Multimed Tools Appl
DOI 10.1007/s11042-017-4717-7

Extracting hierarchical structure of content groups from different social media platforms using multiple social metadata

Daichi Takehara · Ryosuke Harakawa · Takahiro Ogawa · Miki Haseyama

Received: 15 October 2016 / Revised: 23 March 2017 / Accepted: 13 April 2017

Abstract A novel scheme for retrieving users' desired contents, *i.e.*, contents with topics in which users are interested, from multiple social media platforms is presented in this paper. In existing retrieval schemes, users first select a particular platform and then input a query into the search engine. If users do not specify suitable platforms for their information needs and do not input suitable queries corresponding to the desired contents, it becomes difficult for users to retrieve the desired contents. The proposed scheme extracts the hierarchical structure of content groups (sets of contents with similar topics) from different social media platforms, and it thus becomes feasible to retrieve desired contents even if users do not specify suitable platforms and do not input suitable queries. This paper has two contributions: (1) A new feature extraction method, Locality Preserving Canonical Correlation Analysis with multiple social metadata (LPCCA-MSM) that can detect content groups without the boundaries of different social media platforms is presented in this paper. LPCCA-MSM uses multiple social metadata as auxiliary information unlike conventional methods that only use content-based information such as textual or visual features. (2) The proposed novel retrieval scheme can realize hierarchical content structuralization from different social media platforms. The extracted hierarchical structure shows various abstraction levels of content groups and their hierarchical relationships, which can help users select topics related to the input query. To the best of our knowledge, an intensive study on such an application has not been conducted; therefore, this paper has strong novelty. To verify the effectiveness of the above contributions, extensive experiments for real-world datasets containing YouTube videos and Wikipedia articles were conducted.

Keywords Social media platform · Cross-platform application · Hierarchical structure · YouTube · Wikipedia

D. Takehara · R. Harakawa · T. Ogawa · M. Haseyama
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Japan
Tel.: +81-11-706-6078
Fax: +81-11-706-7369
E-mail: {takehara, harakawa, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

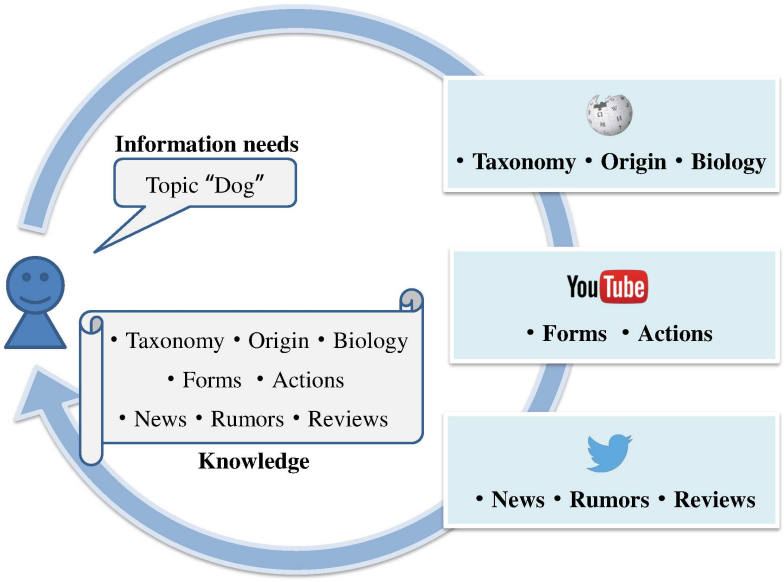


Fig. 1 Illustration of a user model to obtain information through multiple social media platforms.

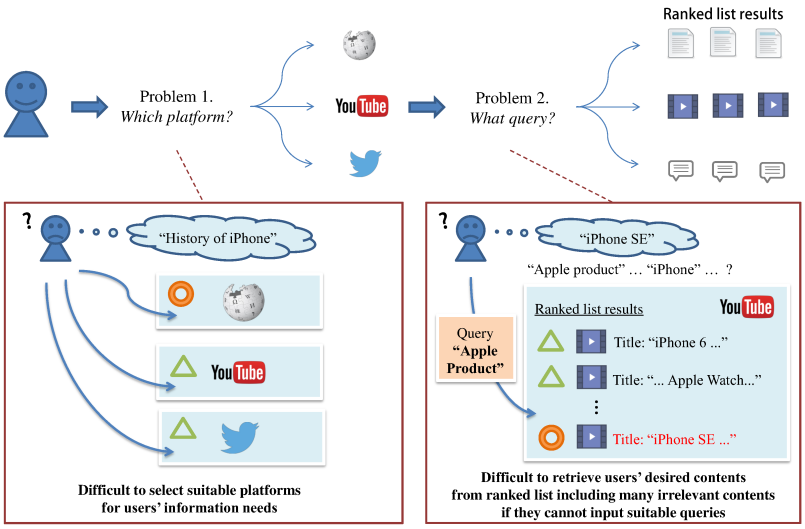
1 Introduction

With the development of various social media platforms such as Twitter[1], YouTube[2] and Wikipedia[3], ways for users to retrieve their desired contents, *i.e.*, contents with topics in which users are interested, have become more diversified. From such a social background, users tend to acquire knowledge through multiple social media platforms [41, 47] (see Fig. 1). For example, if users are interested in actions of animals, it may be desirable to watch videos in video hosting services such as YouTube. Meanwhile, if detailed information of animals such as taxonomy, origin and biology is needed, users should use knowledge bases such as Wikipedia.

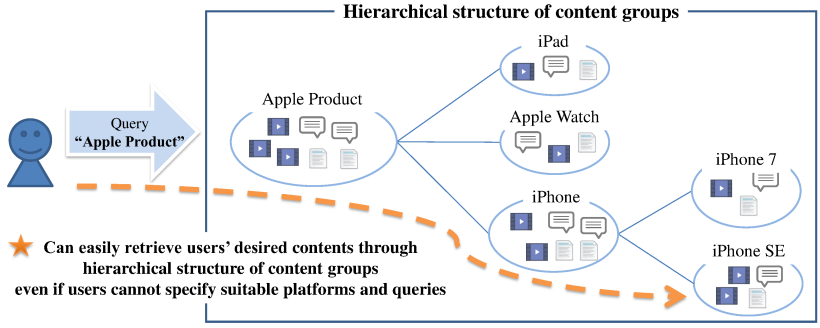
In existing retrieval schemes, when users want to retrieve desired contents from social media platforms, they first select a particular platform and then input a query into the search engine (see Fig. 2(a)). However, there are the following two problems for users to successfully retrieve desired contents from multiple social media platforms.

Problem 1: *Which platform should users select?*
Contents generated by different social media platforms often include different aspects of information about same topics. For example, given a topic "Soccer", users can learn about its rules or history in Wikipedia, while news or trends for soccer can be found in Twitter. However, it is difficult for users to understand such information in all social media platforms due to the large number of social

[1] https://twitter.com
[2] https://www.youtube.com
[3] https://en.wikipedia.org



(a) Conventional scheme: Users first select a particular platform and then input a query into the search engine. The users retrieve their desired contents from the ranked list results returned by each search engine.



(b) Proposed scheme: Users select a content group that includes desired contents from the hierarchical structure of content groups and then retrieve desired contents from the selected content group.

Fig. 2 Conventional and proposed schemes to retrieve desired contents from multiple social media platforms.

media platforms. Therefore, in general, it is difficult for users to select suitable platforms for their own information needs.

Problem 2: *What query should users input?*

Although users desire specific topics, they sometimes input ambiguous queries instead of suitable queries corresponding to the desired contents [7]. This difficulty in inputting suitable queries is increased by the diversity of platforms since suitable queries tend to be different for each platform. If suitable queries are not inputted, it is difficult for users to retrieve desired contents since the ranked list results include many irrelevant contents [19].

To solve the above problems, in this paper, we propose a novel scheme to retrieve desired contents from multiple social media platforms. The proposed scheme extracts the hierarchical structure of content groups from contents of different social media platforms related to the user's input query (see Fig. 2(b)). In this paper, content groups are defined as sets of contents with similar topics, and the hierarchical structure is defined as the property of content groups being divided into sub-groups. An intensive study on applications by hierarchical content structuralization from different social media platforms has not been conducted. Therefore, our work has a strong novelty. In the proposed scheme, users select a content group that includes desired contents from the extracted hierarchical structure and then retrieve the desired contents from the selected content group. The proposed scheme can solve the above problems by the following two strategies.

