Int J Multimed Info Retr (2014) 3:193–205
DOI 10.1007/s13735-014-0056-x

REGULAR PAPER

Topic detection in cross-media: a semi-supervised co-clustering approach

Zhe Xue · Guorong Li · Weigang Zhang · Junbiao Pang · Qingming Huang

Received: 28 July 2013 / Revised: 25 March 2014 / Accepted: 10 April 2014 / Published online: 11 May 2014

Abstract With the rapid development of social media, the topics emerge and propagate in a variety of media websites. Although much work has been done since NIST proposed the problem of topic detection and tracking (TDT), most of them focus on single media data and are mainly based on unsupervised clustering method, which does not use some side information to help detecting topics. Therefore, traditional TDT approaches are not competent for cross-media topic detection. To efficiently use the information contained in multi-modal data from different sources and the prior knowledge, we propose a semi-supervised co-clustering approach for cross-media topic detection by a constrained non-negative matrix factorization. The correctness and convergence of our approach are proved to demonstrate its mathematical rigorousness. Experiments on the cross-media dataset verify the effectiveness of our proposed approach.

Z. Xue (✉) · G. Li (✉) · Q. Huang (✉)
University of Chinese Academy of Sciences, Beijing, China
e-mail: zhxue@jdl.ac.cn

G. Li
e-mail: grli@jdl.ac.cn

Q. Huang
e-mail: qmhuang@jdl.ac.cn

W. Zhang (✉)
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
e-mail: wgzhang@jdl.ac.cn

J. Pang
The College of Computer Science, Beijing University of Technology, Beijing, China
e-mail: jbpang@jdl.ac.cn

Q. Huang
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Keywords Cross-media · Topic detection · Semi-supervised clustering · Non-negative matrix factorization

1 Introduction

As the rapid advancement of Internet and multimedia technologies, social media websites become an important platform for people to access the interested information and share opinions. Concretely, media data emerged from multiple sources, such as news websites, video/photo sharing websites and social network websites; these media data represent different aspects of some real-world topics that people concern and give users a variety of experiences. However, the growing volume of media data makes it difficult for people to find what they are interested in. If topics are discovered from the large amounts of data automatically, users can know what is happening and quickly access the information they concern.

Topic detection is such an effort to discover topics from a collection of documents and group the documents belonging to the same subject. Topic detection for cross-media data integrates data with different modalities from multiple sources and detects hot topics implied in it. Although NIST proposed TDT in the 1990s [1], so far, most related works only focus on single-source media data, such as news data [2–5] and web videos [6,7]. Apparently, it is difficult to understand the world and conceive the topics if we do not exploit the media data from multiple sources. However, the traditional TDT approaches are barely suited for cross-media topic detection for the following reasons. First, the cross-media data have different modalities, and the data representation of each modality varies greatly; besides, the characteristic of the same modality in different sources is not the same. For example, textual information in news articles is abundant, while it (tag information) is sparse and noisy in web videos. Second, most of the TDT approaches are based on unsupervised clustering which ignores the possible side information. Some side information reflects user's intention and could provide strong indications of the real-world topics. User's queries recorded in search engine are such an example [8–10] and many of them are directly related to the topics that happened around us. The traditional TDT approaches do not think of using this side information as guidance for topic detection which makes them fail to detect topics more fully and effectively.

In this paper, we propose a cross-media topic detection approach based on a semi-supervised co-clustering (SSC). Our approach can solve these problems due to the following facts. First, by conducting SSC method, the cross-media data can be co-clustered simultaneously using the multi-modality information, and the information of the same modality from different sources can complement with each other. Second, our approach can use the guidance to help detecting topics. Some topics and part of their related data could be obtained through the guidance directly. To detect the topics more effectively, we should utilize the data co-clustering and the guidance simultaneously for topic detection. Semi-supervised clustering approach is a good way to achieve that purpose because of the following reasons. By introducing constraints (guidance) in co-clustering, the data labeled to the same category (topic) are made as tight as possible, and the unlabeled data that related to the prior known topics could also be clustered; besides, the influence of sparse and noise information on the clustering could be alleviated by adding the label information.

The data co-clustering is achieved by non-negative matrix tri-factorization. The cross-media data with different modalities can be represented by several non-negative matrices and each matrix R^{ci} denotes the relation between the central data and a feature modality. Then the clusters of the central data and every feature modality can be obtained by tri-factorizing these matrices simultaneously (co-clustering), and every cluster of the central data may correspond to an event (see Fig. 1). The co-clustering can integrate the information contained in different modalities so that we can discover the hidden global structure in the multi-modality data. Besides, NMF is a powerful tool for textual data clustering because each axis in the semantic space learned by the NMF has a much more straightforward correspondence with each latent topic than other SVD- and the eigenvector-based clustering methods [11]. Learning from the textual data among different sources simultaneously can make use of the complementary information sufficiently, and the learned axis will be enhanced which provides a better representation for a specific topic.

Some side information on the Internet can be used as the guidance for topic detection. In our case, the queries recorded in search engine are used. When users want to know

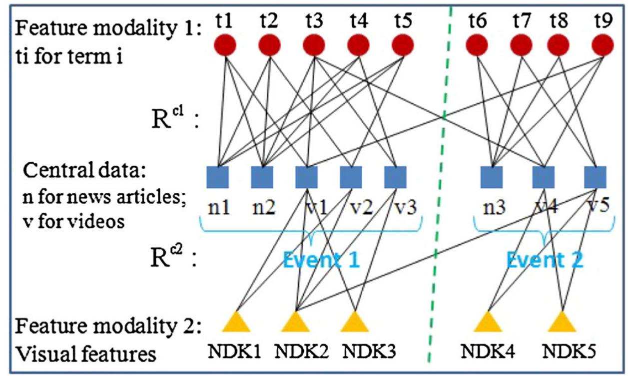

Fig. 1 Illustration of co-clustering cross-media data with different modalities

some topics happened in the real world, they usually search some information through the search engine. So the queries recorded in the search engine are closely related to the topics that happened in the world, and can be used as guidance for topic detection. However, not all the queries are related to real-world topics, so we preprocess the queries to get topic-related queries. Then, these queries are used to find the topic-related data (news articles, web video, etc) directly. Finally, we integrate this prior knowledge into co-clustering through introducing constraints into SSC to make the data belong to the same query (topic) as tight as possible.

