IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 28, 2021, accepted November 28, 2021, date of publication December 1, 2021,
date of current version December 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131979

Feature Integration Through Semi-Supervised Multimodal Gaussian Process Latent Variable Model With Pseudo-Labels for Interest Level Estimation

KYOHEI KAMIKAWA[1], (Member, IEEE), KEISUKE MAEDA[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]Office of Institutional Research, Hokkaido University, Sapporo 060-0808, Japan
[3]Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

Corresponding author: Kyohei Kamikawa (kamikawa@lmd.ist.hokudai.ac.jp)

This study was supported in part by AMED Grant Number JP21zf0127004.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethical committee in Hokkaido University.

ABSTRACT This study presents a novel feature integration method for interest level estimation using a semi-supervised multimodal Gaussian process latent variable model with pseudo-labels (semi-MGPPL). Semi-MGPPL is an extended version of the multimodal Gaussian process latent variable model (mGPLVM). It integrates features calculated from multiple modalities to predict the users' interest levels in content. It is known that reflecting known interest levels of known users in the latent space effectively improves the accuracy of interest level estimation. However, previous methods have difficulty reflecting the interest levels when the number of samples is insufficient. Semi-MGPPL efficiently reflects interest levels in the latent space by pseudo-labeling of unlabeled samples and increasing the number of available pairs among labeled samples. In addition, obtaining behavior features is difficult for a new test sample. However, requirement of features of all modalities by previous mGPLVM-based methods makes the calculation of latent variables of a test sample challenging. Semi-MGPPL solves this problem by training a projection function from the original feature to the latent space. The experimental results on real data demonstrate the effectiveness and robustness of semi-MGPPL.

INDEX TERMS Gaussian process, multimodal analysis, feature integration, semi-supervised learning, pseudo-label.

I. INTRODUCTION

With the spread of multimedia content sharing services, such as YouTube[1] and Spotify,[2] several contents are available on the web [1]. Due to the huge amount of contents available nowadays, users must provide appropriate queries to find their favorite contents [2]. However, it is often difficult or labor-intensive to provide appropriate queries for searching unknown contents. To solve this problem, several recommendation systems requiring no queries have been proposed [3], [4]. Nevertheless, these methods do not reflect the individual interests of users because they use information obtained from contents instead of users. Therefore, to obtain more user-centered recommendations, an approach using users' behavior information while watching contents has been proposed [5], [6]. This approach requires no queries and can reflect users' interests by constructing a latent space with multiple modalities. The latent space enables highly accurate

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

[1]http://www.youtube.com/
[2]https://www.spotify.com/

interest estimation by adequately representing the preferences of each user. Therefore, we focus on a novel feature integration method that utilizes both content and users' behavior information.

To achieve accurate interest level estimation using multiple modalities, such as users' behavior and content information, calculation of effective latent variables that capture the potential relationship between these modalitiesis essential. The traditional approach [6] to estimate latent variables adopts canonical correlation analysis (CCA) [7], which is one of the typical latent variable models. CCA can capture the potential relationship by calculating latent variables that maximize the correlation between multiple modalities through a linear transformation. Furthermore, deep learning-based latent variable models have been proposed, which can flexibly construct the latent space through a nonlinear transformation [8]–[10]; unfortunately, these models have some problems that are difficult to solve. In particular, although the behavior information is useful, it is not practical to collect sufficient training data. In addition, users' behavior is influenced by the content and surrounding environment along with other external factors and emotional state. Therefore, behavior information is likely to contain many noises that are unrelated to the users' interests. However, deterministic models, such as CCA or deep learning-based models, that require a large amount of data may suffer from overfitting since it is difficult to integrate features without the influence of noise in the data [11], [12]. Therefore, the performance of interest level estimation may decrease since these feature integration methods are not suitable for determining the relationship between content and behavior information.

This problem has been solved by proposing a multimodal Gaussian process latent variable model (mGPLVM) [13], which is one of the probabilistic generative approaches. mGPLVM constructs a common latent space by assuming that multiple modalities are probabilistically generated from the common latent space. In particular, mGPLVM maximizes the likelihood of multiple modalities against latent variables. By calculating the common latent space based on the probabilistic approach, correlations between modalities can be represented accurately even in noisy data.

Because mGPLVM is a powerful feature integration method, many researchers have constructed its extended versions [14]–[16]. In addition, to improve the ability for constructing the latent space, several mGPLVM-based methods employing label information have been proposed [17], [18]. However, labeled data are often partial because users assign labels to only some of the content they watch. On the contrary, semi-supervised methods can calculate latent variables even when labels are assigned to a part of the samples. In specific, semi-GPLVM, which is a semi-supervised version of GPLVM, has been proposed [19]. Semi-GPLVM is a feature integration method that uses both labeled and unlabeled samples to calculate the latent space of the data. Furthermore, this method preserves label information in the latent space using pairwise relationships between labeled samples and reflects it in the construction. The unique constraint used in the method places sample pairs with the same label close together, whereas that with different labels far apart in the latent space. Moreover, an extended method that considering the ordinality of the labels has been proposed [20], which reflects the similarity of the labels and distance between the labels in the latent space. However, these methods do not consider the following two problems.

Problem (i): The content labeled by users often amount to small percentages of the content they watch. However, in previous methods, a decrease in the number of labeled samples rapidly reduces the number of pairs between labeled samples; therefore, to efficiently reflect a small amount of label information in the latent space may be difficult.

Problem (ii): A recommendation system estimates users' interest levels when test data are new. In these methods, it is necessary to reconstruct the latent space using all data when calculating the latent variables for test data. Here, we need the features of test data in all modalities. However, to obtain the behavior features of the test data is impossible because users do not watch the content of the test data. Thus, as the latent variables of the test data cannot be calculated, estimation of the interest levels by these methods is difficult.