Strategy for solving problem 1:
The proposed scheme detects the content groups without the boundaries of different social media platforms. Thus, users are not required to select a particular platform and can retrieve contents across multiple platforms seamlessly. To detect content groups from different social media platforms, it is necessary to match distributions of features extracted from them. Usually, to meet this necessity, latent features are extracted from contents of different social media platforms by projecting features extracted from them into a common feature space on the basis of correlation learning [34] or topic modeling [28]. On the other hand, most of the social media contents have multiple social metadata that represents high-level semantics. For example, in YouTube videos, semantic relationships can be mined on the basis of social metadata such as tag-based relationships [22] and uploader-based relationships [6]. However, conventional methods have not utilized multiple social metadata (and there have been only a few works in which a single kind was utilized), and their performance may therefore be limited. To overcome this limitation, we have developed a new feature extraction method, Locality Preserving Canonical Correlation Analysis with multiple social metadata (LPCCA-MSM), that enables direct comparison of contents obtained from different social media platforms by learning a common feature space with preserving locality, *i.e.,* semantic information on the basis of multiple social metadata.

Strategy for solving problem 2:
The proposed scheme extracts the hierarchical structure of content groups related to the input query. Specifically, we first construct a heterogeneous graph for which nodes are contents of different social media platforms, enabling a direct comparison of contents of different social media platforms in a common feature space. We then hierarchically detect content groups in the heterogeneous graph on the basis of a well-known community detection method [4], and the hierarchical structure can thus be extracted. The extracted hierarchical structure shows various abstraction levels of content groups and their hierarchical relationships, which can help users grasp topics related to the input query. Thus, even if users do not input suitable queries, it becomes feasible to easily retrieve desired contents through the hierarchical structure.

Consequently, the proposed scheme enables users to easily retrieve desired contents in multiple social media platforms even if they cannot specify suitable platforms and suitable queries.

This paper has two contributions.

Contribution (1):
A new feature extraction method, LPCCA-MSM, that can detect content groups without the boundaries of different social media platforms is presented in this paper. LPCCA-MSM uses multiple social metadata as auxiliary information unlike conventional methods that only use content-based information such as textual or visual features.

Contribution (2):
The proposed novel retrieval scheme can realize hierarchical content structuralization from different social media platforms. The extracted hierarchical structure shows various abstraction levels of content groups and their hierarchical relationship, which can help users select topics related to the input query. To the best of our knowledge, an intensive study on such an application has not been conducted. Therefore, this paper has a strong novelty.

Experimental results for real-world datasets containing YouTube videos and Wikipedia articles verified the effectiveness of the new feature extraction method, *i.e.,* LPCCA-MSM, and a novel retrieval scheme in a hierarchical way.

2 Related work

In this section, we describe several works related to our work. Some works on the detection of content groups from different social media platforms [2,9,39,40,44,45] are strongly related to the strategy for solving problem 1 (see Section 1). The content groups detected in those works can provide complementary information delivered by multiple platforms, which is much richer information than the information from a single platform. For example, a framework to detect topics on Twitter by introducing New York Times and Flickr as complementary platforms was proposed [2]. Detection of content groups that consist of Web videos and news reports to provide a better description of topics was also reported [44]. In these works, mining of semantic relationships among contents of different social media platforms is a main technical issue. To successfully realize it, not only content-based information such as textual or visual features but also other high-level information such as social metadata was utilized in some works. For example, mining of semantic relationships among multi-modal contents obtained from different social media platforms by fusing two uni-modal graphs, *i.e.* text and visual graphs, with upload-time similarities [9, 45] and user behavior information [39] was performed. In another work, information on hot search queries was used as guidance to calculate similarities between contents of different platforms [40]. However, since multiple types of social metadata were not used in those works, the performance of the methods may be limited. Different from those works [2, 9, 39, 40, 44, 45], our proposed scheme utilizes multiple social metadata to successfully mine semantic relationships among different social media contents. The extension by using multiple social metadata contributes to the improvement of robustness of our feature extraction method.

Some works in which content groups were detected from search result contents [5, 7, 13, 18, 21, 42] are related to the strategy for solving problem 2 (see Section 1). On the basis of the detected content groups, users can get an overview

of the search result contents and easily retrieve desired contents even if they do not input suitable queries. Hierarchical clustering-based approaches [5, 13, 21] enable more effective retrieval by the extracted hierarchical structure that shows various abstraction levels of content groups and their hierarchical relationships. The retrieval of uni-modal contents was first studied [13], and the scope was then extended to multi-modal contents such as Web images [5] or Web videos [21]. However, to the best of our knowledge, those works are limited to applications that focus only on a single platform. Different from those works [5, 7, 13, 18, 21, 42], our scheme enables users to get an overview of contents in multiple social media platforms.

On the other hand, some graph-based approaches for multimedia content analysis [15, 20, 26, 29–31, 43] are also related to our proposed heterogeneous graph-based approach (See Section 3.3). These graph-based approaches, which perform content analysis by utilizing relationships between contents or between components of contents, have been proposed for several applications such as image retrieval [15, 26, 29, 43], online item recommendation [20] and venue recommendation in a location-based social network [31]. The motivation for graph-based analysis for cross-domain image annotation [30], which discovers common knowledge of each semantic concept from user-generated contents in different domains to boost the performance of semantic annotation, is similar to that of our work. Although many applications have been proposed before, hierarchical content structuralization from different social media platforms using heterogeneous graph-based analysis has not been proposed.

Finally, we note that this paper is an extended version of our earlier work [37]. The major difference is that the latent feature extraction method in our earlier work [37] is improved by utilizing multiple social metadata, *i.e.,* "related videos"[4], "tags" and "uploader". Only "related videos" were used in our earlier work.

3 Extracting hierarchical structure of content groups from different social media platforms

In this section, we present our proposed scheme to retrieve desired contents from multiple social media platforms.

3.1 Problem setting

Our scheme extracts the hierarchical structure of content groups from different social media platforms. An overview of our scheme is shown in Fig. 3. The input of our scheme is a set of contents obtained from different social media platforms $\mathcal{H}$. In this paper, we adopt a YouTube video set $\mathcal{Y}$ and a Wikipedia article set $\mathcal{W}$ ($\mathcal{H} = \mathcal{Y} \cup \mathcal{W}$). The output of our scheme is the hierarchical structure of content groups, which shows various abstraction levels of content groups and their hierarchical relationships. The details of the proposed method are explained below.

[4] YouTube videos related to each other are linked as "related videos".

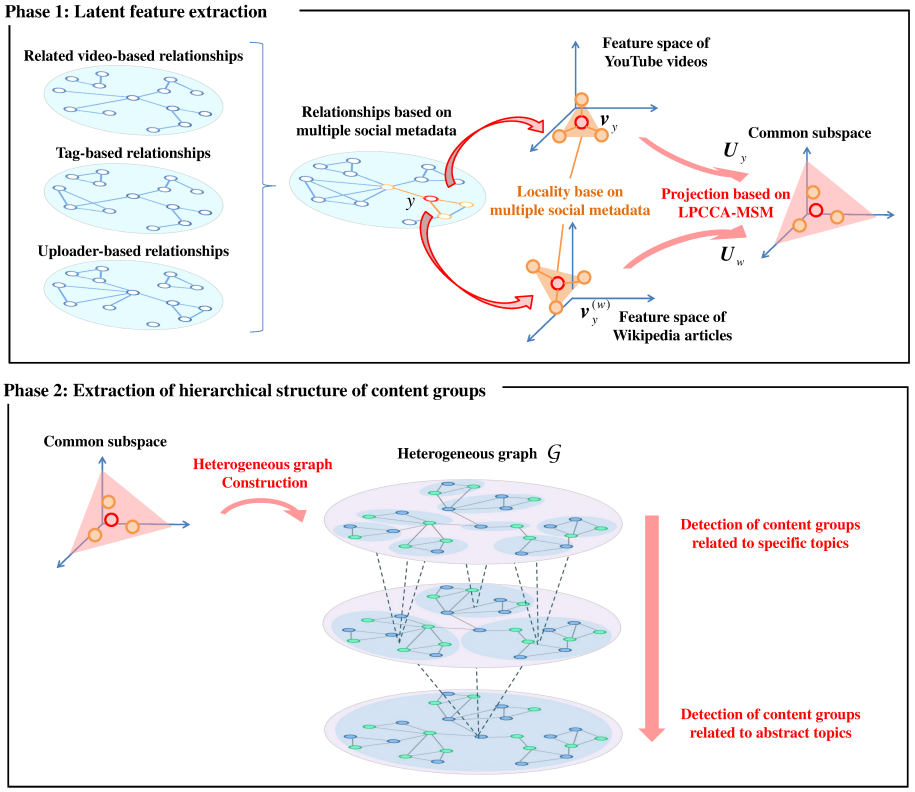


Fig. 3 Overview of the proposed scheme. In Phase 1, we extract latent features from YouTube videos and Wikipedia articles based on LPCCA-MSM. In Phase 2, based on the obtained latent features, we first construct a heterogeneous graph for which nodes are YouTube videos and Wikipedia articles. We then extract the hierarchical structure of content groups by hierarchically detecting content groups in the constructed heterogeneous graph.