Figure 2 shows the framework of the proposed approach. For clarity, the following two terminologies are defined. An *Event* is something that happens in a time slot, which consists of a set of closely related data such as news and videos. A *Topic* is a series of strongly interconnected events happened in several time slots. We know that the data related to the same event are close in time. To reduce the influence of irrelevant data and noise, we perform event detection in each time slot. From the framework we can see the cross-media data are first divided by time. Then in each time slot, the queries that belong to the corresponding time slot are preprocessed and find topic-related data to obtain the guidance information. The guidance information is converted to the label indicator matrix for SSC. Meanwhile, the multi-modality matrices are built based on the feature contained in different modalities. Next, we group the data into clusters under the guidance information by performing SSC, and each data cluster may be an event. After the events have been detected in different time slots, we thread these events together according to their text and time similarity to obtain topics. The major contribution of our work is summarized as follows:

1. Our approach could cluster the cross-media data utilizing the multi-modality features simultaneously which reduces the impact of the insufficient information of single modality on clustering. Besides, NMF-based text clustering could enhance the incomplete information in

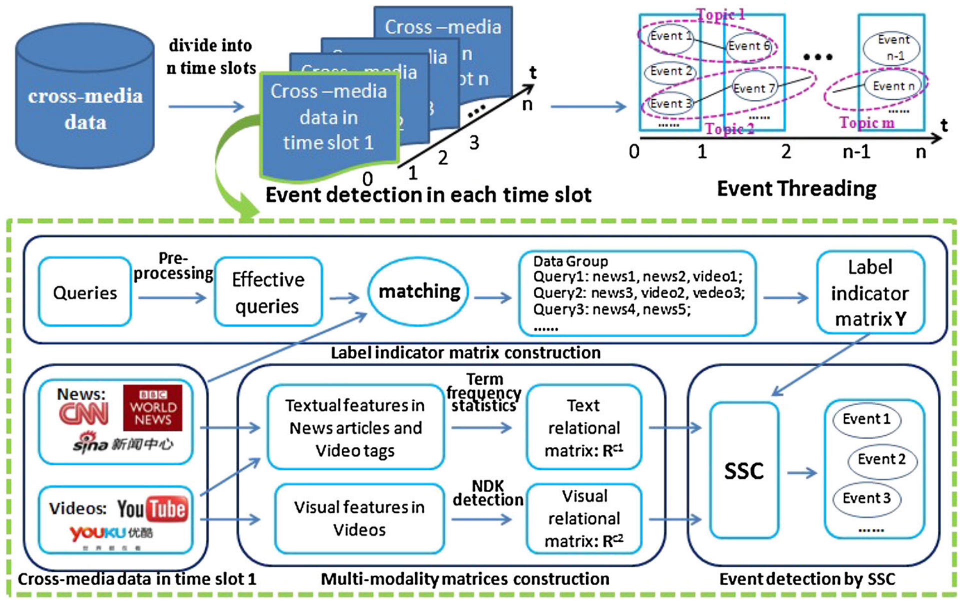

Fig. 2 The framework of our approach

cross-media data by making use of textual information from different sources complement with each other.

2. To the best of our knowledge, this is the first work that uses the guidance information for topic detection in a semi-supervised manner. Since some side information such as queries could give us guidance for topic detection, using them in cross-media topic detection can help us to obtain topics more directly. Through semi-supervised clustering, the topics are detected not only by data clustering, but also under the guidance of prior knowledge. The two ways could promote each other and improve effectiveness of topic detection.

The rest of this paper is organized as follows. We review some related works in Sect. 2. In Sect. 3, the SSC model is introduced in detail. How to obtain the guidance information and convert them into co-clustering constraints is described in Sect. 4. Event threading method is briefly described in Sect. 5. In Sect. 6 we provide the experimental results and finally the paper is concluded in Sect. 7.

2 Related works

Since TDT was proposed in 1990s, it attracts lots of research interests. The objective of TDT is to search, organize and structure the news oriented documents from news media, and it mainly involves five tasks: topic tracking, link detection, topic detection, first story detection and story segmentation. Topic tracking detects the stories that discuss previous known topic. Link detection associates detected stories that belong to the same topic. Topic detection detects previous unknown topics and groups stories discussing the same topic together. First story detection outputs a binary decision on whether the new story discusses a new topic or not. The task of story segmentation is to detect story changes [5].

Since the main source for people getting information is news articles in the early years, most works in topic detection mainly focus on detecting topics from news data [2–4,12,13]. The first benchmark evaluation corpus for TDT research is TDT1 which is proposed by researchers in TDT Pilot Research Project. There are 15,863 chronologically ordered news documents in this corpus which are mostly obtained from Reuters articles and CNN broadcasts. Larger and richer corpora have been developed by LDC [14]. Lots of TDT woks verify their performance through experiments on these corpora. Temporal Discriminative Probabilistic Model (DPM) [15] is an effective TDT work which is an unsupervised topic detection method for single media and it is shown to be theoretically equivalent to the classic vector space model with feature selection and temporal constraints. In the recent years, web video has become increasingly popular for its rich audio-visual content. Web users like to share some important or interesting video on the website. Topic detection for web video can recommend useful information to users and will enhance user experience. Therefore, some methods are developed for web video topic detection [6,7]. However, these efforts are designed for single-source media data. Topic detection for cross-media data has not been fully investigated. One such pioneer work to solve this challenge is [16]. Topics were detected for cross-media data from multiple sources such as news data from Sina [17] and web video from Youku [18].

Several techniques have been proposed on topic detection, and unsupervised clustering is a major way. For news data, Yang et al. [19] used the group-average clustering and single-pass clustering to detect topics. Wang et al. [20] adopted a traditional agglomerative clustering method to cluster news stories into topics. Due to the sparseness of the textual information and the limited visual information, topic detection for web video is a challenging task. Liu et al. [6] proposed a bipartite graph reinforcement model to densify the sparse textual information and reduce the noise. Several approaches try to take advantage of heterogeneous features which are extracted from the data in different modalities. Cao et al. [7] first clustered video tags into groups to get events, and then linked these events into topics based on textual and visual similarity. However, topic detection for cross-media data is more challenging, this is because the data from multiple sources have different modalities and the characteristic of the same feature modality varies among different sources. For example, news articles contain rich textual information. Web video is lack of textual information, but contains rich visual–audio content. How to utilize the complementary information between different sources is one of the key problems for cross-media topic detection. Zhang et al. [16] proposed a multi-graph fusing framework to cluster the data with different modalities for topic detection. Latent Dirichlet allocation (LDA) and near-duplicate keyframe (NDK) are used for measuring the textual similarity and the visual similarity, and two graphs are constructed based on the two similarities, respectively. After fusing the two types of graph, a graph clustering algorithm is conducted for detecting topics.

In recent years, co-clustering has attracted some researchers and is proven to be superior to traditional one-side clustering [21–23]. At first, some works are about homogeneous data co-clustering [22,23], then several co-clustering approaches are proposed for heterogeneous data clustering. Shao et al. [24] developed a star-structured K-partite Graph-based co-clustering framework for web video topic discovery. Textual features, visual features and audio features in videos are integrated for clustering through the K-partite Graph. Some works [25,26] also co-clustered the heterogeneous data based on graph theoretical. However, most co-clustering methods are based on graph model and require solving eigen-problem leading to high computational cost. So they are inapplicable for large-scale datasets. Xu et al. [11] have shown that NMF is more accurate and efficient than spectral methods in document clustering, and co-clustering via NMF has been proposed in [27]. Relational multimanifold co-clustering (RMC) [28] is a recent unsupervised heterogeneous data co-clustering method based on matrix factorization. It learns a convex combination of some predefined manifolds to maximally approach the desired intrinsic manifold. Compared to unsupervised method, semi-supervised learning can learn from both labeled and unlabeled data, which makes it generate better clustering result. Li et al. [29] learn a high-quality sentiment model using a constrained co-clustering approach. The constraints enable learning from partial supervision along both dimensions of the term-document matrix. Then an effective sentiment model is built by appropriately constraining the factors using prior knowledge. Some semi-supervised co-clustering methods based on matrix factorization for heterogeneous data have been proposed in [30,31]. Semi-supervised NMF (SS-NMF) [31] is a semi-supervised heterogeneous data co-clustering approach which computes new relational matrices by incorporating prior knowledge through distance metric learning and modality selection.

