Feature integration via back-projection ordering multi-modal Gaussian process latent variable model for rating prediction

Paper ID: 3047
Session: WP2.L303 Machine Learning for Multidimensional Signal Processing
30th IEEE International Conference on Image Processing

Kyohei Kamikawa†, Keisuke Maeda††, Takahiro Ogawa†† and Miki Haseyama††

† Graduate School of Information Science and Technology, Hokkaido University
†† Faculty of Information Science and Technology, Hokkaido University

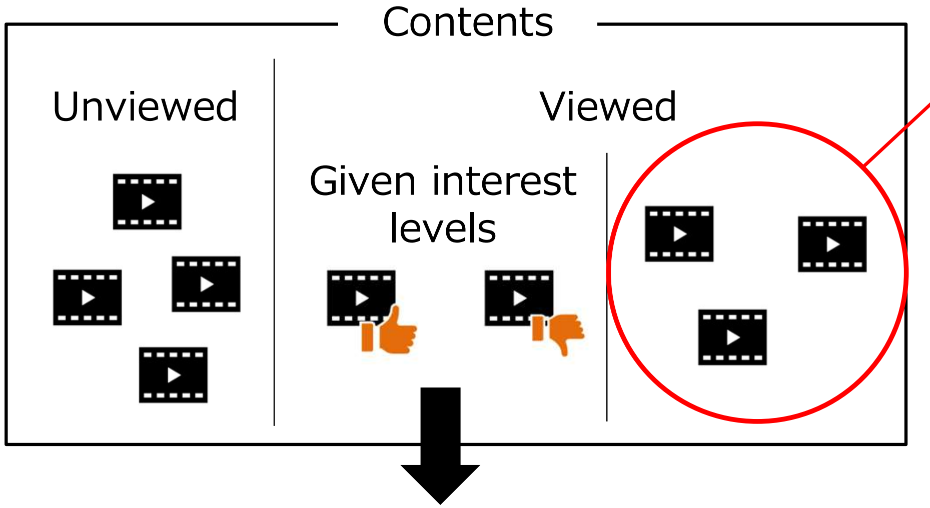
Wednesday, 11 October, 2023

Currently, there are many **multimedia contents** on the Web, such as video, images, and music.

⇒ Choosing what to view is a heavy burden on users.

Necessity of recommending multimedia contents that are estimated to be of high users' interest from unviewed contents

Problems in conventional methods



They can be used to estimate users' interests.

Many contents that users viewed but did not give an interest level to

The conventional methods [1, *1] have used them as viewing history.

These contents are only treated equally, and the conventional methods do not consider users' interests that should be present.

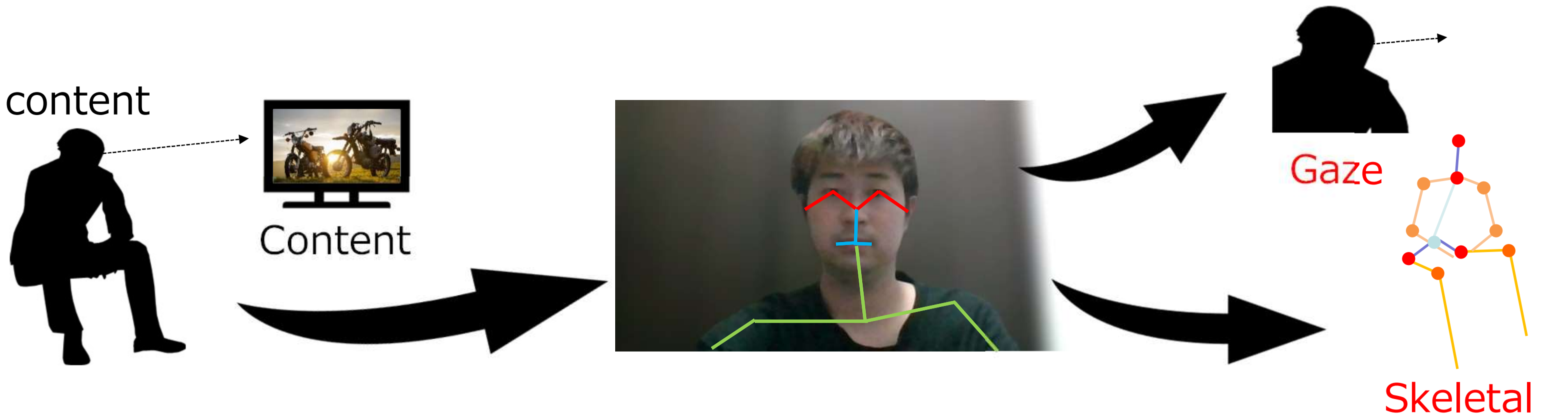
By considering users' interest in the contents not given interest levels, it is expected to improve the accuracy of interest level estimation.