3.2 Latent feature extraction from different social media contents using multiple social metadata

First, in our scheme, we obtain textual feature vectors $\boldsymbol{v}_y^t$ and $\boldsymbol{v}_w^t$ from YouTube videos $y \in \mathcal{Y}$ and Wikipedia articles $w \in \mathcal{W}$, respectively. We also obtain visual feature vectors $\boldsymbol{v}_y^v$ and $\boldsymbol{v}_w^v$ (details shown in Section 4).

Next, we extract latent features from the obtained features of YouTube videos and Wikipedia articles. Usually, mining semantic relationships among contents of different social media platforms is difficult due to the different distributions of features extracted from their contents [48]. Therefore, we extract latent features to directly compare contents of different social media platforms and mine their semantic relationships. We use a new feature extraction method, LPCCA-MSM, in our scheme. LPCCA-MSM can provide latent features that enable direct comparison of contents of different social media platforms by learning a common feature space with preservation of the locality, *i.e.,* semantic information on the basis of multiple social metadata (see Phase 1 in Fig. 3).

Preparation of the input for LPCCA-MSM

We first obtain pairs of features, *i.e.,* $\boldsymbol{v}_y$ and $\boldsymbol{v}_y^{(w)}$, from YouTube videos and Wikipedia articles to obtain the input of LPCCA-MSM. To obtain $\boldsymbol{v}_y$ and $\boldsymbol{v}_y^{(w)}$, we estimate a Wikipedia article set $\mathcal{R}^w(y)$, which is a set of Wikipedia articles similar to a YouTube video y. $\mathcal{R}^w(y)$ is defined as K most similar Wikipedia articles of y based on the following similarity:

$$\mathrm{sim}(\boldsymbol{v}_y, \boldsymbol{v}_w) = \left| \frac{\boldsymbol{v}_y \cdot \boldsymbol{v}_w}{\|\boldsymbol{v}_y\| \|\boldsymbol{v}_w\|} \right|, \tag{1}$$

where $\boldsymbol{v}_y = [\boldsymbol{v}_y^{t\,\mathrm{T}}, \boldsymbol{v}_y^{v\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{v}_w = [\boldsymbol{v}_w^t, \boldsymbol{v}_w^{v\,\mathrm{T}}]^{\mathrm{T}}$. Moreover, we obtain $\boldsymbol{v}_y^{(w)}$ by using features extracted from Wikipedia articles included in $\mathcal{R}^w(y)$ as follows:

$$\boldsymbol{v}_y^{(w)} = \frac{1}{\tau_y} \sum_{w \in \mathcal{R}^w(y)} \mathrm{sim}(\boldsymbol{v}_y, \boldsymbol{v}_w) \boldsymbol{v}_w, \tag{2}$$

$$\tau_y = \sum_{w' \in \mathcal{R}^w(y)} \mathrm{sim}(\boldsymbol{v}_y, \boldsymbol{v}_{w'}). \tag{3}$$

In this way, we obtain pairs of $\boldsymbol{v}_y$ and $\boldsymbol{v}_y^{(w)}$ to input LPCCA-MSM.

Introduction of locality based on multiple social metadata

Next, we define similarity matrices $\boldsymbol{S}_y$ and $\boldsymbol{S}_w$ to preserve the locality of YouTube videos by using multiple social metadata "related videos", "tags" and "uploader". The (p, q)-th elements $s_{y(p,q)}$ and $s_{w(p,q)}$ (of pairs of YouTube videos y_p and y_q) of similarity matrices $\boldsymbol{S}_y$ and $\boldsymbol{S}_w$ are defined as follows:

$$s_{y(p,q)} = \begin{cases} \mathrm{sim}(\boldsymbol{v}_{y_p}, \boldsymbol{v}_{y_q}) & \text{if } f_{p,q} = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

$$s_{w(p,q)} = \begin{cases} \mathrm{sim}(\boldsymbol{v}_{y_p}^{(w)}, \boldsymbol{v}_{y_q}^{(w)}) & \text{if } f_{p,q} = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where $f_{p,q}$ is equal to 1 if

(i) y_p is included in the metadata "related videos" of y_q (and vice versa)
or
(ii) one or more common tags are shared in the metadata "tags" of y_p and y_q
or
(iii) the metadata "uploader" of y_p and that of y_q are the same

and 0 otherwise. In this way, by collaboratively utilizing multiple types of social metadata, we can robustly represent locality even if a part of the social metadata includes noise.

Derivation of LPCCA-MSM

Finally, we derive projection to a common subspace of features extracted from YouTube videos and Wikipedia articles. Specifically, by considering the locality of YouTube videos based on $\boldsymbol{S}_y$ and $\boldsymbol{S}_w$ for estimating the correlation between $\boldsymbol{V}_y =$

Algorithm 1 : Latent feature extraction from different social media contents

Input: YouTube videos $y \in \mathcal{Y}$ and Wikipedia articles $w \in \mathcal{W}$
Output: Latent feature vector $\boldsymbol{v}_y^l$ and $\boldsymbol{v}_w^l$

1: **for** $y \in \mathcal{Y}$ **do**
2: Calculate $\boldsymbol{v}_y = [\boldsymbol{v}_y^{t\,\mathrm{T}}, \boldsymbol{v}_y^{v\mathrm{T}}]^{\mathrm{T}}$.
3: Estimate a related Wikipedia article set $\mathcal{R}^w(y)$ of y from $\mathcal{W}$.
4: Calculate $\boldsymbol{v}_y^{(w)}$ by using $\mathcal{R}^w(y)$ in Eqs. (2) and (3).
5: **end for**
6: Define similarity matrices $\boldsymbol{S}_y$ and $\boldsymbol{S}_w$ using multiple social metadata in Eqs. (4) and (5).
7: Calculate the projection matrices $\boldsymbol{U}_y$ and $\boldsymbol{U}_w$ from $\boldsymbol{V}_y = [\boldsymbol{v}_{y_1}, \cdots, \boldsymbol{v}_{y_{|\mathcal{Y}|}}]$, $\boldsymbol{V}_w = [\boldsymbol{v}_{y_1}^{(w)}, \cdots, \boldsymbol{v}_{y_{|\mathcal{Y}|}}^{(w)}]$, $\boldsymbol{S}_y$ and $\boldsymbol{S}_w$ in Eq. (6).
8: Calculate latent feature vector $\boldsymbol{v}_y^l$ and $\boldsymbol{v}_w^l$ in Eqs. (7) and (8).
9: **return** latent feature vector $\boldsymbol{v}_y^l$ and $\boldsymbol{v}_w^l$