Some efforts have been made to use queries as guidance for topic detection, because they are an important clue for what happened in the world. When some event happens, users will search the related information about it through the search engine. So queries can provide strong indication of the events that users are interested in. Sun et al. [8] detected topics in a user-oriented manner and proposed a query-guided event detection method for news and blogs. However, it needs a search engine to obtain relevant documents, and cannot be applied to the specified dataset. Moreover, detecting topics entirely based on queries cannot fully find topics due to the following facts. The queries obtained from websites are the most popular ones, which can only represent part of the hottest topics concerned by users. Therefore, we cannot get the corresponding queries for all topics. The queries are also used to refine the topic detection results which are obtained by tag group mining in [9]. But the topic detection relies mainly on tag groups. The queries are only used for refining the results of tag group detection which does not consider detecting topics directly based on them. [32] is our previous work which uses hot queries to detect the topics directly, and then the topics that are unrelated to queries are detected by co-clustering. The method to obtain the prior knowledge from queries is applied in this work.

3 SSC for cross-media topic detection

In this section, we first briefly introduce the co-clustering approach via NMF in Sect. 3.1. In Sect. 3.2, the proposed SSC model is described in detail. The correctness and convergence of SSC are proved in appendix.

3.1 Co-clustering via NMF

Lee et al. [33] first proposed NMF for learning the parts of object and then many clustering methods have been proposed based on NMF [27,31,34]. Given a set of m-dimensional data vectors $\mathbf{x}_i$, $i \in [1, n]$, they are placed in the rows of an $n \times m$ matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n)^T \in R^{n\times m}$. To make it easily understandable, we assume that $\mathbf{X}$ is a $n \times m$ document-term matrix. NMF factorizes the input non-negative data matrix X into two non-negative matrices,

$$\mathbf{X} \approx \mathbf{F}\mathbf{G}^T \tag{1}$$

where F is a $n \times k$ cluster indicator matrix, G is a $m \times k$ matrix and k is the number of clusters, which is smaller than m and n. Each row of F represents the degree that a document belong to the cluster k, and the columns of G are the dictionary elements or cluster centers learned by NMF. Although many ways can be used to measure the difference between $\mathbf{X}$ and $\mathbf{FG}^T$, the most commonly used measure is the Frobenius norm [11,27,31]:

$$F(\mathbf{X}, \mathbf{FG}^T) = \left\| \mathbf{X} - \mathbf{FG}^T \right\|_F^2 = \sum_{p=1}^{m} \sum_{q=1}^{n} [\mathbf{X}_{pq} - (\mathbf{FG}^T)_{pq}]^2 \tag{2}$$

where $(\cdot)_{pq}$ represents the element of the matrix in row p and column q. The most popular algorithm for the NMF problem is the multiplicative rules proposed by Lee et al. [33].

Ding et al. [27] provided a 3-factor NMF

$$\mathbf{X} \approx \mathbf{FSG}^T \tag{3}$$

which simultaneously cluster the row and column of data matrix $\mathbf{X}$. $\mathbf{F}$ and $\mathbf{G}$ are the cluster indicator matrices for rows and columns of $\mathbf{X}$, respectively, and matrix $\mathbf{S}$ reflects the association between data clusters and feature clusters. However, this model can only deal with homogeneous data clustering. To cluster heterogeneous data, Chen et al. [31] presented a semi-supervised co-clustering model based on matrix tri-factorization. But they use metric learning to learn a new matrix containing the guidance information making this method not computationally efficient. In our case, we use the guidance directly and obtain satisfactory results.

3.2 Problem formulation

Let $C = \{x_1, x_2, \ldots, x_{n_c}\}$ denote the central dataset and each element in C is a sample from cross-media data (a news article or a video). The h feature modalities are denoted by $V_1, V_2, \ldots, V_h$ so that each set $V_i (1 \leq i \leq h)$ represents one type of feature such as textual feature (terms) or visual feature (NDK). The goal of co-clustering is to partition the dataset C into k_c clusters and divide each feature modality V_i into k_i clusters simultaneously. Let $\mathbf{R}^{(ci)} \in \boldsymbol{R}^{n_c \times n_i}$ represent the relation between the central dataset C and a feature modality set $V_i (1 \leq i \leq h)$, where n_i is the dimension of the feature modality V_i such as the total number of terms or NDKs. For textual relational matrix $\mathbf{R}^{(c1)}$, the p-th row of $\mathbf{R}^{(c1)}$ is the term frequency vector of data x_p. Both news article and video data are constructed in this way. Then, for visual relational matrix $\mathbf{R}^{(c2)}$, the p-th row of $\mathbf{R}^{(c2)}$ is constructed by:

$$\mathbf{R}_{pq}^{(c2)} = \begin{cases} 1 & \text{if data } p \text{ is a video and contains NDK } q \\ 0 & \text{else} \end{cases}$$

So for news articles, the corresponding row of $\mathbf{R}^{(c2)}$ is a zero vector.

Then co-clustering for cross-media data can be achieved by non-negative tri-factorization of $\mathbf{R}^{(ci)}$ as follows:

$$\min_{\mathbf{F} \geq 0, \mathbf{S}^{(ci)} \geq 0, \mathbf{G}^{(i)} \geq 0} \sum_{i=1}^{h} \left\| \mathbf{R}^{(ci)} - \mathbf{FS}^{(i)} \mathbf{G}^{(i)} \right\|_F^2 \tag{4}$$

where $\mathbf{F} \in \boldsymbol{R}^{n_c \times k_c}$ shows the clustering results of the cross-media data in C, $\mathbf{G}^{(i)} \in R^{k_i \times n_i}$ reflects the clustering results of the features in V_i, $\mathbf{S}^{(i)} \in R^{k_c \times k_i}$ indicates the association between the data clusters and the feature clusters of V_i and $h = 2$ represents using two relational matrices (textual and visual).

To make use of the guidance information, most traditional methods formulate the guidance information as pairwise constraints, but it is very complicated to construct constraints using this formulation in our case. After query and data matching, the data matched to the same query are considered to belong to the same topic, so they are labeled to the same category. Since the category of the labeled data is directly presented, we can construct constraints directly and do not need to convert them into pairwise ones.

We introduce label indicator matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_{n_c})^T \in \boldsymbol{R}^{n_c \times k_c}$ where $\mathbf{y}_i \in \{0, 1\}^{k_c \times 1}$ is the label vector of data x_i, and $\mathbf{Y}$ is constructed as follows:

$$\mathbf{Y}_{ij} = \begin{cases} 1 & \text{if data } \mathrm{x_i} \text{ is labeled to the cluster j} \\ 0 & \text{else} \end{cases}$$

To incorporate the prior knowledge $\mathbf{Y}$ into co-clustering, we include an auxiliary constraint on $\mathbf{F}$ to make the clustering results of the labeled data and the original label information as close as possible. So the penalty term is given by

$$\begin{aligned} &\min_{\mathbf{F}} \sum_{p=1}^{n_c} \mathbf{U}_{pp} \sum_{q=1}^{k_c} (\mathbf{F}_{pq} - \mathbf{Y}_{pq})^2 \\ &= \min_{\mathbf{F}} Tr[(\mathbf{F} - \mathbf{Y})^T \mathbf{U} (\mathbf{F} - \mathbf{Y})] \end{aligned} \tag{5}$$

where $\mathbf{U}$ is a diagonal matrix and is defined as:

$$\mathbf{U}_{pp} = \begin{cases} 1 & \text{if data } \mathrm{x}_p \text{ is labeled} \\ 0 & \text{else} \end{cases}$$

Then the SSC model is derived by including this penalty term:

$$\begin{aligned} J = &\min_{\mathbf{F} \geq 0, \mathbf{S}^{(i)} \geq 0, \mathbf{G}^{(i)} \geq 0} \sum_{i=1}^{h} \left\| \mathbf{R}^{(ci)} - \mathbf{FS}^{(i)} \mathbf{G}^{(i)} \right\|_F^2 \\ &+ \alpha Tr[(\mathbf{F} - \mathbf{Y})^T \mathbf{U} (\mathbf{F} - \mathbf{Y})] \end{aligned} \tag{6}$$

where $\alpha \in [0, +\infty)$ is the parameter that controls the enforcement degree of the prior knowledge.

