experiment are presented in this subsection. We show the results obtained by the PM and CMs1–7 in Tables 2–4 and Fig. 3. In these tables, MAEs and the standard deviations in rates of videos that subjects had not watched are shown, and Fig. 3 shows transitions of the MAEs' means along rates of the videos that subjects had not watched in each method, where a and b of $a \pm b$ indicate MAE and standard deviation, respectively. First, we show the effectiveness and robustness of the PM by comparing the PM and CMs1–7 in three conditions that videos had not been watched by subjects. As shown in Tables 2–4, the results show the effectiveness of the PM since we can see that means of the MAEs of the PM are perfectly lower than those of CMs1–7, and standard deviations of the PM are almost lower than those of CMs1–7. Specifically, in Table 2, it is confirmed that the PM is effective in the condition that known data for interest level estimation are sufficient compared with unknown data. Conversely, in Table 4, it is confirmed that the PM is also effective in the condition that most of the data for interest level estimation are unknown, and this condition is similar to situations in which a user's interest levels for videos in actual shops and on the Internet are estimated. Therefore, since the PM is effective for various conditions, the PM is robust with regard to incomplete data and can be used in actual shops and on the Internet.

Furthermore, we show the effectiveness of S2LPCCA and the tensor completion by using the graph that the MAEs' mean is distributed in each rate. As shown in Fig. 3, the

TABLE 2. MAEs of the PM and CMs1–7 in the condition that 10% of videos have not been watched by subjects.

Subject	PM	CM1	CM2	CM3	CM4	CM5	CM6	CM7
A	0.677 ± 0.209	0.794 ± 0.276	0.757 ± 0.211	$\mathbf{0.668 \pm 0.208}$	0.695 ± 0.218	0.753 ± 0.237	0.733 ± 0.202	0.852 ± 0.276
B	$\mathbf{0.538 \pm 0.166}$	0.636 ± 0.225	0.650 ± 0.220	0.600 ± 0.210	0.634 ± 0.205	0.685 ± 0.214	0.709 ± 0.215	0.673 ± 0.265
C	$\mathbf{0.729 \pm 0.254}$	0.731 ± 0.265	0.772 ± 0.276	0.823 ± 0.268	1.025 ± 0.218	0.928 ± 0.295	0.943 ± 0.292	1.117 ± 0.305
D	0.964 ± 0.243	1.049 ± 0.264	1.036 ± 0.235	$\mathbf{0.947 \pm 0.279}$	1.041 ± 0.226	1.105 ± 0.265	1.089 ± 0.243	1.004 ± 0.317
E	$\mathbf{0.581 \pm 0.176}$	0.664 ± 0.229	0.603 ± 0.212	0.612 ± 0.240	0.675 ± 0.181	0.656 ± 0.197	0.649 ± 0.213	0.874 ± 0.311
F	0.976 ± 0.244	1.061 ± 0.269	1.042 ± 0.275	$\mathbf{0.949 \pm 0.252}$	0.999 ± 0.195	1.021 ± 0.275	1.033 ± 0.274	0.992 ± 0.284
G	$\mathbf{0.686 \pm 0.180}$	0.823 ± 0.271	0.830 ± 0.178	0.822 ± 0.244	0.742 ± 0.186	0.771 ± 0.272	0.843 ± 0.221	0.926 ± 0.315
H	0.844 ± 0.209	0.861 ± 0.249	0.902 ± 0.234	0.955 ± 0.304	$\mathbf{0.824 \pm 0.209}$	0.870 ± 0.237	0.885 ± 0.223	0.974 ± 0.309
I	$\mathbf{0.667 \pm 0.186}$	0.708 ± 0.231	0.744 ± 0.238	0.764 ± 0.219	0.729 ± 0.193	0.759 ± 0.243	0.747 ± 0.224	0.786 ± 0.266
J	1.009 ± 0.267	1.053 ± 0.287	1.048 ± 0.268	1.011 ± 0.306	1.060 ± 0.251	1.111 ± 0.272	1.080 ± 0.214	$\mathbf{0.977 \pm 0.356}$
Mean	**0.767**	0.838	0.838	0.815	0.842	0.866	0.871	0.918

TABLE 3. MAEs of the PM and CMs1–7 in the condition that 50% of videos have not been watched by subjects.