To obtain an effective latent space for estimating users' interest levels, a novel probabilistic generative model that simultaneously solves the above problems should be constructed.

Therefore, a semi-MGPPL is proposed in this paper to solve the above problem. The proposed semi-MGPPL introduces two following approaches:

Approach (i): The proposed method assumes that unlabeled samples have pseudo-labels [21], [22], which are generated using labeled samples. Therefore, the proposed semi-MGPPL increases the number of pairs using pseudo-labeled samples and efficiently reflects the information of labels in the latent space even when the number of labeled samples is small.

Approach (ii): The proposed method introduces a mapping scheme that allows learning the projection from the observation space to the common latent space [23]. The mapping scheme obtains latent variables in test data even when it is difficult to obtain behavior information.

Because the proposed method is based on mGPLVM, it can integrate features even when noisy modalities, such as behavior features, are included. In addition, the above two novel points allow the proposed method to project new samples to appropriate latent variables when the number of labeled samples is small. Therefore, a highly accurate recommendation system can be realized using the proposed method for interest level estimation. The experimental results show that semi-MGPPL can improve the prediction accuracy of user's interest levels. This is an extension of our previous study [20].

II. RELATED WORKS

mGPLVM is a probabilistic model capable of integrating nonlinear features and applies to various data. Further, it is

an underfitting model for small samples or noisy data. This model uses M modalities denoted by $\{\boldsymbol{X}^m\}_{m=1}^M$, where $\boldsymbol{X}^m = [\boldsymbol{x}_1^m, \boldsymbol{x}_2^m, \ldots, \boldsymbol{x}_N^m]^\mathsf{T} \in \mathbb{R}^{N\times D_m}$ denotes m-th modality, D_m is the dimension of m-th modality, and N is the number of samples. The aim of mGPLVM is to train the latent variables $\boldsymbol{Z} = [z_1, z_2, \ldots, z_N]^\mathsf{T} \in \mathbb{R}^{N\times Q}$, where Q denotes the dimension of the latent variables. The hyperparameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^M\}$, where $\boldsymbol{\theta}^m$ is the hyperparameter of all participating kernel functions and the variance of a noise. This model assumes that each dimension of the observed value is generated from each potential function f_d^m $(d = 1, \ldots, D_m)$ as follows:

$$x_{nd}^m = f_d^m(z_n) + \epsilon_{nd}^m, \tag{1}$$

where x_{nd}^m is (n, d) element in matrix $\boldsymbol{X}^m$, and $\epsilon_{nd}^m \sim \mathcal{N}(0, \sigma_m^2)$ denotes the Gaussian noise. mGPLVM assumes that these latent functions $\{f_d^m\}_{d=1}^{D_m}$ follow the Gaussian distribution, which is defined as follows: $\boldsymbol{f}_d^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{ZZ}}^m)$, where $\boldsymbol{f}_d^m = [f_d^m(z_1), f_d^m(z_2), \ldots, f_d^m(z_N)]^\mathsf{T} \in \mathbb{R}^N$, and $\boldsymbol{K}_{\boldsymbol{ZZ}}^m$ is the covariance matrix of latent variables. k_{ij}^m, which is (i, j) element in $\boldsymbol{K}_{\boldsymbol{ZZ}}^m$, is defined as follows:

$$k_{ij}^m = k^m(z_i, z_j), \tag{2}$$

where $k^m(\cdot)$ is a kernel function. Depending on the purpose of a task, the kernel function can be selected freely from the linear and radial basis function kernels.

The maximum a posteriori (MAP) method can be used to calculate the hyperparameters $\boldsymbol{\Theta}$ and latent variables $\boldsymbol{Z}$. The Bayesian theorem is given by

$$\begin{aligned} &p(\boldsymbol{Z}|\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M, \boldsymbol{\Theta}) \\ &\quad \propto p(\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M|\boldsymbol{Z}, \boldsymbol{\Theta})p(\boldsymbol{Z}), \end{aligned} \tag{3}$$

where $p(\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M|\boldsymbol{Z}, \boldsymbol{\Theta})$ is the joint marginal likelihood of $\{\boldsymbol{X}^m\}_{m=1}^M$ and $p(\boldsymbol{Z})$ is the prior distribution of $\boldsymbol{Z}$. To maximize the posterior distribution, the value on the right-hand side of Eq. (3) should be maximized. Thus, the following maximization problem is obtained:

$$\{\hat{\boldsymbol{Z}}, \hat{\boldsymbol{\Theta}}\} = \arg\max_{\boldsymbol{Z},\boldsymbol{\Theta}} \; p(\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M|\boldsymbol{Z}, \boldsymbol{\Theta})p(\boldsymbol{Z}), \tag{4}$$

where

$$\begin{aligned} &p(\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M|\boldsymbol{Z}, \boldsymbol{\Theta}) \\ &\quad = \prod_{m=1}^M p(\boldsymbol{X}^m|\boldsymbol{Z}, \boldsymbol{\theta}^m), \end{aligned} \tag{5}$$

$$\begin{aligned} &p(\boldsymbol{X}^m|\boldsymbol{Z}, \boldsymbol{\theta}^m) \\ &\quad = \prod_{d=1}^{D_m} \frac{\exp(-\frac{1}{2}(\boldsymbol{x}_{:,d}^m)^\mathsf{T}(\tilde{\boldsymbol{K}}_{\boldsymbol{ZZ}}^m)^{-1}\boldsymbol{x}_{:,d}^m)}{(2\pi)^{\frac{N}{2}}|\tilde{\boldsymbol{K}}_{\boldsymbol{ZZ}}^m|^{\frac{1}{2}}} \\ &\quad = \frac{1}{(2\pi)^{\frac{ND_m}{2}}|\tilde{\boldsymbol{K}}_{\boldsymbol{ZZ}}^m|^{\frac{D_m}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}_{\boldsymbol{ZZ}}^{m-1}\boldsymbol{X}\boldsymbol{X}^\mathsf{T})\right), \end{aligned} \tag{6}$$

and $\tilde{\boldsymbol{K}}_{\boldsymbol{ZZ}}^m = \boldsymbol{K}_{\boldsymbol{ZZ}}^m + \sigma_m^2\boldsymbol{I}_N$, $\boldsymbol{I}_N$ is the N-dimensional identity matrix, and $\boldsymbol{x}_{:,d}^m$ denotes d-th column of $\boldsymbol{X}^m$.