[1] U. Javed, et al., *International Journal of Emerging Technologies in Learning*, 2021.
[*1] H. Li, et al., in *Proc. International Conference on Computational and Information Sciences*, 2012.

Use of the features that are relevant to users' interest in the contents

Behavior features

User viewing content



- ✔ The relevance to users' interest in the viewed contents has reported. [8]
- ✔ If the content has already been viewed, it can be acquired.
- ✔ The burden on the user in acquisition is small.
- △ Since it is easily affected by the surrounding environment other than content, it is noisy [*2].

The use of multiple features, conventionally used content features and users' behavior features, is useful for interest level estimation.

[8] Z. Ma, et al., *Computer*,2019.
[*2] Z. Ma, et al., *Computer*, 2019.

Feature integration

- Construction of the latent space common to multiple features and acquisition of **new features (latent variables)**
- By using latent features, achievement of improved accuracy in tasks such as classification compared to using features as they are [*3]

Type of feature integration methods

・Models based on statistics [*4]

・Models based on deep learning [2, 3]

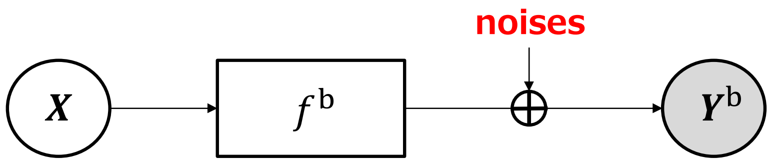
⚠ Behavior features are noisy.

Possibility of overfitting due to noises

Solution: Use of models that can calculate latent variables probabilistically

- **Multimodal Gaussian process latent variable model (mGPLVM) [14]**

It is assumed that behavior features are generated with **noises** from the latent variables as follows:

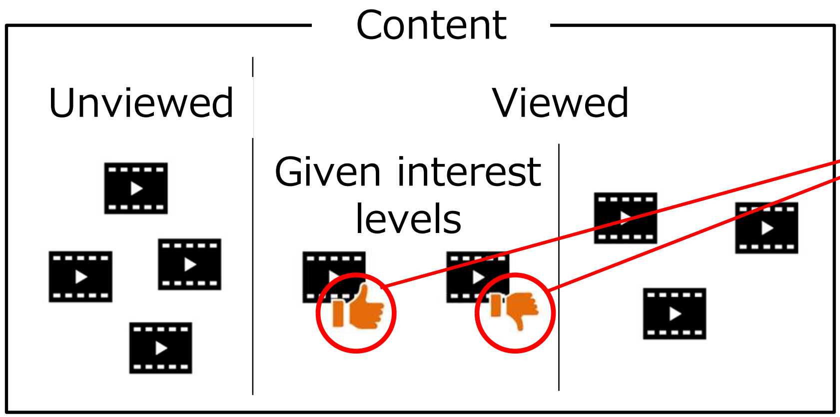


$\boldsymbol{X}$: Latent variables
f^{b}: Latent function
$\boldsymbol{Y}^{\mathrm{b}}$: Behavior features

✔ Overfitting can be avoided.

We integrate content and behavior features based on mGPLVM to avoid overfitting.

[*3] S. Sun, *Neural computing and applications*, 2013.
[2] H.-h. Zhao and H. Liu, *Granular Computing*, 2020.
[3] A. Krizhevsky, et al., in *Proc. Advances in Neural Information Processing Systems*, 2012.
[14] C. H. Ek, et al., in *Proc. International Workshop on Machine Learning for Multimodal Interaction*, 2007.

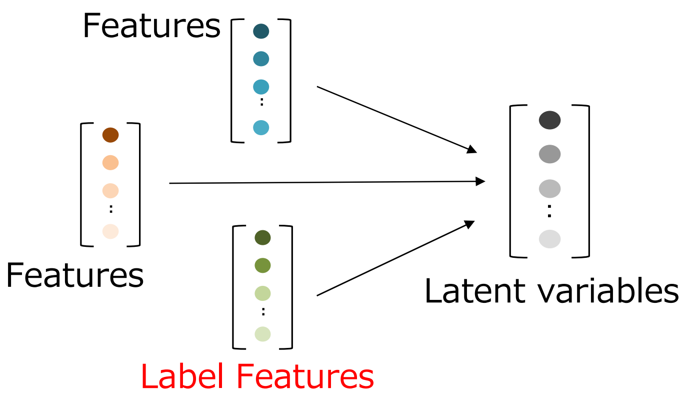


Known interest levels are important information for interest level estimation.

Known interest levels can used as label information.