$[\boldsymbol{v}_{y_1}, \cdots, \boldsymbol{v}_{y_{|\mathcal{Y}|}}]$ and $\boldsymbol{V}_w = [\boldsymbol{v}_{y_1}^{(w)}, \cdots, \boldsymbol{v}_{y_{|\mathcal{Y}|}}^{(w)}]$, we calculate the projection matrices $\boldsymbol{U}_y$ and $\boldsymbol{U}_w$, which maximize the following cost function, *i.e.,* the correlation:

$$\begin{aligned} \max_{\boldsymbol{U}_y, \boldsymbol{U}_w} \quad & \boldsymbol{U}_y^{\mathrm{T}} \boldsymbol{V}_y \boldsymbol{L}_{yw} \boldsymbol{V}_w^{\mathrm{T}} \boldsymbol{U}_w, \\ \text{s.t.} \quad & \boldsymbol{U}_y^{\mathrm{T}} \boldsymbol{V}_y \boldsymbol{L}_{yy} \boldsymbol{V}_y^{\mathrm{T}} \boldsymbol{U}_y = \boldsymbol{1}, \\ & \boldsymbol{U}_w^{\mathrm{T}} \boldsymbol{V}_w \boldsymbol{L}_{ww} \boldsymbol{V}_w^{\mathrm{T}} \boldsymbol{U}_w = \boldsymbol{1}, \end{aligned} \tag{6}$$

where $\boldsymbol{L}_{yw} = \boldsymbol{D}_{yw} - \boldsymbol{S}_y \circ \boldsymbol{S}_w$, $\boldsymbol{L}_{yy} = \boldsymbol{D}_{yy} - \boldsymbol{S}_y \circ \boldsymbol{S}_y$ and $\boldsymbol{L}_{ww} = \boldsymbol{D}_{ww} - \boldsymbol{S}_w \circ \boldsymbol{S}_w$. Note that $\circ$ denotes the Hadamard product, and $\boldsymbol{D}_{yw}$, $\boldsymbol{D}_{yy}$ and $\boldsymbol{D}_{ww}$ are diagonal matrices whose the i-th diagonal elements are equal to the sum of the elements in i-th row of $\boldsymbol{S}_y \circ \boldsymbol{S}_w$, $\boldsymbol{S}_y \circ \boldsymbol{S}_y$ and $\boldsymbol{S}_w \circ \boldsymbol{S}_w$, respectively. In Eq. (6), $\boldsymbol{U}_y$ and $\boldsymbol{U}_w$ are eigenvectors, which can be obtained by solving the generalized eigenvalue problem in the same manner as LPCCA [36]. Based on the obtained projections, we can calculate latent feature vectors $\boldsymbol{v}_y^l$ and $\boldsymbol{v}_w^l$ from $y \in \mathcal{Y}$ and $w \in \mathcal{W}$ as follows:

$$\boldsymbol{v}_y^l = \boldsymbol{U}_y^{\mathrm{T}} \boldsymbol{v}_y, \tag{7}$$

$$\boldsymbol{v}_w^l = \boldsymbol{U}_w^{\mathrm{T}} \boldsymbol{v}_w. \tag{8}$$

In this way, we can extract latent features from YouTube videos and Wikipedia articles. The latent features enable direct comparison of the contents in different social media platforms, and their semantic relationships can be mined on the baisis of multiple social metadata. For detailed procedures of latent feature extraction, refer to Algorithm 1.

3.3 Extraction of hierarchical structure of content groups based on a heterogeneous graph

Next, we extract the hierarchical structure of content groups from a set of contents of different social media platforms $\mathcal{H}$ $(= \mathcal{Y} \cup \mathcal{W})$. In our scheme, we first construct a heterogeneous graph for which nodes are YouTube videos and Wikipedia articles,

which enables direct comparison of the contents of different social media platforms in the common feature space. After that, we hierarchically detect content groups in the constructed heterogeneous graph, and consequently the hierarchical structure of content groups can be obtained (see Phase 2 in Fig. 3).

Specifically, we construct a heterogeneous graph $\mathcal{G} = (\mathcal{H}, \mathcal{E}, \mathcal{O})$, where $\mathcal{H}$, $\mathcal{E}$ and $\mathcal{O}$ represent sets of nodes, edges and edge weights, respectively. On the basis of the extracted latent features, we define a weight $o_{p,q} \in \mathcal{O}$ of an edge $e_{p,q} \in \mathcal{E}$ between nodes $h_p, h_q \in \mathcal{H}$ as follows:

$$o_{h_p,h_q} = \begin{cases} \mathrm{sim}(\boldsymbol{v}^l_{h_p}, \boldsymbol{v}^l_{h_q}) & \text{if } \mathrm{sim}(\boldsymbol{v}^t_{h_p}, \boldsymbol{v}^t_{h_q}) > \tau \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

This equation shows that the edges are first made on the basis of the textual features and then the edge weights are defined on the basis of the latent features. It has been reported that accurate clustering becomes feasible by semantically linking two nodes via textual features and then precisely weighting via visual features [5]. Motivated by that work, we introduce the approach in Eq. (9) into the proposed method.

Next, we hierarchically detect content groups in the constructed heterogeneous graph $\mathcal{G}$ on the basis of a previously reported method [4]. That method is a well-known algorithm that has been used for multimedia content analysis [17]. The method is based on optimization of modularity by iterating the following two processes, which is reported in multimedia content analysis.

Process 1: Detection of content groups

We assign each node $h \in \mathcal{H}$ to each different content group. For each node h, we evaluate the gain of modularity Q when a node h_p is set to a content group including a neighborhood node h_q, and then h_q is re-assigned to a content group for which the positive gain is maximum. Note that modularity Q is an evaluation measure for detecting the group structure from a graph, which is defined as follows [4].

$$Q = \frac{1}{2m} \sum_{h_p \in \mathcal{H}} \sum_{h_q \in \mathcal{H}} (o_{h_p,h_q} - \frac{k_{h_p} k_{h_q}}{2m}) \delta_{p,q}, \tag{10}$$

$$2m = \sum_{h_p \in \mathcal{H}} \sum_{h_q \in \mathcal{H}} o_{h_p,h_q}, \tag{11}$$

$$k_{h_p} = \sum_{h_p \in \mathcal{H}} o_{h_p,h_q}, \tag{12}$$

where $\delta_{p,q}$ is 1 if h_p and h_q belong to the same content group and 0 otherwise. This is applied to all nodes sequentially until there is no gain in the modularity.

Process 2: Construction of a new graph

In the second phase, we construct a new graph for which nodes are the content groups detected in Process 1. Here, the weight of edges between the two new nodes is the sum of the edge weights in the original graph. Also, each new node has a self-loop that is derived from the weighted edges of the corresponding original nodes obtained in the first process.

In this paper, a pair of the first and second processes is represented as "a pass" and this iteration number is denoted by t ($= 1, 2, \cdots, T$; T being the number of all passes). Furthermore, we iterate the passes, *i.e.,* detection of content groups

Algorithm 2 : Extraction of hierarchical structure of content groups

Input: A heterogeneous graph $\mathcal{G}$ whose nodes are YouTube videos y and Wikipedia articles w

Output: Content groups $\mathcal{C}_{n_t}^t$ $(t = 1, 2, \cdots, T;\ n_t = 1, 2, \cdots, N^t)$

```
Assign each node h ∈ H to each different content group.
Set the index of pass as t ← 1.
while True do
  while Modularity Q of G can gain do
    for each node h ∈ H do
      Evaluate the gain of Q when a node is set to each content group containing neighbourhood nodes.
      Re-assign a node to the content group for which the gain of Q is maximum.
    end for
    Calculate Q of G.
  end while
  Denote the detected content groups by C^t_{n_t}.
  if Modularity Q of G cannot gain then
    Break the while loop.
  end if
  Construct the new graph G with a self-loops whose nodes represent the detected content groups, where each edge weight is defined by the sum of the weights of edges in the original graph.
  t ← t + 1
end while
return Content groups C^t_{n_t}
```

and construction of the new graph, until there is no gain in modularity. In these iterations, we can extract the content groups $\mathcal{C}_{n_t}^t$ $(n_t = 1, 2, \cdots, N^t;\ N^t$ being the number of groups when the iteration is $t)$, which capture various abstraction levels based on their iterations. For detailed procedures of extraction of the hierarchical structure, refer to Algorithm 2.

In this way, the hierarchical structure of content groups can be extracted. The content groups are detected without the boundaries of different social media platforms. Thus, users are not required to select a particular platform and can retrieve contents across multiple platforms seamlessly. Moreover, the hierarchical structure shows various abstraction levels of content groups and their hierarchical relationships, which can help users select the topics related to the input queries. Therefore, the proposed scheme enables a user to easily retrieve the desired contents placed in different social media platforms even if the user dose not select suitable platforms and does not input suitable queries.

Table 1 Details of the datasets. Note that the columns of Concepts@*lev* represent the number of concepts on level *lev* in the sub-hierarchy, and the columns of $|\mathcal{Y}|$ and $|\mathcal{W}|$ represent the numbers of YouTube videos and Wikipedia articles, respectively.