Subject	PM	CM1	CM2	CM3	CM4	CM5	CM6	CM7
A	$\mathbf{0.666 \pm 0.077}$	0.761 ± 0.099	0.778 ± 0.109	0.867 ± 0.166	0.686 ± 0.089	0.760 ± 0.114	0.811 ± 0.099	1.064 ± 0.223
B	$\mathbf{0.573 \pm 0.064}$	0.678 ± 0.100	0.693 ± 0.100	0.794 ± 0.216	0.623 ± 0.063	0.732 ± 0.122	0.733 ± 0.113	0.976 ± 0.260
C	$\mathbf{0.802 \pm 0.091}$	0.834 ± 0.115	0.815 ± 0.105	0.947 ± 0.160	1.032 ± 0.086	0.983 ± 0.114	0.984 ± 0.127	1.185 ± 0.173
D	$\mathbf{0.999 \pm 0.082}$	1.079 ± 0.126	1.076 ± 0.130	1.045 ± 0.189	1.041 ± 0.084	1.099 ± 0.120	1.085 ± 0.126	1.178 ± 0.203
E	$\mathbf{0.624 \pm 0.062}$	0.669 ± 0.097	0.638 ± 0.087	0.779 ± 0.154	0.669 ± 0.063	0.680 ± 0.109	0.711 ± 0.090	1.010 ± 0.214
F	$\mathbf{0.975 \pm 0.088}$	1.047 ± 0.118	1.039 ± 0.126	1.015 ± 0.123	0.991 ± 0.085	1.031 ± 0.125	1.036 ± 0.113	1.156 ± 0.205
G	0.718 ± 0.062	0.800 ± 0.108	0.846 ± 0.119	0.912 ± 0.139	$\mathbf{0.717 \pm 0.074}$	0.821 ± 0.105	0.863 ± 0.117	1.086 ± 0.190
H	0.842 ± 0.073	0.854 ± 0.118	0.885 ± 0.104	1.020 ± 0.153	$\mathbf{0.825 \pm 0.077}$	0.864 ± 0.102	0.894 ± 0.124	1.155 ± 0.196
I	$\mathbf{0.689 \pm 0.084}$	0.715 ± 0.098	0.797 ± 0.114	0.894 ± 0.184	0.693 ± 0.087	0.744 ± 0.137	0.800 ± 0.092	1.003 ± 0.246
J	$\mathbf{1.029 \pm 0.098}$	1.110 ± 0.125	1.073 ± 0.137	1.087 ± 0.178	1.060 ± 0.079	1.136 ± 0.138	1.105 ± 0.130	1.179 ± 0.208
Mean	**0.792**	0.855	0.864	0.936	0.834	0.885	0.902	1.099

TABLE 4. MAEs of the PM and CMs1–7 in the condition that 90% of videos have not been watched by subjects.