Generally, mGPLVM assumes that each z_n follows the standard Gaussian distribution of the Q dimension and defines $p(\boldsymbol{Z})$ as follows:

$$p(\boldsymbol{Z}) = \prod_{n=1}^N p(z_n), \tag{7}$$

$$p(z_n) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_Q). \tag{8}$$

Different forms of prior distributions can be introduced in mGPLVM to increase the model's flexibility [23], [24]. Based on the above, mGPLVM performs the latent variable calculation to integrate multiple modalities from the observations.

III. FEATURE INTEGRATION VIA SEMI-MGPPL

In this section, the proposed semi-MGPPL method is explained. Using the proposed semi-MGPPL, common latent variables from multiple modalities is calculated. Furthermore, the proposed method considers label information efficiently by assuming pseudo-labels [21], [22] for unlabeled data and increasing the available pairwise relationships between samples. Furthermore, a back constraint (BC) method [23] can be used to estimate a mapping function from observation to latent space. Therefore, the proposed semi-MGPPL can calculate latent variables even in real-world scenarios with no behavioral features in the test data.

A. OBJECTIVE FUNCTION OF SEMI-MGPPL

In this subsection, the objective function of the proposed semi-MGPPL is described. Using the semi-MGPPL, two kinds of features are obtained. First, labeled content features are defined as follows: $\overline{\boldsymbol{X}}^\mathrm{c} = [\overline{\boldsymbol{x}}_1^\mathrm{c}, \overline{\boldsymbol{x}}_2^\mathrm{c}, \ldots, \overline{\boldsymbol{x}}_{\overline{N}}^\mathrm{c}]^\mathsf{T} \in \mathbb{R}^{\overline{N}\times D^\mathrm{c}}$ ($\overline{N}$ being the number of labeled samples) and unlabeled content features $\underline{\boldsymbol{X}}^\mathrm{c} = [\underline{\boldsymbol{x}}_1^\mathrm{c}, \underline{\boldsymbol{x}}_2^\mathrm{c}, \ldots, \underline{\boldsymbol{x}}_{\underline{N}}^\mathrm{c}]^\mathsf{T} \in \mathbb{R}^{\underline{N}\times D^\mathrm{c}}$ ($\underline{N}$ being the number of unlabeled samples) from contents. Second, labeled behavior features are defined by $\overline{\boldsymbol{X}}^\mathrm{b} = [\overline{\boldsymbol{x}}_1^\mathrm{b}, \overline{\boldsymbol{x}}_2^\mathrm{b}, \ldots, \overline{\boldsymbol{x}}_{\overline{N}}^\mathrm{b}]^\mathsf{T} \in \mathbb{R}^{\overline{N}\times D^\mathrm{b}}$ and unlabeled behavior features $\underline{\boldsymbol{X}}^\mathrm{b} = [\underline{\boldsymbol{x}}_1^\mathrm{b}, \underline{\boldsymbol{x}}_2^\mathrm{b}, \ldots, \underline{\boldsymbol{x}}_{\underline{N}}^\mathrm{b}]^\mathsf{T} \in \mathbb{R}^{\underline{N}\times D^\mathrm{b}}$ from users' behavior while viewing content. In addition, their concatenated matrices are defined as $\boldsymbol{X}^\mathrm{c} \in \mathbb{R}^{N\times D^\mathrm{c}}$ and $\boldsymbol{X}^\mathrm{b} \in \mathbb{R}^{N\times D^\mathrm{b}}$, respectively, where $N = \overline{N} + \underline{N}$. The labels $\overline{\boldsymbol{l}} = [\overline{l}_1, \overline{l}_2, \ldots, \overline{l}_{\overline{N}}]^\mathsf{T} \in \mathbb{R}^{\overline{N}}$ are the known interest levels for different contents viewed by users. $l_n \in \{1, 2, \ldots, L_{\max}\}$ ($L_{\max}$ is the number of types of interest levels), including ordinal grades. The details of features used in the proposed semi-MGPPL are explained in section **IV-A**.

Similar to multimodal similarity Gaussian process latent variable model (m-SimGP) [14], the proposed semi-MGPPL assumes that the similarity matrices of the observed values are generated probabilistically from the latent space. The similarity matrix $\boldsymbol{S}^m = [\boldsymbol{s}_1^m, \boldsymbol{s}_2^m, \ldots, \boldsymbol{s}_N^m]^\mathsf{T} \in \mathbb{R}^{N\times N}$ is defined as follows:

$$s_{pq}^m = \exp\left(-\frac{\|\boldsymbol{x}_p^m - \boldsymbol{x}_q^m\|_2^2}{2\gamma_m}\right), \tag{9}$$

where s^m_{pq} is (p, q) element in the matrix $\boldsymbol{S}^m$ $(m \in \{\mathrm{c}, \mathrm{b}\})$, γ_m is a bandwidth parameter, and $\gamma_m > 0$. Each element of the similarity matrices calculated in Eq. (9) is assumed to be generated from a latent function f^m_q $(q = 1, \ldots, N)$ as follows:

$$s^m_{pq} = f^m_q(\boldsymbol{z}_p) + \epsilon^m_{pq}, \tag{10}$$

$$\epsilon^m_{pq} \sim \mathcal{N}(0, \sigma^2_m). \tag{11}$$

Using the similarity matrices and partial label information, the objective function of the proposed semi-MGPPL is defined as follows:

$$\{\hat{\Psi}, \hat{\boldsymbol{\Theta}}\} = \arg\max_{\Psi, \boldsymbol{\Theta}} p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) p(\Psi) |\underline{\boldsymbol{\Sigma}}|^{-\lambda}, \tag{12}$$

where Ψ is a parameter in the mapping function $g(\cdot)$ that is introduced according to the BC technique, and $\boldsymbol{z}_i = g(\boldsymbol{s}^{\mathrm{c}}_i)$. Furthermore, $p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta})$ is the joint marginal likelihood of $\boldsymbol{S}^{\mathrm{c}}$ and $\boldsymbol{S}^{\mathrm{b}}$, $p(\Psi)$ is the prior distribution of $\boldsymbol{Z}$, $\underline{\boldsymbol{\Sigma}}$ is the covariance matrix of the pseudo-labels, and λ is a trade-off parameter. The definitions of the above characteristics are discussed in detail in sections **III-B** and **III-C**. Similar to the semi-GPLVM-based methods [19], [20], $p(\Psi)$ is calculated using the label information. However, unlike the methods [19], [20] using the obtained labels only, the proposed method uses the obtained and pseudo-labels to calculate $p(\Psi)$. This point corresponds to **Approach (i)**. Unlike mGPLVM, which directly optimizes latent variables, the proposed method optimizes the mapping function parameters that project the observed values into the latent space. This point corresponds to **Approach (ii)**. Therefore, Eq. (12) is solved by alternating between the following two updates: (a) pseudo-labels update based on Gaussian process regression (GPR) and (b) parameters update based on the MAP method. Each procedure is described in detail in the following subsections.

B. UPDATE OF PSEUDO-LABELS BASED ON GPR

Using the proposed method, pseudo-labels $\underline{\boldsymbol{l}}$ of $\underline{\boldsymbol{Z}}$ are generated from $\overline{\boldsymbol{Z}}$ and $\overline{\boldsymbol{l}}$. Because the latent variable $\boldsymbol{Z}$ cannot be observed, it is initialized based on principal component analysis [25]. The kernel function parameters $k'(\cdot)$ used in GPR are optimized as follows:

$$\begin{aligned}\hat{\boldsymbol{\phi}} &= \arg\max_{\boldsymbol{\phi}} \log p(\overline{\boldsymbol{l}}|\overline{\boldsymbol{Z}}, \boldsymbol{\phi}) \\ &= \arg\max_{\boldsymbol{\phi}} \log \mathcal{N}(\overline{\boldsymbol{l}}|\boldsymbol{0}, \tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}}),\end{aligned}$$

where $k'(\cdot)$ is the radial basis function kernel and $\boldsymbol{\phi}$ is the parameter in GPR. Therefore, using the optimized $\hat{\boldsymbol{\phi}}$, the predictive distribution of the pseudo-labels $\mathcal{N}(\underline{\boldsymbol{l}}|\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}})$ are calculated where

$$\underline{\boldsymbol{\mu}} = \tilde{\boldsymbol{K}}'_{\underline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}} (\tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}})^{-1} \overline{\boldsymbol{l}}, \tag{13}$$

$$\underline{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{K}}'_{\underline{\boldsymbol{Z}}\,\underline{\boldsymbol{Z}}} - \tilde{\boldsymbol{K}}'_{\underline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}} (\tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}})^{-1} \tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{Z}}\,\underline{\boldsymbol{Z}}}. \tag{14}$$

The proposed semi-MGPPL defines the prior distribution to reflect the information of the known labels $\overline{\boldsymbol{l}}$ and pseudo-labels $\underline{\boldsymbol{l}}$ in the latent space. The vector $\boldsymbol{l} \in \mathbb{R}^N$ as the concatenation of $\overline{\boldsymbol{l}}$ and $\underline{\boldsymbol{\mu}}$ and the prior distribution $p(\Psi)$ is defined as follows:

$$p(\Psi) = \frac{1}{\text{const.}} \exp\left(-\sum_{i,j=1}^{N} a_{ij} \| g(\boldsymbol{s}^{\mathrm{c}}_i) - g(\boldsymbol{s}^{\mathrm{c}}_j) \|_2 \right), \tag{15}$$

$$a_{ij} = \alpha(\Delta - |l_i - l_j|) \frac{e^t}{1 + e^t}, \tag{16}$$

where const. is a constant term, $t = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$, α is a parameter, and $\Delta = \frac{(L_{\max}-1)}{2}$. Moreover, Eq. (15) can be rewritten as follows:

$$\begin{aligned}p(\Psi) &= \frac{1}{\text{const.}} \exp\left(-\sum_{i,j=1}^{N} a_{ij} \| g(\boldsymbol{s}^{\mathrm{c}}_i) - g(\boldsymbol{s}^{\mathrm{c}}_j) \|_2 \right) \\ &= \frac{1}{\text{const.}} \exp\left(-\mathrm{tr}(\boldsymbol{Z}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{Z})\right) \\ &= \frac{1}{\text{const.}} \exp\left(-\mathrm{tr}(\boldsymbol{A} \boldsymbol{Z} \boldsymbol{Z}^{\mathsf{T}})\right),\end{aligned} \tag{17}$$

where $\boldsymbol{A}$ consists of a_{ij}. However, the use of the pseudo-labels in Eq. (15) does not necessarily improve the interest level estimation accuracy because errors in the pseudo-labels may limit the performance [26]. To overcome this problem, the predictive covariance matrix $\underline{\boldsymbol{\Sigma}}$ is explicitly exploited. In particular, the predictive covariance is minimized by considering Eq. (12) in the objective function, as described in [26].