Id	Root concept	Concepts@1	Concepts@2	Concepts@3	$\lvert\mathcal{Y}\rvert$	$\lvert\mathcal{W}\rvert$
1	"Sport"	20	37	59	1650	298
2	"Fish"	12	19	57	1108	398
3	"Bird"	22	67	181	3878	1218
4	"Mammal"	6	28	76	1191	211
5	"Vehicle"	7	34	83	1004	853
6	"Machine"	40	68	74	2712	1096
7	"Invertebrate"	14	42	74	1033	834
8	"Geological formation"	28	52	47	1155	826
9	"Insect"	34	80	102	2102	1167
10	"Animal"	36	51	130	2220	1445
11	"Beverage"	19	82	76	1587	1222
12	"Foodstuff"	21	74	136	3441	1556

4 Experimental Results

In this section, we verify the effectiveness of our proposed method. The experimental setting is described in Section 4.1, and experimental results are presented in Section 4.2.

4.1 Setting

We first describe the experimental setting. Specifically, we explain the datasets and the features, *i.e.,* textual and visual features, used in this experiment.

4.1.1 Datasets

In the experiment, we used a WordNet [27] hierarchy as a reliable hierarchical structure to verify the accuracy of the extracted hierarchical structure of content groups. The datasets were constructed in the following way. First, we manually selected one concept in the WordNet hierarchy as the "root concept". Next, we obtained the sub-hierarchy of the WordNet hierarchy that included up to three lower levels of concepts from the root concept. An example of the sub-hierarchy is shown in Fig. 4. Finally, in each leaf concept[5] of the obtained sub-hierarchy, we collected YouTube videos and Wikipedia articles by inputting the concepts as queries into search engines. We obtained the top 50 videos and 10 articles from search results when each query was given by using YouTube Data API[6] and Media Wiki API[7]. The details of the datasets are shown in Table 1.

[5] We define a leaf concept as a concept that has no lower levels of concepts in the hierarchy.
[6] https://developers.google.com/youtube/v3/
[7] https://www.mediawiki.org/wiki/API/

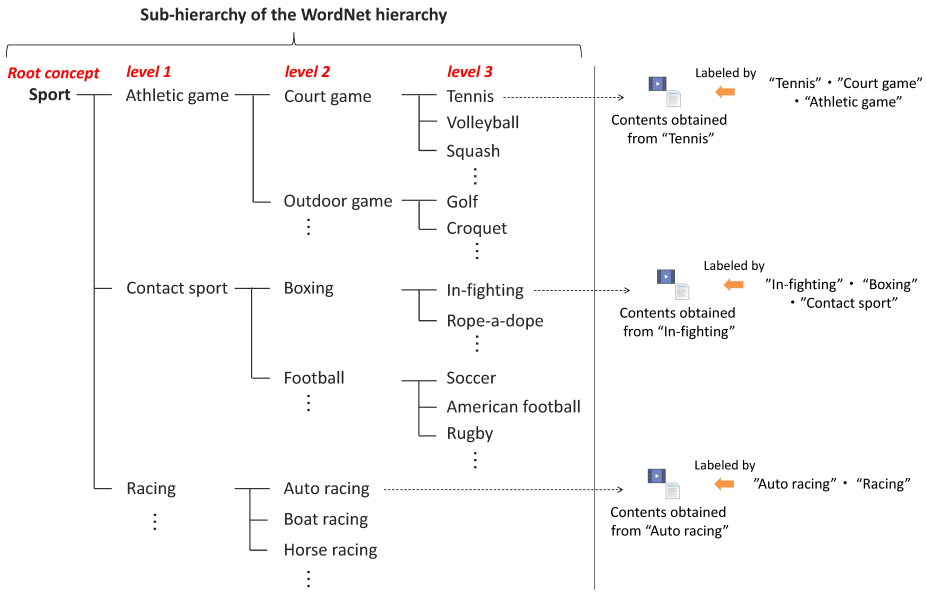


Fig. 4 Example of the sub-hierarchy. In the experiment, the leaf concept and its upper levels of concepts are served as ground truth labels of each YouTube video and Wikipedia article.

4.1.2 Features

The following features were used in the experiment.

Textual features:
We extracted textual features by applying Latent Semantic Indexing (LSI) [10] to text attached to YouTube videos, *i.e.*, titles, descriptions and tags, and text contained in Wikipedia articles. As a result, a 500-dimensional textual feature vector was obtained from each content.

Visual features:
Visual features were extracted from keyframes of YouTube videos and images included in Wikipedia articles. We first calculated Convolutional Neural Network (CNN) features [11] for each image contained in YouTube videos and Wikipedia articles. Then we applied locality-constrained linear coding (LLC) [38] to CNN features of each content. Consequently, a 500-dimensional visual feature vector was obtained from each content.

4.2 Results

In this subsection, we show and evaluate the experimental results. We qualitatively and qualitatively evaluate the results obtained by our proposed scheme.

4.2.1 Qualitative Evaluations

First, we qualitatively evaluate the experimental results to demonstrate the performance of the proposed method, *i.e.*, the effectiveness of contributions (1) and (2)

in this paper (See Section 1). Note that each of the effectiveness of contributions (1) and (2) are shown in Sec 4.2.2 "Quantitative evaluation". The hierarchical structure of content groups obtained from dataset 1, which was extracted by our proposed method, is shown in Fig. 5. From this figure, we can see that the detected content groups include different social media contents with similar topics; in other words, content groups can be detected without the boundaries of different social media platforms. By accessing the detected content groups, users are not required to select a particular platform and can retrieve contents across multiple platforms seamlessly. Thus, our scheme enables users to obtain information through multiple platforms, and such information in mush better than information obtained from only a single platform. Moreover, the extracted hierarchical structure shows various abstraction levels of content groups and their hierarchical relationships. Thus, users can easily select topics related to the input queries and easily access the content groups related to the detailed topics through the hierarchical structure even if they input ambiguous queries.

We explain cases in which users can benefit from the hierarchical structure.

Case 1. Users who input ambiguous queries and do not input specific queries that identify the desired contents:
Even if the user only inputs ambiguous queries such as "Sports" and "Equestrian sports" for desired contents related to "Dressage ", which is one of the equestrian sports, the user can find desired contents by browsing the hierarchical relationships between content groups (see Fig. 5). Moreover, there is a merit of newly discovering content groups related to the desired topic from the same or neighboring hierarchical levels.

Case 2. Users who desire both YouTube videos and Wikipedia articles:
The information a user desires tends to exist in diversified platforms. For example, a user who is interested in "Deer hunting" can learn about its actions or skills from YouTube videos as well as know its historical background from Wikipedia articles. In this way, our scheme enables users to obtain information through multiple platforms, which is much better than information obtained from only a single platform.

4.2.2 Quantitative Evaluations

Next, we quantitatively evaluate the experimental results. In the experiment, we defined the leaf concept and its upper levels of concepts as ground truth labels of each YouTube video and Wikipedia article (see Fig. 4). Thus, each content has hierarchical labels. For example, if a content is obtained by the query "tennis", the content also has upper levels of labels "court game" and "athletic game". Here, the F-measure is used as a typical evaluation measure in the clustering method [1]. Also, hierarchical clustering methods should be evaluated by considering the entire hierarchical structure [46]. Therefore, we evaluated the content groups $\mathcal{C}^t_{n_t}$ ($t = 1, 2, \cdots, T$; $n_t = 1, 2, \cdots, N^t$) by weighted average F-measure@lev (WAF@lev), which is defined as follows:

$$\mathrm{WAF@}lev = \frac{1}{T}\sum_{t=1}^{T}\sum_{n_t=1}^{N^t}\frac{|\mathcal{C}^t_{n_t}|}{\sum_{n_t{}'=1}^{N^t}|\mathcal{C}^t_{n_t{}'}|}\mathrm{F}_{lev}(\mathcal{C}^t_{n_t}), \tag{13}$$

Fig. 5 Hierarchical structure of content groups for dataset 1 extracted by our scheme. Note that the authors checked contents included in each content group and attached names for them manually in this figure. Also, examples of YouTube videos (thumbnails) and Wikipedia articles contained in each content group are shown.