Through updating $\mathbf{S}$, $\mathbf{G}$ and $\mathbf{F}$ iteratively, we could obtain the solution of problem (6). The details of our algorithm are provided in Algorithm 1. After factorization, the data x_p is assigned to the cluster k_i if $k_i = \arg\max_j F_{pj}$.

Algorithm 1. SSC for cross-media data co-clustering

INPUT: Relation matrices $\{\mathbf{R}^{(ci)}\}_{1\le i\le h}$;

Label indicator matrix $\mathbf{Y}$

OUTPUT: Cluster indicator matrices $\mathbf{F}$, $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$ and cluster association matrix $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$.

METHOD:

1. Randomly initialize $\mathbf{F}$, $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$, $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$ with non-negative values;
2. Repeat the following steps till convergence:

(a). Fixing $\mathbf{F}$ and $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$, update $\mathbf{S}^{(i)}$ for each $i\,(1\le i\le h)$ according to:

$$\mathbf{S}_{pq}^{(i)} \leftarrow \mathbf{S}_{pq}^{(i)} \frac{\left(\mathbf{F}^T\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\right)_{pq}}{\left(\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\right)_{pq}} \tag{7}$$

(b). Fixing $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$ and $\mathbf{F}$, update $\mathbf{G}^{(i)}$ for each $i\,(1\le i\le h)$.

$$\mathbf{G}_{pq}^{(i)} \leftarrow \mathbf{G}_{pq}^{(i)} \frac{\left(\mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{R}^{(ci)}\right)_{pq}}{\left(\mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\right)_{pq}} \tag{8}$$

(c). Fixing $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$ and $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$, update $\mathbf{F}$.

$$\mathbf{F}_{pq} \leftarrow \mathbf{F}_{pq} \frac{\left(\sum_{i=1}^{h}\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{Y}\right)_{pq}}{\left(\sum_{i=1}^{h}\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{F}\right)_{pq}} \tag{9}$$

4 Obtain the constraints in SSC

As mentioned in Sect. 1, the queries retained in the search engine are used for guidance information for topic detection. However, there are two problems in using the queries. First, not all queries are related to topics [8]. For example, some of them are website names and they should be removed (denoising) before topic detection. Besides, as different queries may relate to the same topic, these queries should be treated as one (merging). Second, the queries have too little information to find the related data because they are usually composed of just a few terms.

To solve these two problems, we search a query in the search engine to obtain relevant information to enrich the query. The search results of each query are crawled (denoted as background information or BI) and only textual information is used. Based on BI, we propose a query preprocessing method which includes query denoising and merging to obtain effective queries in Sect. 4.1. After that, we conduct query and data matching to find the related data for each query in Sect. 4.1.1.

4.1 Preprocessing

To preprocess the queries, we first remove the queries that are not topic related and then merge the ones related to the same topic.

4.1.1 Denoising

The i-th search result for query q is denoted by a term vector (tf-idf) d_i and the BI of q is represented by $Bq =$

$\{d_1, d_2, \ldots, d_N\}$. We find that the search results of topic-related queries are similar with each other because most of them are the related news, reports, blogs etc. But the search results of topic-unrelated queries are not consistent with each other. For example, the search results of a website name usually contain the hyperlinks of its different pages and the content of the search results differs a lot.

To measure the similarity between the search results in BI, the average cosine distance is used and defined as:

$$AverSim(q) = \frac{1}{n} \sum_{d_i, d_j \in B_q, i \neq j} \cos(d_i, d_j) \tag{10}$$

where n is the total number of calculations. Smaller $AverSim$ means the difference between search results is larger. So if $AverSim$ is smaller than a threshold (we set 0.65 in our case), then q could be treated as topic unrelated and removed.

4.1.2 Merging

If two queries are related to the same topic, they may contain some same terms (a query usually consists of several terms). Besides, their content of BI should similar with each other. These are the two criteria that we use to merge similar queries. The m-th term in query is denoted ast_m. Then the term similarity between two queries q_i, q_j is given by:

$$Sim\_Term(q_i, q_j) = \sum_{t_m \in q_i \cap q_j} t_m \Big/ \left(\sum_{t_n \in q_i \cup q_j} t_n \right) \tag{11}$$

We can see that the greater similarity is, the more same terms that the two queries contain.

The cosine distance is used for defining the BI similarity between q_i and q_j:

$$Sim\_BI(q_i, q_j) = \cos(\overline{Bq_i}, \overline{Bq_j}) \tag{12}$$

where $\overline{Bq_i}$ and $\overline{Bq_j}$ are the average term vectors of all the search result vectors in Bq_i and Bq_j, respectively.

Then the total similarity between $q_i$and $q_j$is defined as:

$$Sim(q_i, q_j) = \lambda \cdot Sim\_BI(q_i, q_j) + (1 - \lambda) \cdot Sim\_Term(q_i, q_j) \tag{13}$$

If the total similarity is greater than the threshold th_1, then the two queries are treated as one. We set $\lambda = 0.6$ and $th_1 = 0.45$.

4.2 Matching with data

Since BI contains some related information of the query, it can enrich and densify textual information for matching. Since the cross-media data from different sources have different characteristics. For example, textual information in news articles is abundant, but the videos only have some limited user-supplied tags. It is difficult for matching query with videos directly. So for videos, the similarity between query q_i and video v_j is defined as:

$$Sim\_QV(q_i, v_j) = \theta \cdot Sim\_BIV(q_i, v_j) + (1 - \theta) \cdot Sim\_Term(q_i, v_j) \tag{14}$$

where $Sim\_Term(\cdot)$ is the term similarity between the query terms and video tags defined by Eq. (11). $Sim\_BIV(\cdot)$ is the similarity between BI of query q_i and the tags of v_j, given by:

$$Sim\_BIV(q_i, v_j) = \sum_{w \in Bq_i \cap v_j} TF - IDF(w) \tag{15}$$

We can see that even there are no common terms between video tags and query terms, they can be matched as long as the BI bridges them. Both $Sim\_BIV(\cdot)$ and $Sim\_Term(\cdot)$ should be normalized to [0, 1] before summing in Eq. (14). $\theta \in [0, 1]$ is the parameter controls the weights of the two terms and we set $\theta = 0.5$ in our case.

Cosine distance is used to measure the similarity between query q_i and news n_j:

$$Sim\_QN(q_i, n_j) = \cos(\overline{Bq_i}, n_j) \tag{16}$$

where $\overline{Bq_i}$ is same as in Eq. (12).

The video and news are assumed to be matched to the query if the similarity score calculated by Eqs. (14) and (16) is larger than the threshold th_2 and th_3, respectively.