The generation of the pseudo-labels enhances the calculation of $a_{i,j}$ for all sample pairs in Eq. (16). Therefore, the problem that the number of sample pairs will be reduced drastically when the number of labeled samples is small can be solved. In addition, the small amount of label information can be reflected efficiently in the latent space. This corresponds to the contribution of **Approach (i)** in this study.

C. UPDATE OF PARAMETERS BASED ON MAP METHOD

According to mGPLVM, the joint marginal likelihood is defined as follows:

$$\begin{aligned}&p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) \\ &= p(\boldsymbol{S}^{\mathrm{c}}|\Psi, \boldsymbol{\theta}^{\mathrm{c}}) p(\boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\theta}^{\mathrm{b}}) \\ &= \frac{1}{(2\pi)^{\frac{ND^{\mathrm{c}}}{2}} |\tilde{\boldsymbol{K}}^{\mathrm{c}}_{\boldsymbol{Z}\boldsymbol{Z}}|^{\frac{D^{\mathrm{c}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^{\mathrm{c}-1}_{\boldsymbol{Z}\boldsymbol{Z}} \boldsymbol{S}^{\mathrm{c}} (\boldsymbol{S}^{\mathrm{c}})^{\mathsf{T}})\right) \\ &\quad \frac{1}{(2\pi)^{\frac{ND^{\mathrm{b}}}{2}} |\tilde{\boldsymbol{K}}^{\mathrm{b}}_{\boldsymbol{Z}\boldsymbol{Z}}|^{\frac{D^{\mathrm{b}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^{\mathrm{b}-1}_{\boldsymbol{Z}\boldsymbol{Z}} \boldsymbol{S}^{\mathrm{b}} (\boldsymbol{S}^{\mathrm{b}})^{\mathsf{T}})\right).\end{aligned} \tag{18}$$

The linear kernel $k(\cdot)$, which is the simplest kernel function, is used to verify the effectiveness of the proposed semi-MGPPL. By rewriting Eq. (12), the following equation can be obtained.

$$\begin{aligned}\{\hat{\Psi}, \hat{\boldsymbol{\Theta}}\} &= \arg\max_{\Psi, \boldsymbol{\Theta}} \log \mathcal{L}, \\ \log \mathcal{L} &= \log p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) + \log p(\Psi) - \lambda \log |\underline{\boldsymbol{\Sigma}}|.\end{aligned} \tag{19}$$

In this study, Eq. (19) is solved using the scaled conjugate gradient method [27] similar to previous methods [24], [28], [29].

FIGURE 1. Experiment's environment. Red and blue squares indicate a sensor of Tobii Eye Tracker 4C and that for OpenPose, respectively.

By solving Eqs. (13) and (19) iteratively, the parameters of the mapping function Ψ and hyperparameters $\boldsymbol{\Theta}$ are optimized. Then, application of the estimated mapping function allows the estimation of the latent variables $\boldsymbol{z}_{(\mathrm{t})}$ of the test sample $\boldsymbol{x}^{\mathrm{c}}_{(\mathrm{t})}$ as follows:

$$\boldsymbol{z}_{(\mathrm{t})} = g(\boldsymbol{s}^{\mathrm{c}}_{(\mathrm{t})}). \tag{20}$$

Therefore, the proposed semi-MGPPL can calculate the latent variables for test data for which no behavioral features are available by assuming the mapping function $g(\cdot)$ that projects the content features into the latent space. In semi-MGPPL, a multilayer perceptron is used as the mapping function. This corresponds to the contribution of **Approach (ii)** in this study.

IV. EXPERIMENTAL RESULTS

In this section, experimental results are presented to verify the effectiveness and robustness of the proposed semi-MGPPL. In IV-A, the dataset used in the experiments is explained. Then, in IV-B, the experimental setup, comparison methods, and evaluation results are described. Finally, the experimental results are presented in IV-C.

A. DATASET

In this subsection, the dataset used in the experiments is explained. In the present experiment, 49 movie trailers obtained from YouTube[3] are used, similar to previous studies [6], [20], [30]. In particular, ten trailers each from "science," "music," "action," and "comedy" genres, and nine from the "sports" genre are used. Each frame of these videos is used as input to Inception-v3 [31] and obtained output vector from the middle layer. Then, their average vectors as content features $\boldsymbol{x}^{\mathrm{c}}_n \in \mathbb{R}^{D^{\mathrm{c}}}$ for each i-th ($i = 1, 2, \ldots, I$; I being the number of the videos) video are calculated, where $D^{\mathrm{c}} = 2048$.