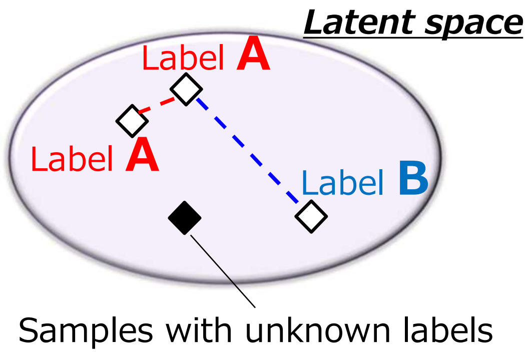
mGPLVM based methods using label information [*4, 11]

supervised GPLVM [*4]



Converting label information into label features

semi-supervised GPLVM [11]



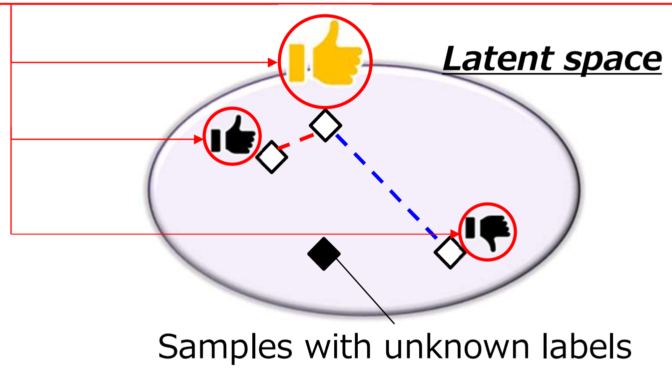
In the latent space, pairs of samples with the same label are trained to be close.

The latent space can improve its representational ability using known interest levels as label information.

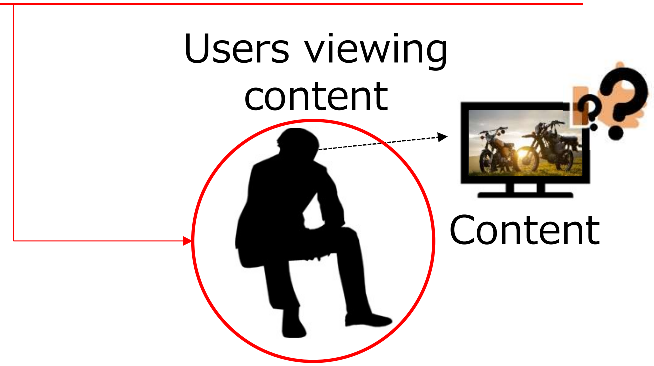
[*4] N. Yamaguchi, *Journal of Advanced Computational Intelligence and Intelligent Informatics,* 2015.
[11] X. Wang, et al., *Neurocomputing,* 2010.

The three purposes of our research are as follows:

- Improvement of the representational power of latent space by using **known interest levels as label information**

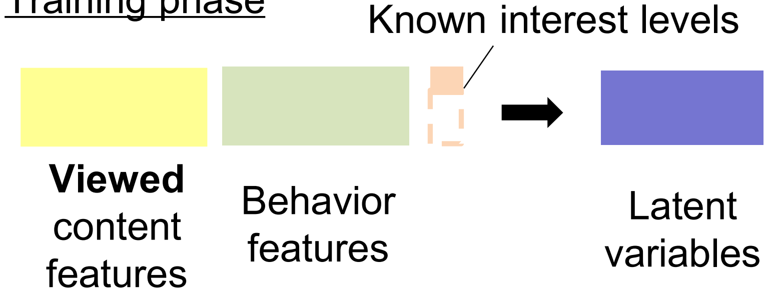


- Consideration of users' interest in content that has not been given an intertest level through **the use of behavior information**



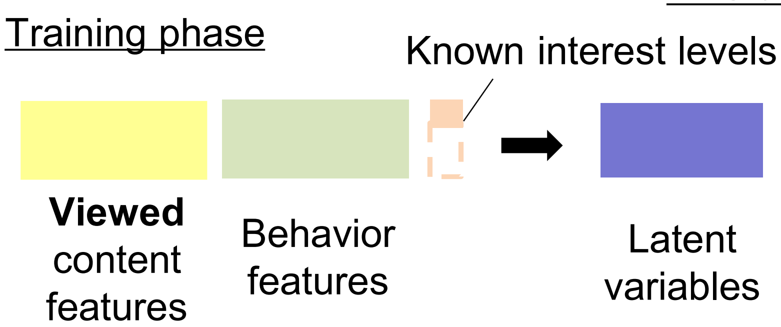
- Assumption of the **content recommendation situation**