where

$$\mathrm{F}_{lev}(\mathcal{C}^t_{n_t}) = \max_{l \in \mathcal{L}^{(lev)}} \left\{ \frac{2 \cdot \mathrm{Recall}_l(\mathcal{C}^t_{n_t}) \cdot \mathrm{Precision}_l(\mathcal{C}^t_{n_t})}{\mathrm{Recall}_l(\mathcal{C}^t_{n_t}) + \mathrm{Precision}_l(\mathcal{C}^t_{n_t})} \right\}, \tag{14}$$

and $\mathrm{Recall}_l(\mathcal{C}^t_{n_t})$ and $\mathrm{Precision}_l(\mathcal{C}^t_{n_t})$ are respectively defined as follows:

$$\mathrm{Recall}_l(\mathcal{C}^t_{n_t}) = \frac{|\mathcal{C}^t_{n_t} \cap \mathcal{D}_l|}{|\mathcal{D}_l|}, \tag{15}$$

$$\mathrm{Precision}_l(\mathcal{C}^t_{n_t}) = \frac{|\mathcal{C}^t_{n_t} \cap \mathcal{D}_l|}{|\mathcal{C}^t_{n_t}|}. \tag{16}$$

In the above equations, $\mathcal{D}_l$ is a set of contents with a label l in the dataset, and $\mathcal{L}^{(lev)}$ is a set of all kinds of labels on a hierarchical level lev. WAF@lev represents the weighted average of the F-measures for each content group based on labels of a hierarchical level lev. Hence, the higher the WAF@lev value is, the more accurate is the extraction of the hierarchical structure of content groups in terms of the hierarchical level lev.

Verification of the effectiveness of LPCCA-MSM.

First, we compare our novel feature extraction method *i.e.,* LPCCA-MSM, with other feature extraction methods to demonstrate the effectiveness of contribution (1) in this paper (See Section 1).

Table 2 shows WAF@1, 2 and 3 of our method and the following comparative methods.

LPCCA-SM [37]:
This is a method in our earlier work [37], *i.e.,* LPCCA with social metadata. This method extracts latent features from contents in the same manner; however, this method only uses the metadata "related videos".

CCA [24]:
This method extracts latent features from contents by canonical correlation analysis (CCA) [24]. Specifically, this method maximizes the correlations between features extracted from contents of different social media platforms without considering the locality of contents based on the social metadata.

Textual:
This method defines edge weights by similarities between contents based only on textual features $\boldsymbol{v}^t_y$ and $\boldsymbol{v}^t_w$.

Visual:
This method defines edge weights by similarities between contents based only on visual features $\boldsymbol{v}^v_y$ and $\boldsymbol{v}^v_w$.

TCA [32]:
This method extracts latent features from contents by transfer component analysis (TCA) [32]. TCA learns a common subspace across platforms as some transfer components in a reproducing kernel Hilbert space using maximum mean discrepancy.

GFK [16]:
This method extracts latent features from contents by a geodesic flow kernel (GFK) [16]. GFK integrates the inner products in an infinite sequence of feature subspaces that interpolates between different platforms.

Table 2 Comparison of the feature extraction methods based on WAF@1, 2 and 3 of the extracted hierarchical structure. Note that these comparative methods use the same clustering as that in our scheme but use different feature extraction methods. Bold emphases denote the highest evaluation values in each row.

(a) WAF@1

Dataset	Ours	LPCCA-SM [37]	CCA [24]	Textual	Visual	TCA [32]	GFK [16]	MSDA [8]
1	0.358	0.340	0.351	0.338	0.315	**0.366**	0.329	0.351
2	0.211	0.216	0.212	0.215	0.176	0.231	0.213	**0.232**
3	0.201	0.208	0.211	0.226	0.221	**0.268**	0.244	0.259
4	0.105	0.106	0.105	0.106	0.086	0.106	0.106	0.106
5	0.141	0.152	0.137	0.165	0.127	**0.189**	0.167	0.169
6	0.359	0.361	0.362	0.381	0.332	**0.392**	0.374	0.358
7	0.221	0.220	0.228	0.248	0.185	0.263	0.251	**0.267**
8	0.315	0.272	0.302	**0.325**	0.234	0.318	0.288	0.306
9	0.286	**0.313**	0.301	0.287	0.220	0.299	0.283	0.304
10	0.241	0.239	0.239	0.273	0.253	**0.332**	0.287	0.312
11	0.228	0.239	0.223	0.233	0.214	**0.291**	0.248	0.254
12	0.191	0.182	0.191	0.182	0.189	**0.200**	0.177	0.198
Mean	0.238	0.237	0.239	0.248	0.213	**0.271**	0.247	0.260

(b) WAF@2

Dataset	Ours	LPCCA-SM [37]	CCA [24]	Textual	Visual	TCA [32]	GFK [16]	MSDA [8]
1	0.424	0.417	0.429	0.421	0.387	**0.449**	0.407	0.424
2	0.317	**0.345**	0.332	0.321	0.268	0.341	0.329	0.319
3	0.419	**0.434**	0.423	0.385	0.334	0.367	0.366	0.367
4	0.340	0.337	0.337	**0.344**	0.260	**0.344**	0.343	**0.344**
5	0.364	0.372	0.357	0.383	0.309	**0.411**	0.372	0.369
6	**0.509**	0.485	0.475	0.488	0.418	0.486	0.463	0.478
7	0.441	**0.451**	0.427	0.440	0.363	0.429	0.430	0.431
8	0.475	0.419	0.476	**0.478**	0.347	0.453	0.408	0.428
9	**0.477**	0.475	0.463	0.416	0.328	0.407	0.395	0.399
10	0.372	0.378	0.373	0.391	0.347	**0.435**	0.396	0.426
11	0.321	0.316	**0.330**	0.315	0.246	0.295	0.276	0.288
12	**0.293**	0.290	0.288	0.274	0.254	0.271	0.253	0.281
Mean	**0.396**	0.393	0.392	0.388	0.322	0.391	0.370	0.380

(c) WAF@3

Dataset	Ours	LPCCA-SM [37]	CCA [24]	Textual	Visual	TCA [32]	GFK [16]	MSDA [8]
1	**0.616**	0.607	0.592	0.590	0.542	0.596	0.562	0.599
2	0.539	**0.542**	**0.542**	0.499	0.433	0.499	0.469	0.510
3	**0.559**	0.547	0.546	0.448	0.396	0.379	0.403	0.381
4	**0.701**	0.689	0.699	0.669	0.550	0.681	0.654	0.657
5	0.486	**0.490**	0.485	0.468	0.400	0.462	0.462	0.458
6	**0.478**	0.470	0.452	0.466	0.414	0.437	0.428	0.418
7	0.552	**0.564**	0.556	0.548	0.454	0.562	0.505	0.535
8	0.550	0.462	0.511	0.546	0.369	**0.551**	0.461	0.540
9	0.478	0.484	**0.488**	0.428	0.337	0.405	0.377	0.391
10	**0.504**	0.503	0.495	0.434	0.358	0.371	0.389	0.370
11	0.383	**0.407**	0.387	0.394	0.310	0.365	0.352	0.341
12	0.387	**0.401**	0.372	0.361	0.278	0.314	0.303	0.318
Mean	**0.519**	0.514	0.510	0.488	0.403	0.469	0.447	0.460

MSDA [8]:

This method extracts latent features from contents by marginalized stacked denoising autoencoders (MSDA) [8]. MSDA simplifies the denoising auto-encoder as a single linear denoiser for feature learning from different platforms.

It should be noted that the above comparative methods adopt the same scheme as that in our proposed method for extracting the hierarchical structure, but the latent features are different. Also, the parameters of our method and the comparative methods "LPCCA-SM [37]" and "CCA [24]" were set as $K = 5$, which is the number of the related Wikipedia articles $\mathcal{R}^w(y)$. For our method and the above comparative methods, the threshold τ for constructing the heterogeneous graph was set to the 99th percentile of the similarities $\mathrm{sim}(\boldsymbol{v}^t_{h_p}, \boldsymbol{v}^t_{h_q})$ of all pairs of contents on each graph.