5 Event threading

A topic may contain several related events that happened at different time slots. So the detected events by SSC in each time slot may relate to the same topic and should be threaded into topics. These events are both similar in textual content and very close in time. We define the similarity between two events e_i and e_j as follows:

$$Sim\_event(e_i, e_j) = Sim\_content(e_i, e_j) \times Sim\_time(e_i, e_j) \tag{17}$$

The textual information of each event is represented by the average term vector of all the documents clustered into this event. Cosine distance is used as content similarity $Sim\_content(e_i, e_j)$ between two events.

The time similarity is given by:

$$Sim\_time(e_i, e_j) = e^{-\beta \cdot \frac{dt(e_i, e_j)}{T}} \quad (18)$$

where $dt(e_i, e_j) = \begin{cases} \left|t(e_i) - t(e_j)\right| & if \left|t(e_i) - t(e_j)\right| \leq T \\ 0 & else \end{cases}$, $t(e_i)$ is the average time of documents in event e_i. Based on Eq. (17), the events will be threaded when the similarity is greater than the threshold th_4. We set $T = 6$, $\beta = 0.4$ and $th_4 = 0.47$. Finally, the events threaded together are treated as a topic.

6 Experiments

We conduct several experiments in this section to compare the topic detection performance of our approach (SSC) with some representative topic detection and co-clustering approaches, including tag group (TG) method [9], MMG [16], DPM [15], RMC [28] and SS-NMF [31]. TG and DPM are proposed for single media topic detection and MMG is a topic detection framework for cross-media data with different modalities. RMC and SS-NMF are unsupervised and semi-supervised co-clustering methods for heterogeneous data, respectively.

To test the performance of our method for cross-media topic detection, we construct a cross-media dataset YKSN. On the basis of original YKS dataset constructed in [16], we collected more data from the websites so that it contains richer content and much more topics. YKSN contains 8,660 news articles and 6,912 web videos from 1 May to 31 May, 2012. The data are crawled from Youku [18] and Sina [17], respectively. The 430 ground-truth topics of the dataset are manually labeled by 4 assessors. Table 1 shows the composition of the two databases. Although the total number of videos in YKSN is larger than that in YKS, only 13% of videos are topic related, which brings greater challenge for topic detection due to the higher proportion of noise. Considering most of the data are in Chinese, we use a natural language processing (NLP) tool [35] to parse the content first and then filter out the stop words. For visual features of the videos, the NDKs are extracted by the method in [36].

Since Baidu is a major search engine in China, the daily hot search queries [37] recorded by it are used in this experiment. It provides top-20 hottest queries everyday and we collect these queries from the corresponding period of the dataset. So a total of 620 queries are obtained. For each query, we crawl the top 30 Baidu search results as its BI. Both the queries and their BI are also parsed by the NLP tool [35] before using.

To evaluate the results quantitatively, we use Precision, Recall and F-Measure defined by Eq. (19, 20) to assess the performance, which is the same as in [9,16]. The average Precision, Recall and F-Measure of the top-20 topics ranked by F-Measure are used for evaluation. Besides, the number of detected topics [16] is also an important measure of the performance if a cluster whose F-measure is bigger than 0.5, then it is treated as a detected topic.

$$P(S_D, S_G) = \frac{|S_D \cap S_G|}{|S_D|} \qquad R(S_D, S_G) = \frac{|S_D \cap S_G|}{|S_G|} \quad (19)$$

$$F(S_D, S_G) = \frac{2 \times P(S_D, S_G) \times R(S_D, S_G)}{P(S_D, S_G) + R(S_D, S_G)} \quad (20)$$

where S_D is the dataset of a cluster, S_G is the dataset of the best-matched ground-truth topic.

We found that the duration of an event is usually 1–2 days in our dataset. The time slot should be larger than that value, but not too large which will introduce more noise. So the time slot is set to be a time span of 3 days and the 1-month dataset is divided into 11 time slots. The cross-media data are first divided by time, and then the event detection is conducted in each time slot.

6.1 Topic detection for video data of YKSN

We first analyze the performance of our approach (SSC) in detecting topics for the video data. We compare with TG, MMG, DPM, SS-NMF and RMC on the video data of YKSN. Besides, since the video data contain both textual and visual features, detecting topics from them could also verify the effectiveness of our method in integrating multi-modal features. So we conduct another experiment (SSC-V) which does not use the visual relational matrix by removing $\left\|\mathbf{R}^{(c2)} - \mathbf{F}\mathbf{S}^{(2)}\mathbf{G}^{(2)}\right\|_F^2$ in the objective function (6). It should be noted that when detecting topics from single-modality

Table 1 The comparison between YKS and YKSN

	News articles			Web videos		
	Total number	Topic-related number	Ratio (%)	Total number	Topic-related number	Ratio (%)
YKS	7,325	4,373	59.7	2,131	543	25.5
YKSN	8,660	4,481	51.7	6,912	898	13.0

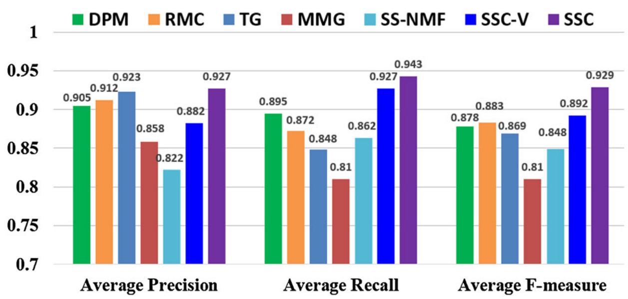

Fig. 3 The average performances of top-20 detected topics on video data of YKSN

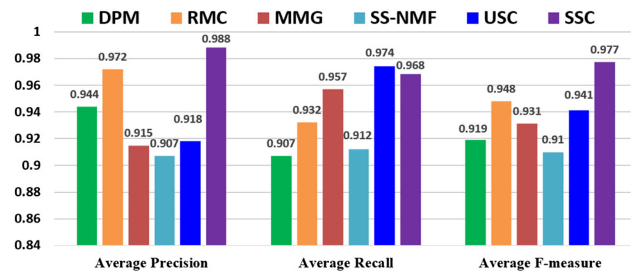

Fig. 4 The average performances of top-20 detected topics on total data of YKSN

data, the formulation of SSC-V is similar to the method proposed in [29] when only the constraints on document dimension is used.

For all approaches, the number of sample or feature clusters is set to the 1/4 of the total number of data in each time slot. The other parameters for each approach are: for SSC-V and SSC: $\alpha = 100$, $th_2 = 0.19$ and $th_3 = 0.49$; for TG: $\theta = 0.2$, $\beta = 0.05$ and $\eta = 0.43$; for MMG: the parameters are set to the same as they are in [16]; for DPM, the offline topic detection model is used and the bursty score threshold is set to be 0.1; for RMC, we construct eleven manifolds containing one binary weighting graph, nine Gaussian graphs and one cosine similarity graph, and the parameters are set to the same as they are in [28]; for SS-NMF, we use the identical prior knowledge as SSC.

The results are shown in Fig. 3 and through F-measure we can see that the two semi-supervised methods SSC and SSC-V outperform the unsupervised methods. By integrating visual information, SSC achieves better results than SSC-V which proves that our approach can effectively cluster the video data using the information contained in different modalities. SS-NMF projects the original data into a new space by the learned distance metric and then co-clusters the data in this new space. However, as the labeled data cover only a small portion of topics, the learned space is only discriminate for those labeled topics. For the other topics, the new space maybe not as appropriate as the original space, leading to unsatisfying co-clustering performance. The rest four methods are all unsupervised ones. RMC achieves best performance among them which demonstrates that the co-clustering performance could be improved by multimanifolds embedding. But its performance is not as good as that of SSC-V and SSC because it cannot make use of the prior knowledge. Although DPM is a good topic detection method for single media such as news articles, its performance is limited because it only use the single-modality feature. TG detects topics by mining the dense tag groups and its precision is high, but it misses some data because the tag groups do not contain enough tags to represent that topic for matching with data. MMG, however, may be affected by the default parameters and the performance is not so good.