When acquiring user behavior information, the subjects sat in a chair in front of a screen and watched videos, as shown in Fig. 1. First, the subjects were given 10 s as preparation time. Then, one of the videos was shown on the display, and the subjects watched it for 30 s. The subjects then had 5 s to record their interest levels[4] for the watched video. For all videos, the subjects were asked to repeat these actions. Note that the subjects included eight men and two women; they were approximately 22 years old. In this study, while the subjects were watching the videos, Tobii Eye Tracker 4C[5] and OpenPose [32] are used to obtain their behavior information. Tobii Eye Tracker 4C can detect the two-dimensional (2D) eye gaze position of users, and the gaze information correlates closely with users' interest [33]. OpenPose is one of the latest methods for estimating 2D body skeleton positions; it has been recently used in several studies. Based on an affinity for body parts, deep neural networks can estimate them positions [32]. Tobii Eye Tracker 4C and OpenPose are used to obtain information on the gaze and skeletal positions of each body part, respectively. Then, averages and variances over movements of those positions for two axes in the 2D space are calculated and user behavior features $\boldsymbol{x}^{\mathrm{b}}_n \in \mathbb{R}^{D^{\mathrm{b}}}$ are obtained for each i-th video watched by each j-th ($j = 1, 2, \ldots, J$; J being the number of the subjects) subject, where $D^{\mathrm{b}} = 64$, as listed in Table 1. Because behavior features include facial information, information on the facial expression toward the video can be obtained. Therefore, it is expected to acquire information closer to users' interests similar to biometric devices, such as a smartwatch.

B. EXPERIMENTAL CONDITIONS

Unlike mGPLVM, the proposed semi-MGPPL can calculate latent variables of new test data using the projection function from observation to latent space. The effectiveness of the proposed semi-MGPPL is confirmed by comparing the accuracy of interest level estimation on the new test data where only content features are available in the experiment. Because the proposed method aims for feature integration, interest level estimation is performed using a different method. Thus, tensor completion used in previous studies is adopted [6], [20]. In particular, an incomplete tensor was constructed using latent variables of all samples and known labels. Then, labels for test data were estimated using tensor completion. To confirm the robustness of semi-MGPPL, conducted experiments in three situations were conducted. First, 10%, 60%, and 30% of the data are selected randomly as labeled training, unlabeled training, and test data, respectively. Second, 20%, 50%, and 30% of the data are selected randomly as labeled training, unlabeled training, and test data. Third, 30%, 40%, and 30% of the data are selected randomly as labeled training, unlabeled training, and test data. Note that all labeled samples are included in the training data. Figure 2 shows the construction of the dataset used in the first condition.

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

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

[5] https://gaming.tobii.com

TABLE 1. Behavioral features in the experiments.

	Details of features	Dimensions
Eye Tracker	Averages and variances over motions of gaze positions for two axes	4
OpenPose	Averages and variances over motions of the neck, nose, and center of the hip positions for two axes	12
	Averages and variances over motions of both ears, eyes, shoulders, wrists, elbows, and hips for two axes	48
Total		**64**

TABLE 2. Mean absolute error of each subject in the first condition.

Sub	**semi-MGPPL**	BC-mGPLVM [13]	BC-m-SimGP [14]	BC-semi-OMGP [20]	MVCCA [34]	BCCA [35]	Deep CCA [8]
A	**0.713**	0.785	1.013	0.759	1.421	0.724	1.553
B	0.756	**0.747**	1.065	0.789	1.476	0.769	1.507
C	**0.747**	0.761	0.982	0.833	1.387	0.770	1.507
D	0.683	**0.642**	1.171	0.783	1.426	0.793	1.652
E	**0.718**	0.781	1.093	0.768	1.369	0.755	1.480
F	0.791	**0.767**	1.132	0.779	1.372	0.795	1.453
G	0.791	**0.662**	1.067	0.725	1.433	0.794	1.527
H	**0.650**	0.773	1.038	0.730	1.409	0.773	1.372
I	0.729	0.737	1.178	**0.718**	1.421	0.815	1.419
J	**0.694**	0.729	1.030	0.746	1.409	0.716	1.553
Mean ±SD	**0.727** ±0.0432	0.738 ±0.0467	1.077 ±0.0629	0.763 ±0.0334	1.412 ±0.0302	0.770 ±0.0301	1.502 ±0.0742

The performance of the proposed semi-MGPPL is compared with the following six methods: BC-mGPLVM, BC-m-SimGP, BC-semi-supervised ordinally multimodal Gaussian process latent variable model (BC-semi-OMGP), multi-view CCA (MVCCA) [34], Bayesian CCA (BCCA) [35], and Deep CCA [8]. BC-mGPLVM, BC-m-SimGP, and BC-semi-OMGP are methods that introduce the BC technique in mGPLVM [13], m-SimGP [14], and semi-OMGP [20], respectively, to calculate latent variables of new test data. mGPLVM is the baseline method; m-SimGP and semi-OMGP are the mGPLVM-based methods. In addition, three types of extended CCA are selected because CCA is one of the most widely used feature integration methods. MVCCA is a deterministic method for calculating the projection that maximizes the sum of the correlations between multiple modalities. BCCA is a fully Bayesian approach to CCA by assuming an appropriate prior distribution for the model parameters. It also has the advantage of being robust to small sample sizes. Deep CCA is a method for maximizing correlations between multiple modalities using a multilayer perceptron. It has been used for several tasks, not limited to interest level estimation. Note that the parameters in semi-MGPPL λ, Q, γ_m, and α were set as 100, 000, 20, 2, and 100, respectively.

Mean absolute error (MAE), which is defined by the following equation, was used for evaluation.

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{test}}} \sum_{s=1}^{N_{\mathrm{test}}} |l_s^{\mathrm{PRE}} - l_s^{\mathrm{GT}}|, \tag{21}$$

where l_s^{PRE} is the estimated interest level of s-th sample, l_s^{GT} is its ground truth, and N_{test} is the number of test samples.

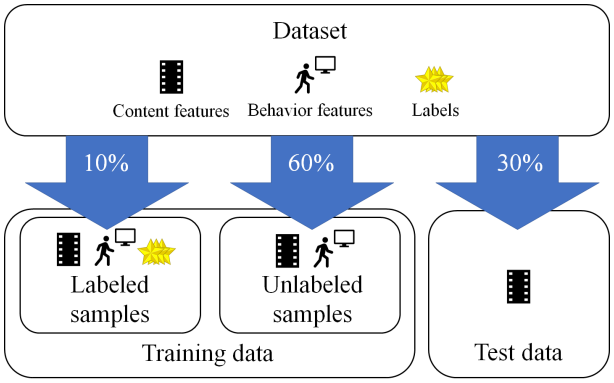

FIGURE 2. Construction of dataset used in the first condition.