Subject	PM	CM1	CM2	CM3	CM4	CM5	CM6	CM7
A	$\mathbf{0.725 \pm 0.049}$	0.811 ± 0.146	1.352 ± 0.127	1.297 ± 0.132	0.732 ± 0.065	0.809 ± 0.173	1.374 ± 0.135	1.401 ± 0.144
B	0.684 ± 0.069	0.801 ± 0.197	1.537 ± 0.194	1.448 ± 0.165	$\mathbf{0.644 \pm 0.071}$	0.779 ± 0.231	1.417 ± 0.144	1.429 ± 0.134
C	$\mathbf{0.891 \pm 0.088}$	0.920 ± 0.122	1.011 ± 0.131	1.072 ± 0.150	1.035 ± 0.097	1.003 ± 0.121	1.284 ± 0.186	1.403 ± 0.185
D	1.072 ± 0.060	1.159 ± 0.140	1.551 ± 0.184	1.508 ± 0.175	$\mathbf{1.041 \pm 0.074}$	1.136 ± 0.158	1.454 ± 0.181	1.448 ± 0.184
E	$\mathbf{0.650 \pm 0.048}$	0.728 ± 0.142	1.218 ± 0.122	1.194 ± 0.119	0.699 ± 0.072	0.738 ± 0.165	1.348 ± 0.156	1.390 ± 0.137
F	$\mathbf{1.005 \pm 0.049}$	1.092 ± 0.104	1.404 ± 0.138	1.378 ± 0.134	1.008 ± 0.060	1.082 ± 0.127	1.410 ± 0.172	1.412 ± 0.173
G	$\mathbf{0.774 \pm 0.052}$	0.854 ± 0.147	1.376 ± 0.135	1.340 ± 0.148	0.777 ± 0.054	0.852 ± 0.173	1.380 ± 0.149	1.397 ± 0.146
H	$\mathbf{0.852 \pm 0.047}$	0.886 ± 0.117	1.301 ± 0.094	1.296 ± 0.123	0.860 ± 0.061	0.890 ± 0.131	1.392 ± 0.142	1.422 ± 0.158
I	$\mathbf{0.730 \pm 0.056}$	0.815 ± 0.169	1.403 ± 0.160	1.340 ± 0.150	0.737 ± 0.082	0.825 ± 0.183	1.402 ± 0.130	1.388 ± 0.144
J	1.103 ± 0.062	1.201 ± 0.142	1.617 ± 0.187	1.572 ± 0.173	$\mathbf{1.053 \pm 0.066}$	1.162 ± 0.157	1.467 ± 0.184	1.421 ± 0.184
Mean	**0.848**	0.927	1.377	1.344	0.859	0.927	1.393	1.411

accuracy of estimation by the PM is highest among the eight methods for all rates of videos that subjects have not watched. By comparing the PM with CMs1–3, it is confirmed that the use of S2LPCCA is more effective than the use of SLPCCA, SemiCCA and CCA for accurate estimation of interest levels. From these results, it is confirmed that both the introduction of the PCA's framework into CCA and preservation of local structures in the same class are effective for interest level estimation. This verification can be also indicated in the case of comparing CM4 with CMs5–7. Moreover, by comparing the PM with CM4, it is confirmed that the tensor completion, which can be applied to three contexts, videos, users and "canonical features and interest levels," is effective for accurate estimation of interest levels. Therefore, it is verified that the proposed method realizes accurate estimation of interest levels by using S2LPCCA and the tensor completion with partially paired users' behavior and videos.

VI. CONCLUSIONS

A novel method for interest level estimation based on tensor completion via feature integration for partially paired users' behavior and videos has been presented in this paper. The proposed method newly defines a CCA framework called S2LPCCA that is suitable for interest level estimation using data for users' behavior. Since S2LPCCA is a hybrid version of the factors of SemiCCA and SLPCCA, it is robust for both partially paired samples, *e.g.*, users' behavior and videos in the real world, and interest level estimation. The experimental results have shown the effectiveness of the proposed method. The experimental results also show that the proposed method

is robust for incomplete data and is effective for interest level estimation in actual shops and on the Internet.