6.2 Topic detection for cross-media data of YKSN

We compare the performance of our method with topic detection and co-clustering methods on the total data of YKSN, including SS-NMF, DPM, RMC, and MMG. Besides, to better understand the help of the guidance information for cross-media topic detection, we also conduct experiments in an unsupervised way (USC) which does not use the guidance information. The parameters are: for USC, we set $\alpha = 0$ to make the guidance information useless for detecting; for the other methods, the parameters are determined as same as they are in Sect. 6.1. The average performances of top-20 detected topics are shown in Fig. 4.

From Fig. 4 we can see that SSC performs best among all the methods, which demonstrates the effectiveness of our approach for cross-media topic detection. Compared with USC, SSC greatly improves the performance using the guidance information. This indicates that SSC could effectively integrate the guidance information into clustering and improve the quality of clustering. RMC achieves the highest performance among all the unsupervised methods as they are in video topic detection (Sect. 6.1). For the same reason in video topic detection, the performance of DPM is not satisfactory. Table 2 provides the comparison results of the number of detected topics with these approaches. MMG only detects the dense subgraph which greatly limits the number of detected topics, so the number of the detected topics is the smallest. Using the guidance information, SSC can easily detect the topics that are prior labeled, and obtains more topics than USC. It is the same for SS-NMF. Although its F-measure is the lowest which indicates it does not have good co-clustering results on the cross-media data, the prior labeled topics are easily detected making it detect more topics than DPM and MMG.

Table 2 The comparison of the number of detected topics with different methods

	DPM	RMC	MMG	SS-NMF	USC	SSC
Number of detected topics	179	196	141	187	198	230

Table 3 The comparison of the number of video-related topics in NV and NT

	NV	NT
Number of video-related topics	63	78

We further analyze whether our NMF-based co-clustering approach (without the guidance) could make the information of cross-media data from different sources complement with each other. If a detected topic contains at least one video that belong to the topic, then this topic is video related. We first conduct USC only on the video data of YKSN and count the number of detected topics (NV) which are all video related. Then, the number of video-related topics is also counted from the detection results on total data of YKSN (NT), that is, the number of video-related topics from all the topics detected by USC (Table 2). Table 3 shows the comparison results. We can see that more video-related topics are detected in total data (NT) than in video data only (NV). Since videos are lack of textual information, it is difficult to directly cluster on these incomplete data. However, after adding some news articles, the related videos will be clustered together because the added textual information bridges them. This result shows that our approach can make the information of multi-modal data from different sources complement with each other.

6.3 Parameter analysis

There are some important parameters in SSC and we analyze the influences of these parameters on the performance of SSC in this section. α controls the enforcement degree of the prior knowledge. Figure 5 shows the influence of α on the average Precision, Recall and F-measure of top-20 topics detected by SSC on YKSN. We can see that when α is small, the constraints in SSC will disturb the clustering in certain degree. When α is bigger, the constraints start to guide the clustering and the performance becomes better. However, when $\alpha$is larger than 100, the performance begins to degrade. So we choose $\alpha = 100$ in our experiments.

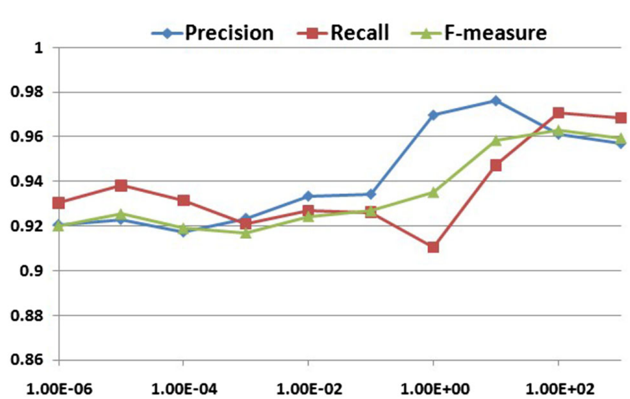

Fig. 5 The influence of α on the average performance of SSC on YKSN

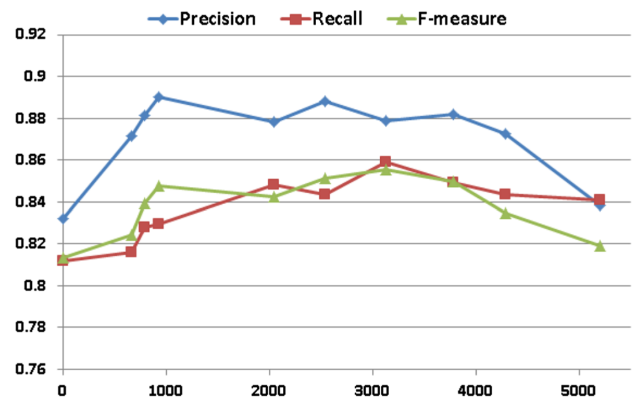

Fig. 6 The influence of different th_2 and th_3 on the average performance of SSC on YKSN

When the queries match with the data, th_2 and th_3 are the parameters that control the threshold of matching so that they can control the number and accuracy of the guidance information. The higher of these thresholds, the accuracy of matched data will be increased but the number is reduced. We now analyze how the different th_2 and th_3 influence the performance of SSC on YKSN. Table 4 shows the number of the matched data (guidance information) under different th_2 and th_3.

To analyze how the guidance information influences the total performance of topic detection, we use the average detection performance top-80 detected topics, but not top-20. Figure 6 shows the average performance of top-80 detected topics by SSC with the values of th_2 and th_3 shown in Table 4. For better illustration, we use the number of matched data as the horizontal axis. From Fig. 6 we can see that when the values of th_2 and $th_3$are high, the data matched with query, which is used for guidance, is accurate, but its quantity is small. When the values of $th_2$and th_3become lower in a certain extent, we can obtain more guidance information while ensuring the accuracy. So we can see that the performance is becoming better and better. However, when the value of th_2 and th_3 continues to decrease, the number of inaccurate guidance increases which influences the performance seriously. So the proper values of $th_2$and th_3 should be neither too high nor too low. The total performance achieves best at $th_2 = 0.19$, $th_3 = 0.49$ which provides both accurate and rich guidance information.

7 Conclusion

In this paper, we propose a novel cross-media topic detection approach which is based on semi-supervised co-clustering. Our approach can detect topics from cross-media data with different modalities from multiple sources. To mining topics contained in the data sufficiently, we also use some side infor-

Table 4 The different values of th_2 and th_3 and the corresponding number of the matched data

th_2	0.46	0.43	0.4	0.25	0.25	0.19	0.13	0.13	0.1
th_3	0.86	0.83	0.8	0.65	0.55	0.49	0.53	0.43	0.4
Number	670	790	928	2,045	2,542	3,135	3,786	4,291	5,202

mation to guide the detection. Experiments demonstrate the superior performance of our approach for cross-media topic detection. In our future work, we will find some new auxiliary information to help topic detection such as hyperlinks between the cross-media data. Besides, we will also improve our approach by selecting the parameters automatically and so on.