C. EXPERIMENTAL RESULTS

In this subsection, the experimental results are presented. The results obtained from the experiments in Tables 2, 3, and 4 are shown. These tables list MAEs for each subject in the interest level estimation of the test data in the first, second, and third conditions, respectively. The experimental results confirm that the proposed semi-MGPPL is effective in a realistic scenario of interest level estimation for test data without behavior features.

Because the proposed semi-MGPPL and other mGPLVM-based methods outperform CCA-based methods, it can be confirmed that feature integration based on mGPLVM is effective in interest level estimation. Furthermore, by comparing semi-MGPPL and BC-semi-OMGP with BC-mGPLVM and BC-m-SimGP, the effectiveness of reflecting label information in the latent space is confirmed. In addition, by comparing semi-MGPPL with

TABLE 3. Mean absolute error of each subject in the second condition.

Sub	**semi-MGPPL**	BC-mGPLVM [13]	BC-m-SimGP [14]	BC-semi-OMGP [20]	MVCCA [34]	BCCA [35]	Deep CCA [8]
A	0.761	0.744	1.014	**0.683**	1.330	0.779	1.494
B	**0.709**	0.757	1.074	0.751	1.374	0.768	1.444
C	**0.728**	0.746	0.950	0.785	1.361	0.764	1.361
D	0.715	**0.639**	0.999	0.764	1.320	0.740	1.425
E	**0.671**	0.760	1.031	0.777	1.320	0.754	1.447
F	0.793	0.787	1.059	**0.715**	1.290	0.733	1.416
G	**0.687**	0.748	0.976	0.778	1.191	0.731	1.488
H	0.698	0.748	1.021	**0.677**	1.328	0.762	1.467
I	0.745	0.792	1.027	0.741	1.318	**0.731**	1.298
J	**0.705**	0.736	1.001	0.742	1.258	0.717	1.458
Mean ±SD	**0.721** ±0.0345	0.746 ±0.0396	1.015 ±0.0348	0.741 ±0.0366	1.309 ±0.0497	0.748 ±0.0191	1.430 ±0.0568

TABLE 4. Mean absolute error of each subject in the third condition.

Sub	**semi-MGPPL**	BC-mGPLVM [13]	BC-m-SimGP [14]	BC-semi-OMGP [20]	MVCCA [34]	BCCA [35]	Deep CCA [8]
A	0.713	**0.677**	0.985	0.704	1.336	0.710	1.339
B	**0.717**	0.748	0.936	0.762	1.330	0.726	1.265
C	**0.691**	0.760	1.005	0.726	1.323	0.717	1.302
D	0.766	**0.748**	0.997	0.808	1.354	0.784	1.419
E	**0.698**	0.729	0.971	0.754	1.260	0.765	1.387
F	0.720	**0.714**	0.992	0.737	1.303	0.737	1.379
G	**0.664**	0.732	0.932	0.711	1.371	0.693	1.419
H	**0.681**	0.746	0.907	0.724	1.351	0.696	1.406
I	0.717	0.754	0.949	0.820	1.285	**0.709**	1.389
J	**0.717**	0.788	0.901	0.730	1.438	0.730	1.376
Mean ±SD	**0.708** ±0.0261	0.740 ±0.0282	0.957 ±0.0359	0.748 ±0.0372	1.335 ±0.0468	0.727 ±0.0277	1.368 ±0.0485

BC-semi-OMGP, the effectiveness of pseudo-labeling for unlabeled samples and its use for the latent space construction is confirmed. Because the estimation accuracy of the proposed semi-MGPPL is better than that of the comparison methods in all situations, robustness of the proposed method to changes in the number of unlabeled samples is confirmed. This robustness is essential for interest level estimation since it is not easy to collect several labeled samples. Tables 2 and 4 show that the proposed semi-MGPPL is more effective than the comparison methods when users have labeled very little and nearly half of the content, respectively.

Therefore, the present experiment confirms the effectiveness of the proposed semi-MGPPL.

V. CONCLUSION

This study presented a novel feature integration method through a semi-supervised multimodal Gaussian process latent variable model with pseudo-labels for interest level estimation. We define a new mGPLVM-based framework called semi-MGPPL suitable for interest level estimation using users' behavior information. Because the proposed semi-MGPPL assumes pseudo-labels for unlabeled samples and can increase the pairwise relationship between labeled samples, the proposed method can efficiently reflect a small amount of label information in the latent space. Furthermore, since semi-MGPPL introduces BC, the proposed method can calculate the latent variables of the newly obtained test data. The experimental results confirmed the effectiveness and robustness of the proposed semi-MGPPL to changes in the number of unlabeled samples.

REFERENCES

[1] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," *IDC Anal. Future*, vol. 2007, pp. 1–16, Dec. 2012.

[2] M. Haseyama, T. Ogawa, and N. Yagi, "[Survey paper] a review of video retrieval based on image and video semantic understanding," *ITE Trans. Media Technol. Appl.*, vol. 1, no. 1, pp. 2–9, 2013.

[3] S. Gong, "A collaborative filtering recommendation algorithm based on user clustering and item clustering," *J. Softw.*, vol. 5, no. 7, pp. 745–752, Jul. 2010.

[4] H. Li, F. Cai, and Z. Liao, "Content-based filtering recommendation algorithm using HMM," in *Proc. 4th Int. Conf. Comput. Inf. Sci.*, Aug. 2012, pp. 275–277.