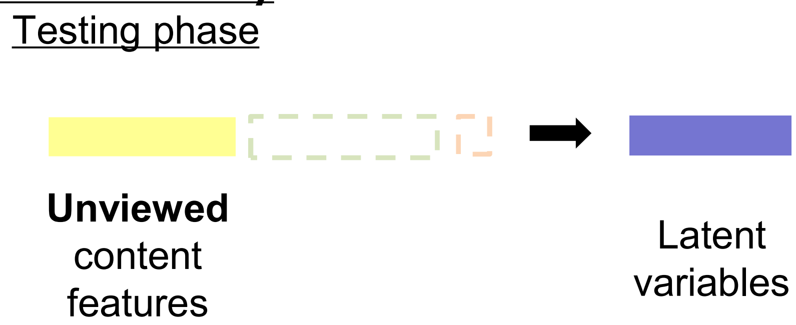
Training phase

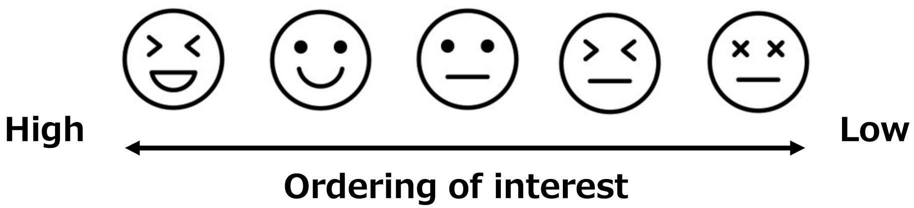


Testing phase

The assumed situation in our study







With conventional methods, it is difficult to reflect the ordering of the interest levels in the latent space.

Approach (i):
Defining the prior distribution of latent variables that introduces the ordering of interest levels

Since behavior features cannot be acquired in testing phase, it is difficult to calculate latent variables in the same way as in training phase.

Approach (ii):
Defining the projection from content features to latent variables and optimized in training phase

The approaches solve the problems of the conventional method.

The assumed situation in our study

Training phase

Known interest levels

Viewed content features

Behavior features

Latent variables

Testing phase

Unviewed content features

Latent variables

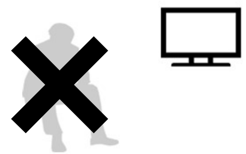
High

Low

Ordering of interest

With conventional methods, it is difficult to reflect the ordering of the interest levels in the latent space.

Approach (i):
Defining the prior distribution of latent variables that introduces the ordering of interest levels