From Table 2(b) and 2(c), we can confirm that our method outperforms all of the comparative methods in WAF@2 and 3, which are calculated by lower (specific) levels of the label. It can be seen that our method enables retrieval of desired contents with specific topics even if the user only inputs ambiguous queries. Note that, our method, "LPCCA-SM [37]" and "CCA [24]" show better performance than comparative methods "TCA [32]", "GFK [16]" and "MSDA [8]", which are state-of-the-art unsupervised domain adaptation methods. We consider that this is because the preprocessing for correlation learning in our method, "LPCCA-SM [37]" and "CCA [24]" obtains pairs of features with high relevance in advance (see Section. 3.1). Moreover, our method performs better than the comparative method "LPCCA-SM [37]", which only uses the social metadata "related videos". It was therefore confirmed that the use of multiple social metadata improves the robustness even if a part of social metadata includes noise.

On the other hand, our method shows worse results than some comparative methods in WAF@1, which is calculated by higher (abstract) levels of the label. The probable reason is that our method may ignore abstract semantic relationships by using social metadata, which tends to capture specific relationships among contents. This is an issue to be solved in the future. Specifically, our clustering method optimizes modularity as an evaluation measure for extracting the hierarchical structure, which is mainly calculated by the graph structure, although the edge weights capture semantic information based on the latent feature. Thus, our method cannot directly consider a measure based on abstraction levels of topics such as entropy of textual features [12]. Therefore, we consider that the above drawback can be solved by introducing a measure based on abstraction levels of topics to our clustering method.

Verification of the effectiveness of our hierarchical structure extraction approach.
Next, we compare our hierarchical structure extraction approach with other clustering methods to demonstrate the effectiveness of contribution (2) in this paper (See Section 1).

Table 3 shows WAF@1, 2 and 3 of our method and the following comparative methods.

AP [14]:
This method flatly detects content groups by affinity propagation (AP) [14], which is widely used in many research fields (*e.g.,* Web video clustering [21] and Web image clustering [23]).

RB [35]:
This method flatly detects content groups by repeated bisection (RB) [35], which is used for multimedia content clustering (*e.g.,* Web video clustering [25]).

Table 3 Comparison of the clustering methods based on WAF@1, 2 and 3 of the extracted hierarchical structure. Note that these comparative methods extract latent features in the same manner as that in our scheme but use different clustering methods. Bold emphases denote the highest evaluation values in each row.

(a) WAF@1

Dataset	Ours	AP [14]	RB [35]
1	**0.358**	0.224	0.248
2	**0.211**	0.107	0.109
3	**0.201**	0.077	0.083
4	0.105	0.093	**0.116**
5	**0.141**	0.048	0.052
6	**0.359**	0.214	0.227
7	**0.221**	0.081	0.075
8	**0.315**	0.167	0.165
9	**0.286**	0.116	0.117
10	**0.241**	0.122	0.121
11	**0.228**	0.099	0.099
12	**0.191**	0.072	0.074
Mean	**0.238**	0.118	0.124

(b) WAF@2

Dataset	Ours	AP [14]	RB [35]
1	**0.424**	0.324	0.360
2	**0.317**	0.164	0.168
3	**0.419**	0.270	0.279
4	**0.340**	0.316	**0.340**
5	**0.364**	0.179	0.172
6	**0.509**	0.292	0.321
7	**0.441**	0.220	0.194
8	**0.475**	0.286	0.257
9	**0.477**	0.263	0.257
10	**0.372**	0.188	0.185
11	**0.321**	0.250	0.228
12	**0.293**	0.183	0.186
Mean	**0.396**	0.245	0.246

(c) WAF@3

Dataset	Ours	AP [14]	RB [35]
1	**0.616**	0.466	0.496
2	**0.539**	0.347	0.345
3	**0.559**	0.492	0.491
4	**0.701**	0.677	0.685
5	**0.486**	0.344	0.327
6	**0.478**	0.303	0.338
7	**0.552**	0.337	0.279
8	**0.550**	0.244	0.235
9	**0.478**	0.312	0.300
10	**0.504**	0.463	0.467
11	**0.383**	0.257	0.242
12	**0.387**	0.328	0.318
Mean	**0.519**	0.381	0.377

From Table 3, we can see the effectiveness of our hierarchical structure extraction approach. We can quantitatively confirm that the hierarchical structure navigates

users to the desired contents with specific topics even if users only input ambiguous queries.

From the above, it can be seen that the hierarchical structure of content groups from different social media platforms can be extracted successfully. Therefore, the proposed scheme enables users to easily retrieve desired contents in different social media platforms even if users do not select suitable platforms and do not input suitable queries.

Although we used WordNet concepts as the input queries in this experiment to quantitatively evaluate our scheme in detail, our scheme can be applied to various queries (including a phrase or with multiple keywords such as "tennis in Japanese schools"). Specifically, on the basis of the input queries, we first obtain contents through the keyword search. We then obtain related contents of the obtained contents by following link relationships (such as hyperlink). From the obtained contents, we can extract the hierarchical structure of content groups.

4.3 Discussion on computational cost

In this subsection, we discuss the computational cost of our proposed scheme. First of all, we assume that offline clustering is unnecessary before the usage of the system since clustering and content group generation are performed after a user inputs the query into the system. Thus, there are some issues for real-world deployment in terms of computational cost. Specifically, the computational cost of the two phases constituting our scheme, *i.e.*, "Phase 1: Latent feature extraction" and "Phase 2: Extraction of hierarchical structure of content groups" (see Fig. 3), and their issues are shown as follows.

Phase 1: Latent feature extraction
The computational order is $O(n^3)$, where n is the number of contents in a target dataset, since we have to solve the generalized eigenvalue problem in LPCCA-MSM. Although it becomes a computational issue to deal with almost all data in social media platforms, this issue will be solved by implementing an efficient algorithm for CCA such as the algorithm in [17], which can reduce the computational costs by selecting a small number of representative contents for performing CCA.

Phase 2: Extraction of hierarchical structure of content groups
The computational order is $O(n\log n)$ in our adopted graph-based hierarchical clustering algorithm [4]. This algorithm is very efficient. Extraction of the hierarchical structure from 118 million contents took only 152 minutes on a standard PC [4]. Parallelization approaches [3, 33] for this algorithm will be useful to further improve the efficiency of this phase.

In this way, although there are some issues for real-world deployment at present, these issues are expected to be solved by increasing the efficiency of our scheme.

5 Conclusions

In this paper, we have proposed a novel scheme for retrieving desired contents from multiple social media platforms. Even if users do not specify suitable platforms

and do not input suitable queries, the proposed scheme enables users to retrieve desired contents by extracting the hierarchical structure of content groups from different social media platforms. To successfully extract the hierarchical structure, we introduce a new feature extraction method, *i.e.,* LPCCA-MSM, which enables direct comparison of contents obtained from different social media platforms by learning a common feature space with preservation of locality, *i.e.,* semantic information on the basis of multiple social metadata. Consequently, the content groups are detected without the boundaries of different platforms, and users are therefore not required to select a particular platform and can retrieve contents across multiple platforms seamlessly. Moreover, the hierarchical structure shows various abstraction levels of content groups and their hierarchical relationships, which helps users to select the topics related to the input query. Experimental results for real-world datasets containing YouTube videos and Wikipedia articles verified the effectiveness of hierarchical content structuralization from different social media platforms by our scheme.

Acknowledgements This work was partly supported by JSPS KAKENHI Grant Numbers JP25280036, JP24120002. We are grateful that a publisher, Springer permits us to deposit this accepted manuscript in the open access repository. The final publication is available at "https://link.springer.com/article/10.1007/s11042-017-4717-7".