[5] 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.

[6] T. Kushima, S. Takahashi, T. Ogawa, and M. Haseyama, "Interest level estimation based on tensor completion via feature integration for partially paired user's behavior and videos," *IEEE Access*, vol. 7, pp. 148576–148585, 2019.

[7] H. Hotelling, "The most predictable criterion," *J. Educ. Psychol.*, vol. 26, no. 2, p. 139, 1935.

[8] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1247–1255.

[9] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1083–1092.

[10] M. Suzuki, K. Nakayama, and Y. Matsuo, "Joint multimodal learning with deep generative models," 2016, *arXiv:1611.01891*.
[11] P. Rai and H. Daume, "Multi-label prediction via sparse infinite CCA," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 1518–1526.
[12] M. Cogswell, F. Ahmed, R. Girshick, L. Zitnick, and D. Batra, "Reducing overfitting in deep networks by decorrelating representations," 2015, *arXiv:1511.06068*.
[13] C. H. Ek, P. H. Torr, and N. D. Lawrence, "Gaussian process latent variable models for human pose estimation," in *Proc. Int. Workshop Mach. Learn. Multimodal Interact.*, 2007, pp. 132–143.
[14] G. Song, S. Wang, Q. Huang, and Q. Tian, "Multimodal similarity Gaussian process latent variable model," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4168–4181, Sep. 2017.
[15] P. Li and S. Chen, "Shared Gaussian process latent variable model for incomplete multiview clustering," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 61–73, Jan. 2020.
[16] J. Li, G. Lu, B. Zhang, J. You, and D. Zhang, "Shared linear encoder-based multikernel Gaussian process latent variable model for visual classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 534–547, Feb. 2021.
[17] X. Gao, X. Wang, D. Tao, and X. Li, "Supervised Gaussian process latent variable model for dimensionality reduction," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 41, no. 2, pp. 425–434, Apr. 2011.
[18] N. Yamaguchi, "Visualizing state of time-series data by supervised Gaussian process dynamical models," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 19, no. 5, pp. 688–696, Sep. 2015.
[19] X. Wang, X. Gao, Y. Yuan, D. Tao, and J. Li, "Semi-supervised Gaussian process latent variable model with pairwise constraints," *Neurocomputing*, vol. 73, pp. 2186–2195, Jun. 2010.
[20] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Feature integration via semi-supervised ordinally multi-modal Gaussian process latent variable model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 4130–4134.
[21] J. Wang, C. H. Q. Ding, S. Chen, C. He, and B. Luo, "Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label," *Remote Sens.*, vol. 12, no. 21, p. 3603, Nov. 2020.
[22] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn.*, 2013, vol. 3, no. 2, p. 896.
[23] N. D. Lawrence and J. Qui nonero-Candela, "Local distance preservation in the GPLVM through back constraints," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 513–520.
[24] R. Urtasun and T. Darrell, "Discriminative Gaussian process latent variable model for classification," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 927–934.
[25] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.
[26] R. V. Babu and V. M. Patel, "Learning to count in the crowd from limited labeled data," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 212–229.
[27] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Nov. 1993.
[28] G. Song, S. Wang, Q. Huang, and Q. Tian, "Harmonized multimodal learning with Gaussian process latent variable models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 858–872, Mar. 2021.
[29] M. Matsumoto, K. Maeda, N. Saito, T. Ogawa, and M. Haseyama, "Multimodal label dequantized Gaussian process latent variable model for ordinal label estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 3985–3989.
[30] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Interest level estimation based on feature integration considering distribution of partially paired User's behavior, videos and posters," in *Proc. IEEE 9th Global Conf. Consum. Electron. (GCCE)*, Oct. 2020, pp. 944–945.
[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
[32] 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*.
[33] 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.
[34] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.
[35] A. Klami, S. Virtanen, and S. Kaski, "Bayesian canonical correlation analysis," *J. Mach. Learn. Res.*, vol. 14, pp. 965–1003, Apr. 2013.

KYOHEI KAMIKAWA (Member, IEEE) received the B.S. degree in electronics and information engineering from Hokkaido University, Japan, in 2021, where he is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. His research interests include machine learning and its applications.

KEISUKE MAEDA (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2015, 2017, and 2019, respectively. He is currently a Specially Appointed Assistant Professor with the Office of Institutional Research, Hokkaido University. His research interests include multimodal signal processing, machine learning, and their applications. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

TAKAHIRO OGAWA (Senior Member, IEEE) 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 joined the Graduate School of Information Science and Technology, Hokkaido University, in 2008, where he is currently an Associate Professor with the Faculty of Information Science and Technology. His research interests include artificial intelligence, the Internet of Things, and big data analysis for multimodal signal processing, and their applications. He was a Special Session Chair of IEEE ISCE2009, a Doctoral Symposium Chair of ACM ICMR2018, an Organized Session Chair of IEEE GCCE2017-2019, a TPC Vice Chair of IEEE GCCE2018, and a Conference Chair of IEEE GCCE2019. He has also been an Associate Editor of *Information and Television Engineers (ITE) Transactions on Media Technology and Applications*. He is a member of ACM, IEICE, and ITE.

MIKI HASEYAMA (Senior Member, IEEE) 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. She was a Visiting Associate Professor with Washington University, USA, from 1995 to 1996. She is currently a Professor with the Faculty of Information Science and Technology Division of Media and Network Technologies, Hokkaido University. Her research interests include image and video processing and its development into the semantic analysis. She has been the Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE) and the Director of the International Coordination and Publicity of IEICE. She is a member of IEICE, ITE, and Acoustical Society of Japan.