Acknowledgments This work was supported in part by National Basic Research Program of China (973 Program): 2012CB316400, in part by National Natural Science Foundation of China: 61025011, 61332016, 61202322, 61202234 and 61303153, by Municipal Natural Science Foundation of Beijing: 4132010 and KZ201310005006, and by China Postdoctoral Science Foundation: 2012M520436.

Appendix

The proof of the correctness and convergence of Algorithm 1 is given in this section.

Correctness of the algorithm

We first prove the correctness of Algorithm1 and have the following theorem.

Theorem 1 *If the update rule of* $\mathbf{F}$, $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$ *and* $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$ *in Algorithm 1 converges, then the final solution satisfies the KKT condition.*

Proof Following the standard theory of constrained optimization, we introduce the Lagrangian multipliers λ_f, λ_{S^i} and λ_{G^i} and minimize the Lagrangian function

$$L(\mathbf{F},\ \mathbf{S}^{(i)},\ \mathbf{G}^{(i)},\ \lambda_f,\ \lambda_{S^1},\ldots,\lambda_{S^h},\ \lambda_{G^1},\ldots,\lambda_{G^h})$$
$$= \sum_{i=1}^{h} \left\| \mathbf{R}^{(ci)} - \mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)} \right\|_F^2 + \alpha Tr[(\mathbf{F}-\mathbf{Y})^T\mathbf{U}(\mathbf{F}-\mathbf{Y})]$$
$$- Tr(\lambda_f \mathbf{F}) - \sum_{i=1}^{h}\left[Tr(\lambda_{S^i}\mathbf{S}^{(i)}) + Tr(\lambda_{G^i}\mathbf{G}^{(i)})\right] \quad (21)$$

The zero gradient condition gives

$$\frac{\partial L}{\partial \mathbf{F}} = -2\sum_{i=1}^{h}\left(\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{Y}\right)$$
$$+ 2\left(\sum_{i=1}^{h}\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{F}\right) - \lambda_f = 0,$$

$$\frac{\partial L}{\partial \mathbf{S}^{(i)}} = -2\mathbf{F}^T\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T} + 2\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T} - \lambda_{S^i} = 0,$$
$$\frac{\partial L}{\partial \mathbf{G}^{(i)}} = -2\mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{R}^{(ci)} + 2\mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)} - \lambda_{G^i} = 0.$$

From the complementary slackness condition, we obtain

$$\lambda_f \mathbf{F}_{pq} = \left[-\sum_{i=1}^{h}(\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{Y}) + \sum_{i=1}^{h}\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T} + \alpha\mathbf{U}\mathbf{F}\right]_{pq}\mathbf{F}_{pq} = 0 \quad (22)$$

$$\lambda_{S^i}\mathbf{S}^{(i)}_{pq} = \left(-\mathbf{F}^T\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T} + \mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\right)_{pq} \times \mathbf{S}^{(i)}_{pq} = 0 \quad (23)$$

$$\lambda_{G^i}\mathbf{G}^{(i)}_{pq} = \left(-\mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{R}^{(ci)} + \mathbf{S}^{(i)^T}\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\right)_{pq} \times \mathbf{G}^{(i)}_{pq} = 0 \quad (24)$$

Equations (22), (23) and (24) are the fixed point equations that the solution must satisfy at convergence. It is obvious that the limiting solutions of the update rules of (7), (8) and (9) satisfy these fixed point equations. For example, we have $\mathbf{S}^{(i)^\infty} = \mathbf{S}^{(i)^{t+1}} = \mathbf{S}^{(i)^t}$ at convergence, so the update rule of (7) becomes

$$\mathbf{S}^{(i)}_{pq} = \mathbf{S}^{(i)}_{pq}\frac{(\mathbf{F}^T\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T})_{pq}}{(\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T})_{pq}}$$

then we obtain

$$(-\mathbf{F}^T\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T} + \mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T})_{pq}\mathbf{S}^{(i)}_{pq} = 0$$

which is identical to Eq. (23). Similarly, we can see the update rule of (8) and (9) is identical to Eqs. (24) and (22), respectively. So the solution satisfies the KKT condition at convergence. The proof is completed. □

Convergence of the algorithm

We now prove the convergence of the algorithm. First, we assume that $\{\mathbf{S}^{(i)}\}_{1\le i\le h}$ and $\{\mathbf{G}^{(i)}\}_{1\le i\le h}$ are fixed matrices, then the objective function (6) can be written as

$$\min_{\mathbf{F}\geq 0} J_1(\mathbf{F}) = Tr\left[-2\left(\sum_{i=1}^{h}\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}+\alpha\mathbf{UY}\right)\mathbf{F}^T\right] + \sum_{i=1}^{h} Tr(\mathbf{F}^T\mathbf{F}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}) + \alpha Tr(\mathbf{F}^T\mathbf{UF})$$

where we ignore the constant $\mathbf{R}^{(ci)^T}\mathbf{R}^{(ci)}$ and $\mathbf{Y}^T\mathbf{UY}$.

Theorem 2 *The objective function $J_1(\mathbf{F})$ is monotonically nonincreasing under the update rule (9).*

To prove Theorem 2, we will make use of auxiliary function approach in [37]. We first introduce the definition of auxiliary function.

Definition 1 $Z(H, \widetilde{H})$ is an auxiliary function for $X(H)$ if the conditions

$$Z(H, \widetilde{H}) \geq X(H), \qquad Z(H, H) = X(H) \tag{25}$$

are satisfied.

Lemma 1 *If Z is an auxiliary, then X is nonincreasing under the update*

$$H^{t+1} = \arg\min_{H} Z(H, H^t) \tag{26}$$

Proof $X(H^t) = Z(H^t, H^t) \geq Z(H^{t+1}, H^t) \geq X(H^{t+1})$. The proof is completed. □

So the key is to find an appropriate auxiliary function $Z(\mathbf{F}, \widetilde{\mathbf{F}})$ of $J_1(\mathbf{F})$.

Proof of Theorem 2 Now we show that

$$Z(\mathbf{F}^{t+1}, \mathbf{F}^{t}) = -\sum_{pq} 2\left(\sum_{i=1}^{h}\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}+\alpha\mathbf{UY}\right)_{pq}\mathbf{F}^{t+1}_{pq} + \sum_{i=1}^{h}\sum_{pq}\frac{(\mathbf{F}^{t}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T})_{pq}(\mathbf{F}^{t+1}_{pq})^2}{\mathbf{F}^{t}_{pq}} + \sum_{pq}\frac{(\mathbf{UF}^{t})_{pq}(\mathbf{F}^{t+1}_{pq})^2}{\mathbf{F}^{t}_{pq}} \tag{27}$$

is an auxiliary function of $J_1(\mathbf{F}^{t+1})$. First, it is obvious that the equality $J_1(\mathbf{F}^t) = Z(\mathbf{F}^t, \mathbf{F}^t)$ holds. Then the inequality $Z(\mathbf{F}^{t+1}, \mathbf{F}^t) \geq J_1(\mathbf{F}^{t+1})$ holds because: the first term in $J_1(\mathbf{F}^{t+1})$ and $Z(\mathbf{F}^{t+1}, \mathbf{F}^t)$ is equal, and the second and third term in $Z(\mathbf{F}^{t+1}, \mathbf{F}^t)$ is always bigger than in $J_1(\mathbf{F}^{t+1})$ [27].