References

1. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information retrieval **12**(4), 461–486 (2009)
2. Bao, B.K., Xu, C., Min, W., Hossain, M.S.: Cross-platform emerging topic detection and elaboration from multimedia streams. ACM Transactions on Multimedia Computing, Communications, and Applications **11**(4), 54 (2015)
3. Bhowmick, S., Srinivasan, S.: A template for parallelizing the louvain method for modularity maximization. In: Dynamics On and Of Complex Networks, vol. 2, pp. 111–124. Springer (2013)
4. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), P10,008 (2008)
5. Cai, D., He, X., Li, Z., Ma, W.Y., Wen, J.R.: Hierarchical clustering of www image search results using visual, textual and link information. In: ACM International Conference on Multimedia, pp. 952–959 (2004)
6. Cao, J., Zhang, Y., Ji, R., Xie, F., Su, Y.: Web video topics discovery and structuralization with social network. Neurocomputing **172**(C), 53–63 (2016)
7. Carpineto, C., Osiński, S., Romano, G., Weiss, D.: A survey of web clustering engines. ACM Computing Surveys **41**(3), 17 (2009)
8. Chen, M., Xu, Z., Sha, F., Weinberger, K.Q.: Marginalized denoising autoencoders for domain adaptation. In: ACM International Conference on Machine Learning, pp. 767–774 (2012)
9. Chu, L., Zhang, Y., Li, G., Wang, S., Zhang, W., Huang, Q.: Effective multi-modality fusion framework for cross-media topic detection. IEEE Transactions on Circuits and Systems for Video Technology **26**(3), 556–569 (2014)
10. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American society for information science **41**(6), 391 (1990)
11. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013)
12. Fang, Q., Xu, C., Sang, J., Hossain, M.S., Ghoneim, A.: Folksonomy-based visual ontology construction and its applications. IEEE Transactions on Multimedia **18**(4), 702–713 (2016)

13. Ferragina, P., Gulli, A.: A personalized search engine based on web-snippet hierarchical clustering. Software: Practice and Experience **38**(2), 189–225 (2008)
14. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
15. Gao, K., Zhang, Y., Luo, P., Zhang, W., Xia, J., Lin, S.: Visual stem mapping and geometric tense coding for augmented visual vocabulary. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3234–3241 (2012)
16. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2066–2073 (2012)
17. Harakawa, R., Ogawa, T., Haseyama, M.: [accurate and efficient extraction of hierarchical structure of web communities forweb video retrieval. ITE Transactions on Media Technology and Applications **4**(1), 49–59 (2016)
18. Harakawa, R., Ogawa, T., Haseyama, M.: A web video retrieval method using hierarchical structure of web video groups. Multimedia Tools and Applications **75**(24), 17,059–17,079 (2016)
19. Haseyama, M., Ogawa, T., Yagi, N.: A review of video retrieval based on image and video semantic understanding. ITE Transactions on Media Technology and Applications **1**(1), 2–9 (2013)
20. He, X., Zhang, H., Kan, M.Y., Chua, T.S.: Fast matrix factorization for online recommendation with implicit feedback. In: ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 549–558 (2016)
21. Hindle, A., Shao, J., Lin, D., Lu, J., Zhang, R.: Clustering web video search results based on integration of multiple features. World Wide Web **14**(1), 53–73 (2011)
22. Hong, R., Tang, J., Tan, H.K., Ngo, C.W., Yan, S., Chua, T.S.: Beyond search: Event-driven summarization for web videos. ACM Transactions on Multimedia Computing, Communications, and Applications **7**(4), 35 (2011)
23. Hong, R., Zha, Z.J., Gao, Y., Chua, T.S., Wu, X.: Multimedia encyclopedia construction by mining web knowledge. Signal Processing **93**(8), 2361–2368 (2013)
24. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3), 321–377 (1936)
25. Kamie, M., Hashimoto, T., Kitagawa, H.: Effective web video clustering using playlist information. In: Annual ACM Symposium on Applied Computing, pp. 949–956 (2012)
26. Liu, A., Nie, W., Gao, Y., Su, Y.T.: Multi-modal clique-graph matching for view-based 3d model retrieval. IEEE Transactions on Image Processing **25**(5), 2103–2116 (2016)
27. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
28. Min, W., Bao, B.K., Xu, C., Hossain, M.S.: Cross-platform multi-modal topic modeling for personalized inter-platform recommendation. IEEE Transactions on Multimedia **17**(10), 1787–1801 (2015)
29. Nie, L., Wang, M., Zha, Z.J., Chua, T.S.: Oracle in image search: A content-based approach to performance prediction. ACM Transactions on Information Systems **30**(2), 13 (2012)
30. Nie, W., Liu, A., Su, Y.: Cross-domain semantic transfer from large-scale social media. Multimedia Systems **22**(1), 75–85 (2016)
31. Nie, W., Liu, A., Zhu, X., Su, Y.: Quality models for venue recommendation in location-based social network. Multimedia Tools and Applications **75**(20), 12,521–12,534 (2016)
32. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. IEEE Transactions on Neural Networks **22**(2), 199–210 (2011)
33. Que, X., Checconi, F., Petrini, F., Gunnels, J.A.: Scalable community detection with the louvain algorithm. In: IEEE International Parallel and Distributed Processing Symposium, pp. 28–37. IEEE (2015)
34. Rasiwasia, N., Costa Pereira, J., Coviello, E., Doyle, G., Lanckriet, G.R., Levy, R., Vasconcelos, N.: A new approach to cross-modal multimedia retrieval. In: ACM International Conference on Multimedia, pp. 251–260 (2010)
35. Steinbach, M., Karypis, G., Kumar, V., et al.: A comparison of document clustering techniques. In: Workshop on Text Mining at ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 400, pp. 525–526 (2000)
36. Sun, T., Chen, S.: Locality preserving cca with applications to data visualization and pose estimation. Image and Vision Computing **25**(5), 531–543 (2007)
37. Takehara, D., Harakawa, R., Ogawa, T., Haseyama, M.: Hierarchical content group detection from different social media platforms using web link structure. In: IEEE International Conference on Image Processing, pp. 479–483 (2016)

38. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3360–3367 (2010)
39. Wang, S., Wang, Z., Jiang, S., Huang, Q.: Cross media topic analytics based on synergetic content and user behavior modeling. In: IEEE International Conference on Multimedia and Expo, pp. 1–6 (2014)
40. Xue, Z., Jiang, S., Li, G., Huang, Q., Zhang, W.: Cross-media topic detection associated with hot search queries. In: ACM International Conference on Internet Multimedia Computing and Service, pp. 403–406 (2013)
41. Zelenkauskaite, A.: Remediation, convergence, and big data conceptual limits of cross-platform social media. Convergence: The International Journal of Research into New Media Technologies p. 1354856516631519 (2016)
42. Zeng, H.J., He, Q.C., Chen, Z., Ma, W.Y., Ma, J.: Learning to cluster web search results. In: ACM SIGIR International Conference on Research and development in Information Retrieval, pp. 210–217 (2004)
43. Zhang, H., Shang, X., Luan, H., Wang, M., Chua, T.S.: Learning from collective intelligence: Feature learning using social images and tags. ACM Transactions on Multimedia Computing, Communications, and Applications **13**(1), 1 (2016)
44. Zhang, W., Chen, T., Li, G., Pang, J., Huang, Q., Gao, W.: Fusing cross-media for topic detection by dense keyword groups. Neurocomputing **169**, 169–179 (2015)
45. Zhang, Y., Li, G., Chu, L., Wang, S., Zhang, W., Huang, Q.: Cross-media topic detection: a multi-modality fusion framework. In: IEEE International Conference on Multimedia and Expo, pp. 1–6 (2013)
46. Zhao, Y., Karypis, G.: Evaluation of hierarchical clustering algorithms for document datasets. In: ACM International Conference on Information and Knowledge Management, pp. 515–524 (2002)
47. Zhou, X., Liang, X., Zhang, H., Ma, Y.: Cross-platform identification of anonymous identical users in multiple social media networks. IEEE Transactions on Knowledge and Data Engineering **28**(2), 411–424 (2016)
48. Zhuang, Y.T., Yang, Y., Wu, F.: Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval. IEEE Transactions on Multimedia **10**(2), 221–229 (2008)

Daichi Takehara received his B.S. degree in Electronics and Information Engineering from Hokkaido University, Japan in 2015. He is currently pursuing an M.S. degree at the Graduate School of Information Science and Technology, Hokkaido University. His research interests include audiovisual processing and Web mining. He is a student member of the IEEE and IEICE.

Ryosuke Harakawa received his B.S., M.S. and Ph.D degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2013, 2015 and 2016 respectively. He is currently a postdoctoral fellow at the Graduate School of Information Science and Technology, Hokkaido University. His research interests include audiovisual processing and Web mining. He is a member of the IEEE, IEICE, and Institute of Image Information and Television Engineers (ITE).

Takahiro Ogawa received his 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 in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of ITE Transactions on Media Technology and Applications. He is a member of the IEEE, EURASIP, IEICE, and Institute of Image Information and Television Engineers (ITE).

Miki Haseyama received her 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 of Washington University, USA from 1995 to 1996. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She has been a Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE), an Editor-in-Chief of ITE Transactions on Media Technology and Applications, a Director, International Coordination and Publicity of The Institute of Electronics, Information and Communication Engineers (IEICE). She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).