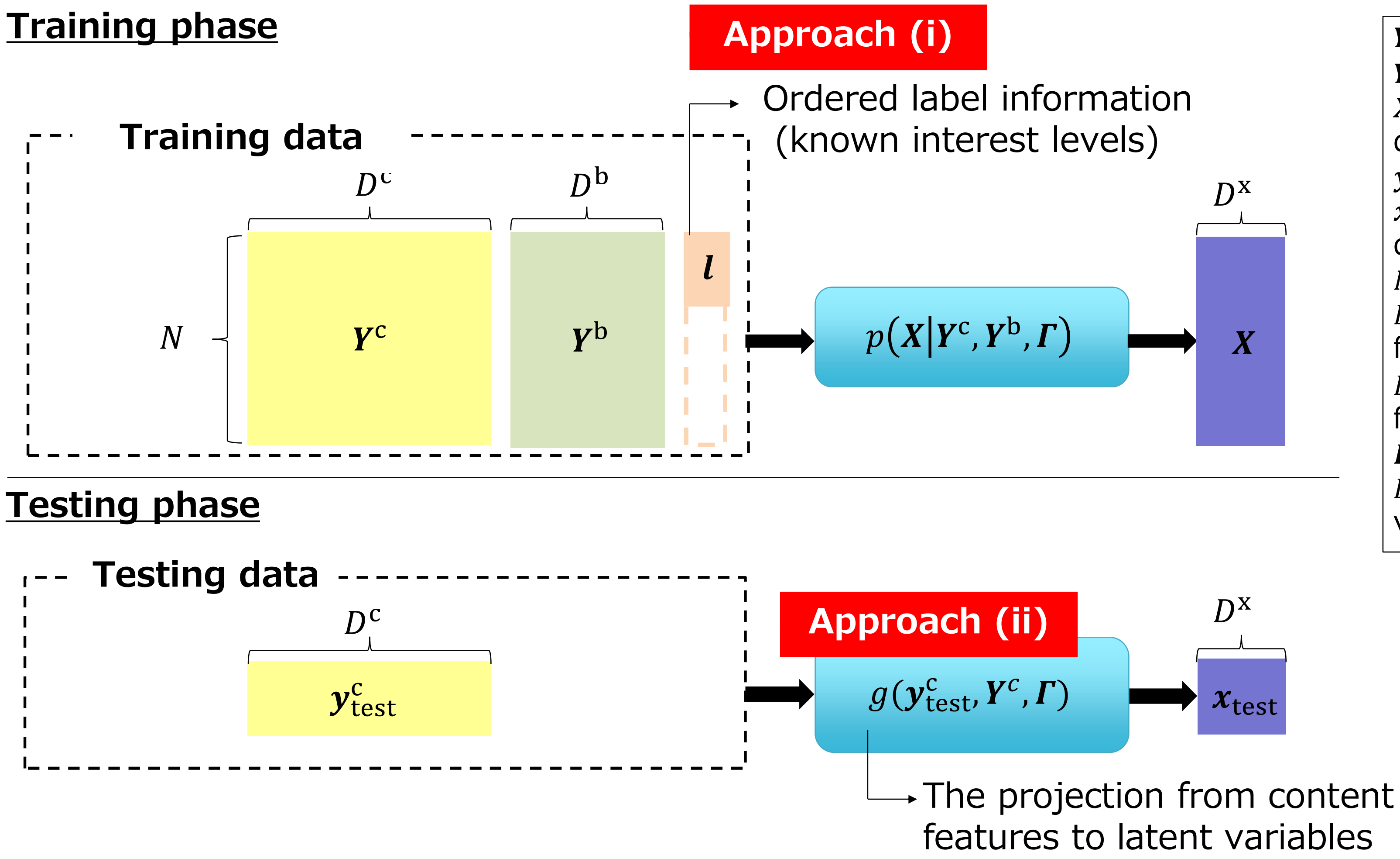


Since behavior features cannot be acquired in testing phase, it is difficult to calculate latent variables in the same way as in training phase.

Approach (ii):
Defining the projection from content features to latent variables and optimized in training phase

The approaches solve the problems of the conventional method.

The proposed method: Back-projection Ordering mGPLVM (BPomGP)



$\boldsymbol{Y}^{c}$: Viewed content features
$\boldsymbol{Y}^{b}$: Behavior features
$\boldsymbol{X}$: Latent variables for training data
$\boldsymbol{y}^{c}_{test}$: Unviewed content features
$\boldsymbol{x}_{test}$: Latent variables for testing data
$\boldsymbol{N}$: The number of samples
D^{c}: The dimension of content features
D^{b}: The dimension of behavior features
$\boldsymbol{\Gamma}$: Projection parameters
D^{x}: The dimension of latent variables

Proposed method: Approach (i)

HOKKAIDO UNIVERSITY

Objective function

$$\arg\min_{\boldsymbol{X},\boldsymbol{\Theta},\boldsymbol{\Gamma}} p(\boldsymbol{Y}^{c},\boldsymbol{Y}^{b}|\boldsymbol{X},\boldsymbol{\Theta})\, p(\boldsymbol{X}|\boldsymbol{Y}^{c},\boldsymbol{Y}^{b},\boldsymbol{\Gamma})\, \boxed{p(\boldsymbol{X}|\boldsymbol{l})}$$

Optimized parameters on the training phase

Likelihood function (Similar to mGPLVM)

Prior distribution of latent variables

$\boldsymbol{x}_i$: Vector of i-th row of $\boldsymbol{X}$
$\boldsymbol{\Theta}$: Projection parameters that transform latent variables into each feature
$\boldsymbol{\Gamma}$: Projection parameters that transform each feature to a latent variable
Z: constant term
α: Hyper-parameters
Δ: Constant term calculated from $\boldsymbol{l}$

Approach (i)

Define the prior distribution of latent variables as follows:

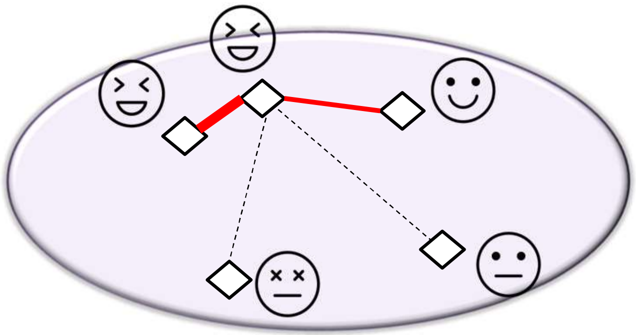
$$p(\boldsymbol{X}|\boldsymbol{l}) = \frac{1}{Z}\exp\left(-\sum_{i,j=1}^{N} w_{i,j}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|\right)$$

Defining the prior distribution of latent variables that introduces the ordering of interest levels

$$w_{i,j} = \begin{cases} \alpha(\Delta - |l_i - l_j|)\dfrac{e^{t_{ij}}}{1+e^{t_{ij}}} & \text{(if } l_i \text{ and } l_j \text{ exist)} \\ 0 & \text{(In cases other than the above)} \end{cases}$$

$$t_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$$

The ordering of known interest levels can be reflected as the distance between latent variables.