REFERENCES

[1] J. Liu, Y. Gu, and S. Kamijo, "Customer behavior recognition in retail store from surveillance camera," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2015, pp. 154–159.
[2] K. Kamei, T. Ikeda, H. Kidokoro, M. Shiomi, A. Utsumi, K. Shinozawa, T. Miyashita, and N. Hagita, "Effectiveness of cooperative customer navigation from robots around a retail shop," in *Proc. IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 235–241.
[3] M. C. Popa, L. J. M. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers, "Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information," *Pattern Recognit. Lett.*, vol. 34, no. 7, pp. 809–819, 2013.
[4] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.
[5] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of quality of experience of video-on-demand services: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 401–418, 1st Quart., 2016.
[6] Y. Ito, T. Ogawa, and M. Haseyama, "Personalized video preference estimation based on early fusion using multiple users' viewing behavior," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3006–3010.
[7] T. Kushima, S. Takahashi, T. Ogawa, and M. Haseyama, "Interest level estimation of items via matrix completion based on adaptive user matrix construction," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.
[8] J. Liu, Y. Gu, and S. Kamijo, "Customer behavior classification using surveillance camera for marketing," *Multimedia Tools Appl.*, vol. 76, no. 5, pp. 6595–6622, 2017.
[9] Y.-C. Wang and C.-C. Yang, "3S-cart: A lightweight, interactive sensor-based cart for smart shopping in supermarkets," *IEEE Sensors J.*, vol. 16, no. 17, pp. 6774–6781, Sep. 2016.
[10] Z. Ma, J. Wu, S.-H. Zhong, J. Jiang, and S. J. Heinen, "Human eye movements reveal video frame importance," *Computer*, vol. 52, no. 5, pp. 48–57, May 2019.
[11] Y. Ding, X. Hu, Z. Xia, Y.-J. Liu, and D. Zhang, "Inter-brain EEG feature extraction and analysis for continuous implicit emotion tagging during video watching," *IEEE Trans. Affect. Comput.*, to be published.
[12] X. Zhang, S. Li, and R. R. Burke, "Modeling the effects of dynamic group influence on shopper zone choice, purchase conversion, and spending," *J. Acad. Marketing Sci.*, vol. 46, no. 6, pp. 1089–1107, 2018.
[13] Y. Hao, J. Yang, M. Chen, M. S. Hossain, and M. F. Alhamid, "Emotion-aware video QoE assessment via transfer learning," *IEEE Multimedia*, vol. 26, no. 1, pp. 31–40, Jan./Mar. 2019.
[14] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends Cognit. Sci.*, vol. 7, no. 11, pp. 498–504, 2003.
[15] S. Mehrkanoon and J. A. K. Suykens, "Regularized semipaired kernel CCA for domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3199–3213, Jul. 2018.
[16] X. Guo, S. Wang, Y. Tie, L. Qi, and L. Guan, "Joint intermodal and intramodal correlation preservation for semi-paired learning," *Pattern Recognit.*, vol. 81, pp. 36–49, Sep. 2018.
[17] A. Kimura, H. Kameoka, M. Sugiyama, T. Nakano, H. Sakano, E. Maeda, and K. Ishiguro, "SemiCCA: Efficient semi-supervised learning of canonical correlations," *Inf. Media Technol.*, vol. 8, no. 2, pp. 311–318, 2013.
[18] J. Yang and X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognit. Lett.*, vol. 33, no. 5, pp. 623–628, Apr. 2012.
[19] X. Chen, S. Chen, H. Xue, and X. Zhou, "A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data," *Pattern Recognit.*, vol. 45, no. 5, pp. 2005–2018, May 2012.
[20] X. Shen and Q. Sun, "A novel semi-supervised canonical correlation analysis and extensions for multi-view dimensionality reduction," *J. Vis. Commun. Image Represent.*, vol. 25, no. 8, pp. 1894–1904, 2014.
[21] B. Zhang, J. Hao, G. Ma, J. Yue, and Z. Shi, "Semi-paired probabilistic canonical correlation analysis," in *Proc. Int. Conf. Intell. Inf. Process.*, 2014, pp. 1–10.
[22] J. Wan, H. Wang, and M. Yang, "Cost sensitive semi-supervised canonical correlation analysis for multi-view dimensionality reduction," *Neural Process. Lett.*, vol. 45, no. 2, pp. 411–430, Apr. 2017.
[23] S. Hou, H. Liu, and Q. Sun, "Sparse regularized discriminative canonical correlation analysis for multi-view semi-supervised learning," *Neural Computing and Applications*. London, U.K.: Springer, 2018, pp. 1–9.
[24] T. Matsuura, K. Saito, Y. Ushiku, and T. Harada, "Generalized Bayesian canonical correlation analysis with missing modalities," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–16.
[25] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
[26] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *ACM Trans. Knowl. Discovery Data*, vol. 13, no. 1, Jan. 2019, Art. no. 6.
[27] J. Liao, J. Tang, W. Zeng, and X. Zhao, "Efficient and accurate traffic flow prediction via incremental tensor completion," *IEEE Access*, vol. 6, pp. 36897–36905, 2018.
[28] Z. Zhu, J. Wang, Y. Zhang, and J. Caverlee, "Fairness-aware recommendation of information curators," Sep. 2018, *arXiv:1809.03040*. [Online]. Available: https://arxiv.org/abs/1809.03040