According to Eq. (26), we find the minimum of $Z(\mathbf{F}^{t+1},\mathbf{F}^t)$ fixing $\mathbf{F}^t$. The minimum is obtained by

$$\frac{\partial Z(\mathbf{F}^{t+1}, \mathbf{F}^{t})}{\partial \mathbf{F}^{t+1}_{pq}} = -2\left(\sum_{i=1}^{h}\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}+\alpha\mathbf{UY}\right)_{pq} + 2\sum_{i=1}^{h}\frac{(\mathbf{F}^{t}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T})_{pq}\mathbf{F}^{t+1}_{pq}}{\mathbf{F}^{t}_{pq}} + 2\frac{(\mathbf{UF}^{t})_{pq}\mathbf{F}^{t+1}_{pq}}{\mathbf{F}^{t}_{pq}} = 0$$

then we can get the update rule (9) by solving F^{t+1}:

$$\mathrm{F}^{t+1}_{pq} = \mathbf{F}^{t}_{pq}\frac{(\sum_{i=1}^{h}\mathbf{R}^{(ci)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}+\alpha\mathbf{UY})_{pq}}{(\sum_{i=1}^{h}\mathbf{F}^{t}\mathbf{S}^{(i)}\mathbf{G}^{(i)}\mathbf{G}^{(i)^T}\mathbf{S}^{(i)^T}+\alpha\mathbf{UF}^{t})_{pq}}$$

So under this update rule, function $J_1(\mathbf{F})$ is monotonically nonincreasing. The proof is completed. □

So far we assume $\{\mathbf{S}^{(i)}\}_{1\leq i\leq h}$ and $\{\mathbf{G}^{(i)}\}_{1\leq i\leq h}$ are fixed matrices. Alternatively, we can also fix $\mathbf{F}$ and $\{\mathbf{G}^{(i)}\}_{1\leq i\leq h}$ or $\mathbf{F}$ and $\{\mathbf{S}^{(i)}\}_{1\leq i\leq h}$, and the proof for convergence of the update rules of (7) and (8) is similar.

References

1. LDC (1999) TDT3 evaluation specification version 2.7
2. He Q, Chang K, Lim EP (2007) Analyzing feature trajectories for event detection. In: ACM SIGIR conference
3. Mei Q, Zhai C (2005) Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In: ACM SIGKDD international conference on knowledge discovery and data mining
4. Allan J, Carbonell J, Doddington G, Yamron J, Yang Y (1998) Topic detection and tracking pilot study: final report. In: Proceedings of the DARPA broadcast news transcription and understanding, workshop, pp 194–218
5. Fiscus JG, Doddington GR (2002) Topic detection and tracking evaluation overview. In: Topic detection and tracking: event-based information organization
6. Liu L, Sun L, Rui Y, Shi Y, Yang S (2008) Web video topic discovery and tracking via bipartite graph reinforcement model. In: International world wide web conference
7. Cao J, Ngo CW, Zhang YD, Li JT (2011) Tracking web video topics: discovery, visualization and monitoring. IEEE Trans Circuits Syst Video Technol 21(12):1835–1846
8. Sun AX, Hu MS (2011) Query-guided event detection from news and blog streams. IEEE Trans Syst Man Cyber Part A Syst Hum 41(5):834–839
9. Chen TL, Liu CX, Huang QM (2012) An effective multi-clue fusion approach for web video topic detection. In: ACM multimedia
10. Zhao Q, Liu T-Y, Bhowmick SS, Ma W-Y (2006) Event detection from evolution of click-through data. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 484–493
11. Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conferece research and development in information retrieval, pp 267–273
12. Yang Y, Carbonell J, Brown R, Pierce T, Archibald BT, Liu X (1999) Learning approaches for detecting and tracking news events. IEEE Intell Syst 14(4):32–43

13. Saha A, Sindhwani V (2012) Learning evolving and emerging topics in social media: a dynamic NMF approach with temporal regularization in WSDM, pp 693–702
14. http://projects.ldc.upenn.edu/TDT/. Accessed June 2013
15. He Q, Chang K, Lim E, Banerjee A (2010) Keep it simple with time: a re-examination of probabilistic topic detection models. IEEE Trans Pattern Anal Mach Intell 32(10):1795–1808
16. Zhang YY, Li GR, Chu LY, Wang SH, Zhang WG, Huang QM (2013) Cross-media topic detection: a multi-modality fusion framework. In: ICME
17. http://news.sina.com.cn/. Accessed June 2013
18. http://www.youku.com/. Accessed June 2013
19. Yang Y, Pierce T, Carbonell J (1998) A study on retrospective and online event detection. In: Proceedings of the 21st annual international ACM SIGIR conference on reserach and development in information retrieval, Melbourne, Australia, pp 28–36
20. Wang CH, Zhang M, Ma SP, Ru LY (2008) Automatic online news issue construction in web environment. In: International world wide web conference
21. Banerjee A, Dhillon IS, Ghosh J, Merugu S, Modha DS (2004) A generalized maximum entropy approach to bregman co-clustering and matrix approximation. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 509–514
22. Dhillon IS et al (2001) Co-clustering documents and words using bipartite spectral graph partitioning. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 269–274
23. Dhillon IS, Mallela S, Modha DS (2003) Information-theoretical coclustering. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 89–98
24. Shao J, Ma S, Lu WM, Zhuang YT (2012) A unified framework for web video topic discovery and visualization. Pattern Recogn Lett 33(4):410–419
25. Gao B, Liu T-Y, Feng G, Qin T, Cheng Q-S, Ma W-Y (2005) Hierarchical taxonomy preparation for text categorization using consistent bipartite spectral graph copartitioning. IEEE Trans Knowl Data Eng 17(9):1263–1273
26. Long B, Wu X, Zhang Z, Yu PS (2006) Spectral clustering for multi-type relational data. In: International conference on machine learning, pp 585–592
27. Ding C, Li T, Peng W, Park H (2006) Orthogonal nonnegative matrix tri-factorizations for clustering. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 126–135
28. Li P, Bu J, Chen C, He Z, Cai D (2013) Relational multimanifold coclustering. IEEE Trans Cybern 43(6):1871–1881
29. Li T, Zhang Y, Sindhwani V (2009) A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In: Proceedings of the joint conference of the 47th annual meeting of the association of computational linguistics, Suntec, Singapore, pp 244–252
30. Wang F, Li T, Zhang C (2008) Semi-supervised clustering via matrix factorization. In: Proceedings of the SIAM international conference on data mining (SDM), Atlanta, GA
31. Chen Y, Wang L, Dong M (2009) Non-negative matrix factorization for semi-supervised heterogeneous data co-clustering. IEEE Trans Knowl Data Eng 22(10):1459–1474
32. Xue Z, Jiang S, Li G, Huang Q, Zhang W (2013) Cross-media topic detection associated with hot search queries. In: Proceedings of fifth international conference on internet multimedia computing and service, pp 403–406
33. Lee D, Seung H (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791
34. Li T, Ding C, Jordan MI (2006) Solving consensus and semi-supervised clustering problems using non-negative matrix factorization. In: Proceedings of ICDM, pp 362–371
35. http://ictclas.org/ictclas_download.aspx. Accessed June 2013
36. Xie L, Natsev A, Kender JR, Hill ML, Smith JR (2011) Visual memes in social media: tracking realworld news in youtube videos. In: ACM multimedia pp. 53–62
37. http://hot.news.baidu.com/. Accessed June 2013