Objective function

$$\underset{X,\Theta,\Gamma}{\arg\min}\; p(Y^{c}, Y^{b} | X, \Theta)\; p(X | Y^{c}, Y^{b}, \boxed{\Gamma})\; p(X | l)$$

Optimized parameters on the training phase

Likelihood function (Similar to mGPLVM)

Projecting from each feature to the latent variable

Θ: Projection parameters that transform latent variables into each feature

Γ: Projection parameters that transform each feature to a latent variable

$K_{y^{c}_{\mathrm{test}} Y^{c}}$: The kernel matrix calculated from y^{c}_{test} and Y^{c}

$K_{Y^{c}}$: The kernel matrix calculated from Y^{c}

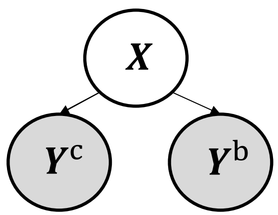
Approach (ii)

x_{test} is calculated as follows:

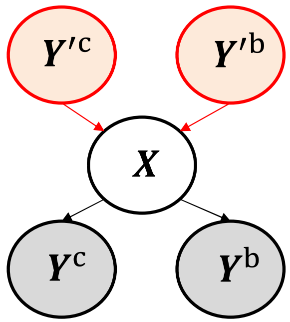
$$x_{\mathrm{test}} = g(y^{c}_{\mathrm{test}}, Y^{c}, \Gamma) = K_{y^{c}_{\mathrm{test}} Y^{c}} (K_{Y^{c}})^{-1} X$$