[29] Y. Yang, L. Han, Z. Gou, B. Duan, J. Zhu, and H. Yan, "Tagrec-CMTF: Coupled matrix and tensor factorization for tag recommendation," *IEEE Access*, vol. 6, pp. 64142–64152, 2018.
[30] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
[31] D. Zhang, Y. Hu, J. Ye, X. Li, and X. He, "Matrix completion by truncated nuclear norm regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2192–2199.
[32] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
[33] Z.-F. Han, C.-S. Leung, L.-T. Huang, and H. C. So, "Sparse and truncated nuclear norm based tensor completion," *Neural Process. Lett.*, vol. 45, no. 3, pp. 729–743, Jun. 2017.
[34] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," 2018, *arXiv:1812.08008*. [Online]. Available: https://arxiv.org/abs/1812.08008
[35] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2001, pp. 301–308.
[36] P. Qvarfordt and S. Zhai, "Conversing with the user based on eye-gaze patterns," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2005, pp. 221–230.
[37] J. Pettersson, A. Albo, J. Eriksson, P. Larsson, K. W. Falkman, and P. Falkman, "Cognitive ability evaluation using virtual reality and eye tracking," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2018, pp. 1–6.
[38] Y. Yu, Q. Wu, Y. Feng, T. Guo, J. Yang, S. Takahashi, Y. Ejima, and J. Wu, "A central-scotoma simulator based on low-cost eye tracker," in *Proc. IEEE Int. Conf. Mechatronics Automat. (ICMA)*, Aug. 2018, pp. 1–6.
[39] X. Luo, J. Shen, H. Zeng, A. Song, B. Xu, H. Li, P. Wen, and C. Hu, "Interested object detection based on gaze using low-cost remote eye tracker," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 1101–1104.
[40] M. L. Gavrilova, Y. Wang, F. Ahmed, and P. P. Paul, "Kinect sensor gesture and activity recognition: New applications for consumer cognitive systems," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 88–94, Dec. 2017.
[41] H. Kim, S. Lee, Y. Kim, S. Lee, D. Lee, J. Ju, and H. Myung, "Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system," *Expert Syst. Appl.*, vol. 45, pp. 131–141, Mar. 2016.
[42] W.-H. Lee and R. Lee, "Implicit sensor-based authentication of smartphone users with smartwatch," in *Proc. Hardw. Archit. Support Secur. Privacy*, 2016, pp. 1–8.
[43] A. R. Smith, "Color gamut transform pairs," *ACM Siggraph Comput. Graph.*, vol. 12, no. 3, pp. 12–19, 1978.
[44] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop Stat. Learn. Comput. Vis. (ECCV)*, vol. 1, 2004, pp. 1–22.
[45] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[49] J. Lee, Y. Koh, and J. Yang, "A deep learning based video classification system using multimodality correlation approach," in *Proc. 17th Int. Conf. Control, Automat. Syst. (ICCAS)*, Oct. 2017, pp. 2021–2025.
[50] J. Song, Z. Guo, L. Gao, W. Liu, D. Zhang, and H. T. Shen, "Hierarchical LSTM with adjusted temporal attention for video captioning," 2017, *arXiv:1706.01231*. [Online]. Available: https://arxiv.org/abs/1706.01231
[51] E.-S. Kim, K.-W. On, J. Kim, Y.-J. Heo, S.-H. Choi, H.-D. Lee, and B.-T. Zhang, "Temporal attention mechanism with conditional inference for large-scale multi-label video classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jun. 2018, pp. 1–12.
[52] H. Wang, H. Yu, P. Chen, R. Hua, C. Yan, and L. Zou, "Unsupervised video highlight extraction via query-related deep transfer," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2971–2976.
[53] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1625–1634.
[54] T. Fujii, S. Yoshida, and M. Muneyasu, "Video retrieval by reranking and relevance feedback with tag-based similarity," in *Proc. IEEE 7th Global Conf. Consum. Electron. (GCCE)*, Oct. 2018, pp. 1–2.
[55] A. B. Mahjoub and M. Atri, "An efficient end-to-end deep learning architecture for activity classification," *Analog Integr. Circuits Signal Process.*, vol. 99, no. 1, pp. 23–32, 2019.
[56] M. A. Hassan, S. Saleem, M. Z. Khan, and M. U. G. Khan, "Story based video retrieval using deep visual and textual information," in *Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Mar. 2019, pp. 166–171.
[57] M. A. S. Boksem and A. Smidts, "Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success," *J. Marketing Res.*, vol. 52, no. 4, pp. 482–492, 2015.
[58] S. H. Fairclough, A. J. Karran, and K. Gilleade, "Classification accuracy from the perspective of the user: Real-time interaction with physiological computing," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 3029–3038.
[59] S. Liu, J. Lv, Y. Hou, T. Shoemaker, Q. Dong, K. Li, and T. Liu, "What makes a good movie trailer?: Interpretation from simultaneous EEG and eyetracker recording," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 82–86.
[60] Y. Sasaka, T. Ogawa, and M. Haseyama, "A novel framework for estimating viewer interest by unsupervised multimodal anomaly detection," *IEEE Access*, vol. 6, pp. 8340–8350, 2018.