Graphical model

mGPLVM



Ours



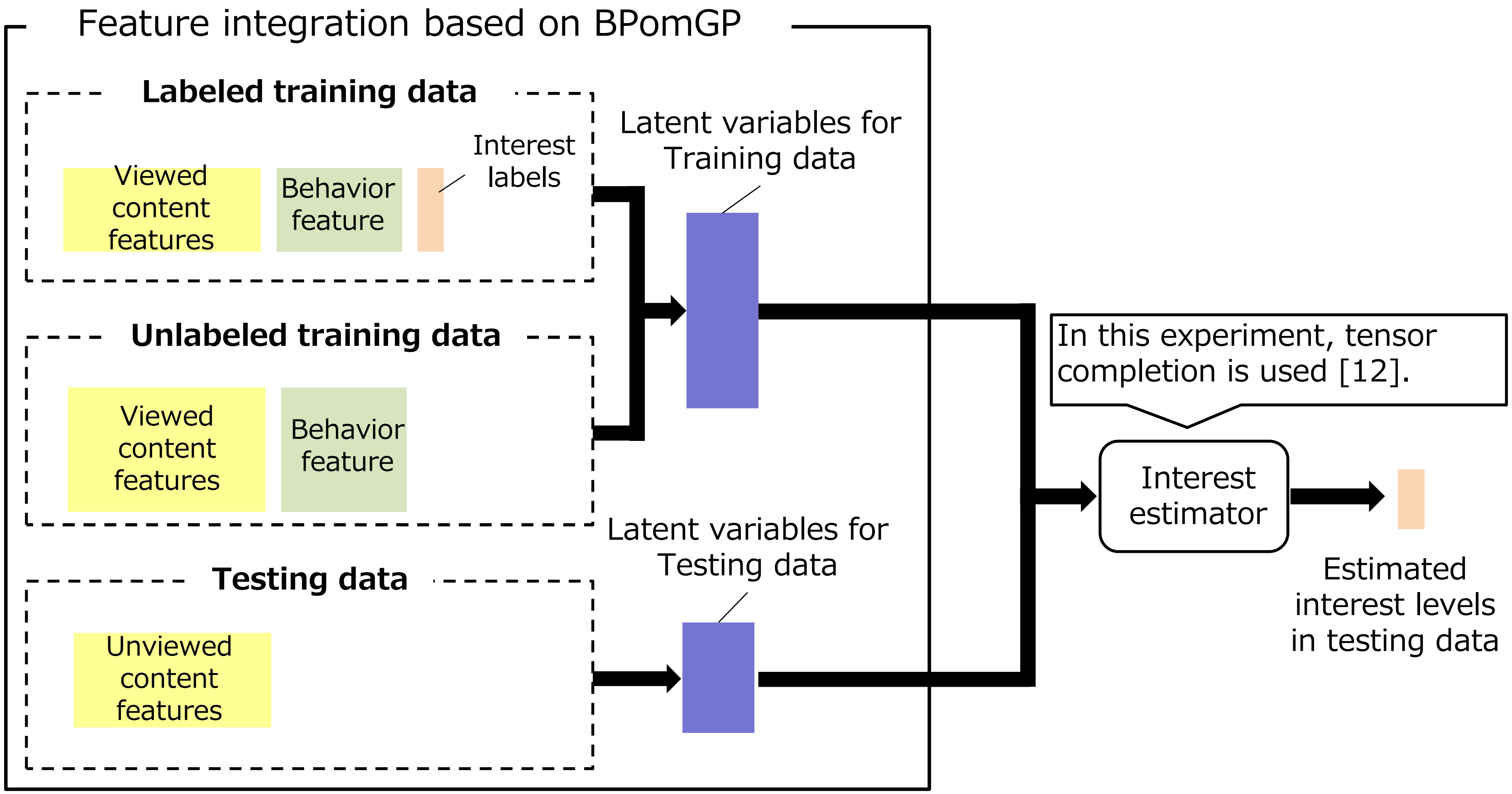
$(\mathrm{s.t.}\ Y'^{m} \approx Y^{m})$

Defining the projection from content features to latent variables and optimized in training phase

Latent variables for unviewed content can be calculated.

Purpose

Verification of interest level estimation accuracy using feature integration based on BPomGP



[12] T. Kushima, et al., *IEEE Access*, 2019.

Experiment: Dataset

Content features (2,048 dimensions)

49 movie trailers of five genres obtained from YouTube* → By inputting each frame of these videos, we calculated outputs from the middle layer of Inception-v3 [*5].

Behavior features (64 dimensions)

Behavior information obtained by using OpenPose [21] and Tobii Eye Tracker 4C** → We extracted users' behavior features from the body skeleton positions and the eye-gaze positions.

The details:

	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**

Tasks to subjects

The subjects: Six men and two women (Average 22 years old)

1. Subjects watched one video for 30 seconds.
2. They evaluated the video in four ordinal classes.
3. They repeated 1) and 2) until they watched all videos.

A web camera for Tobii Eye Tracker 4C

A web camera for OpenPose

※ Four ordinal classes: (4) very interesting, (3) a little interesting, (2) not interesting, (1) not interesting at all

*https://www.youtube.com ** https://gaming.tobii.com/tobii-eye-tracker-4c/

[*5] C. Szegedy, et al., *in Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
[21] Z. Cao, et al., *arXiv preprint arXiv:1812.08008*, 2018.

Experiment: Conditions

Setting

- 490 samples
(49 contents × 10 people)
  - Labeled training data (20%)
  - Unlabeled training data (50%)
  - Testing data (30%)

Evaluation metrics

$$\mathrm{MAE} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left| l_i^{\mathrm{P}} - l_i^{\mathrm{GT}} \right|$$

N_{test} : The number of testing data
l_i^{P} : The estimated interest level of i-th sample
l_i^{GT} : The actual interest level of i-th sample

Compared methods

SAGPLVM [19]
Our method without **Approach (i)**
semiOMGP-MLP [10]
Our method without **Approach (ii)**
mGPLVM-MLP [14]
The baseline

→ To confirm effectiveness of the **our approaches**

Baysian canonical correlation analysis (BCCA) [27]
Probabilistic method introducing Bayesian inference into CCA

→ To confirm effectiveness of the **mGPLVM-based model**

Multi-view CCA (MVCCA) [26]
A statistical methods
Deep CCA [28]
A method based on deep learning

→ To confirm effectiveness of the **probabilistic model**

※semiOMGP-MLP and mGPLVM-MLP are methods that introduce the projection based on the MLP in semiOMGP [10] and mGPLVM [14], respectively, in order to calculate the integrated feature of the new test sample by the content feature.

[19] J. Li, et al., *IEEE Transactions on Neural Networks and Learning Systems*, 2017.
[10] K. Kamikawa, et al., in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021.
[14] C. H. Ek, et al., in *Proc. International Workshop on Machine Learning for Multimodal Interaction*, 2007.
[27] A. Klami, et al., *Journal of Machine Learning Research*, 2013.
[26] J. Rupnik and J. Shawe-Taylor, in *Proc. Conference on Data Mining and Data Warehouses*, 2010.
[28] G. Andrew, et al., in *Proc. International Conference on Machine Learning*, 2013.

Subject	BPomGP	SAGPLVM	semiOMGP-MLP	mGPLVM-MLP	BCCA	MVCCA	Deep CCA
A	**0.632**	<u>0.714</u>	0.751	0.757	0.768	1.374	1.444
B	0.707	**0.670**	<u>0.683</u>	0.744	0.779	1.330	1.494
C	<u>0.651</u>	0.702	0.764	**0.639**	0.740	1.320	1.425
D	**0.690**	<u>0.738</u>	0.785	0.746	0.764	1.361	1.361
E	**0.624**	<u>0.651</u>	0.715	0.787	0.733	1.290	1.416
F	**0.640**	<u>0.719</u>	0.777	0.760	0.754	1.320	1.447
G	**0.667**	0.749	<u>0.677</u>	0.748	0.762	1.328	1.467
H	**0.643**	0.740	0.778	0.748	<u>0.731</u>	1.191	1.488
I	<u>0.695</u>	**0.686**	0.742	0.736	0.717	1.258	1.458
J	**0.606**	<u>0.726</u>	0.741	0.792	0.731	1.318	1.298
Average	**0.656**	<u>0.710</u>	0.741	0.746	0.748	1.309	1.430

BPomGP vs. MVCCA
Deep CCA (Non-probabilistic feature integration)

The effectiveness of being probabilistic models is confirmed.

Subject	BPomGP	SAGPLVM	semiOMGP-MLP	mGPLVM-MLP	BCCA	MVCCA	Deep CCA
A	**0.632**	0.714	0.751	0.757	0.768	1.374	1.444
B	0.707	**0.670**	0.683	0.744	0.779	1.330	1.494
C	0.651	0.702	0.764	**0.639**	0.740	1.320	1.425
D	**0.690**	0.738	0.785	0.746	0.764	1.361	1.361
E	**0.624**	0.651	0.715	0.787	0.733	1.290	1.416
F	**0.640**	0.719	0.777	0.760	0.754	1.320	1.447
G	**0.667**	0.749	0.677	0.748	0.762	1.328	1.467
H	**0.643**	0.740	0.778	0.748	0.731	1.191	1.488
I	0.695	**0.686**	0.742	0.736	0.717	1.258	1.458
J	**0.606**	0.726	0.741	0.792	0.731	1.318	1.298
Average	**0.656**	0.710	0.741	0.746	0.748	1.309	1.430

BPomGP vs. SAGPLVM (Our method without **Approach (i)**)

Defining the prior distribution of latent variables that introduces the ordering of interest levels

The effectiveness of Approach (i) is confirmed.

Subject	BPomGP	SAGPLVM	semiOMGP-MLP	mGPLVM-MLP	BCCA	MVCCA	Deep CCA
A	**0.632**	<u>0.714</u>	0.751	0.757	0.768	1.374	1.444
B	0.707	**0.670**	<u>0.683</u>	0.744	0.779	1.330	1.494
C	<u>0.651</u>	0.702	0.764	**0.639**	0.740	1.320	1.425
D	**0.690**	<u>0.738</u>	0.785	0.746	0.764	1.361	1.361
E	**0.624**	<u>0.651</u>	0.715	0.787	0.733	1.290	1.416
F	**0.640**	<u>0.719</u>	0.777	0.760	0.754	1.320	1.447
G	**0.667**	0.749	<u>0.677</u>	0.748	0.762	1.328	1.467
H	**0.643**	0.740	0.778	0.748	<u>0.731</u>	1.191	1.488
I	<u>0.695</u>	**0.686**	0.742	0.736	0.717	1.258	1.458
J	**0.606**	<u>0.726</u>	0.741	0.792	0.731	1.318	1.298
Average	**0.656**	<u>0.710</u>	0.741	0.746	0.748	1.309	1.430

BPomGP vs. semiOMGP-MLP (Our method without **Approach (ii)**)

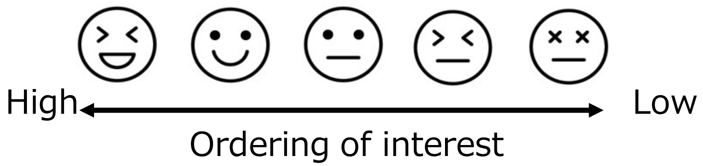
Defining the projection from content features to latent variables and optimized in training phase

The effectiveness of Approach (ii) is confirmed.

Conclusion

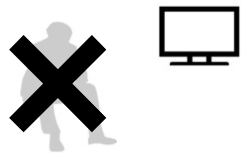
The Proposed method: Back-projection Ordering mGPLVM (BPomGP)

- Probabilistic feature integration model with label information available
- To adapt to the content recommendation situation, the following two problems can be addressed:



With conventional methods, it is difficult to reflect the ordering of the interest levels in the latent space.

Approach (i):
Defining the prior distribution of latent variables that introduces the ordering of interest levels



Since behavior features cannot be acquired in testing phase, it is difficult to calculate latent variables in the same way as in training phase.

Approach (ii):
Defining the projection from content features to latent variables and optimized in training phase

BPomGP can improve the accuracy of interest level estimation in content.

Contact Information

If you have any questions, please contact me.

Kyohei Kamikawa

Email: kamikawa@lmd.ist.hokudai.ac.jp