TETSUYA KUSHIMA (S'17) received the B.S. degree in electronics and information engineering from Hokkaido University, Japan, in 2018, where he is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. His research interest includes multimodal signal processing. He is a Student Member of the IEICE.

SHO TAKAHASHI (S'09–M'13) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2008, 2010, and 2013, respectively, where he joined the Graduate School of Information Science and Technology as an Assistant Professor, in 2013. He was an Associate Professor with the Education and Research Center for Mathematical and Data Science, Hokkaido University, from 2017 to 2018. He was a Visiting Researcher with the Media Integration and Communication Center, University of Florence, from 2018 to 2019. He is currently an Associate Professor with the Faculty of Engineering, Hokkaido University. His research interest includes semantic analysis and visualization in various data and its applications. He is a member of the IEICE, the Institute of Image Information and Television Engineers (ITE), the Japan Society of Civil Engineering (JSCE), the Society of Automotive Engineering of Japan (JSAE), the Japan Society of Traffic Engineers (JSTE), and the Japan Association for Human and Environmental Symbiosis (JAHES).

TAKAHIRO OGAWA (S'03–M'08–SM'18) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively. He is currently an Associate Professor with the Faculty of Information Science and Technology, Hokkaido University. His research interest includes multimedia signal processing and its applications. He has been an Associate Editor of the *ITE Transactions on Media Technology and Applications*. He is a member of the ACM, EURASIP, IEICE, and the Institute of Image Information and Television Engineers (ITE).

MIKI HASEYAMA (S'88–M'91–SM'06) received the B.S., M.S., and Ph.D. degrees in electronics from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor, in 1994, where she is currently a Professor with the Faculty of Information Science and Technology. She was a Visiting Associate Professor with Washington University, USA, from 1995 to 1996. Her research interest includes image and video processing and its development into semantic analysis. She is a member of the IEICE and Information Processing Society of Japan (IPSJ) and a Fellow of the Institute of Image Information and Television Engineers (ITE). She has been a Vice-President of the ITE and the Director of the International Coordination and Publicity of the IEICE. She has been the Editor-in-Chief of the *ITE Transactions on Media Technology and Applications*.