Multimedia News Summarization in Search

ZECHAO LI, JINHUI TANG, and XUEMING WANG, School of Computer Science and Engineering, Nanjing University of Science and Technology
JING LIU and HANQING LU, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

It is a necessary but challenging task to relieve users from the proliferative news information and allow them to quickly and comprehensively master the information of the whats and hows that are happening in the world every day. In this article, we develop a novel approach of multimedia news summarization for searching results on the Internet, which uncovers the underlying topics among query-related news information and threads the news events within each topic to generate a query-related brief overview. First, the hierarchical latent Dirichlet allocation (hLDA) model is introduced to discover the hierarchical topic structure from query-related news documents, and a new approach based on the weighted aggregation and max pooling is proposed to identify one representative news article for each topic. One representative image is also selected to visualize each topic as a complement to the text information. Given the representative documents selected for each topic, a time-bias maximum spanning tree (MST) algorithm is proposed to thread them into a coherent and compact summary of their parent topic. Finally, we design a friendly interface to present users with the hierarchical summarization of their required news information. Extensive experiments conducted on a large-scale news dataset collected from multiple news Web sites demonstrate the encouraging performance of the proposed solution for news summarization in news retrieval.

Categories and Subject Descriptors: H.3.3 [**Information Search and Retrieval**]: Search Process

General Terms: Design, Algorithms, Performance, Human Factors

Additional Key Words and Phrases: News summarization, topic structure, multimodal, hierarchical latent Dirichlet allocation, maximum spanning tree

ACM Reference Format:
Zechao Li, Jinhui Tang, Xueming Wang, Jing Liu, and Hanqing Lu. 2016. Multimedia news summarization in search. ACM Trans. Intell. Syst. Technol. 7, 3, Article 33 (February 2016), 20 pages.
DOI: http://dx.doi.org/10.1145/2822907

1. INTRODUCTION

With the proliferation of news articles on the Internet, an increasing number of people access news information online rather than from newspapers.[1] According to the report

[1] http://mashable.com/2011/03/15/online-versus-newspaper-news/.

This work was partially supported by the 973 Program (project 2014CB347600) and the National Natural Science Foundation of China (grants 61522203 and 61402228).
Authors' addresses: Z. Li, J. Tang (corresponding author), and X. Wang, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; emails: {zechao.li, jinhuitang}@njust.edu.cn, wangxueming516@gmail.com; J. Liu and H. Lu, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; emails: {jliu, luhq}@nlpr.ia.ac.cn.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2016 ACM 2157-6904/2016/02-ART33 $15.00
DOI: http://dx.doi.org/10.1145/2822907

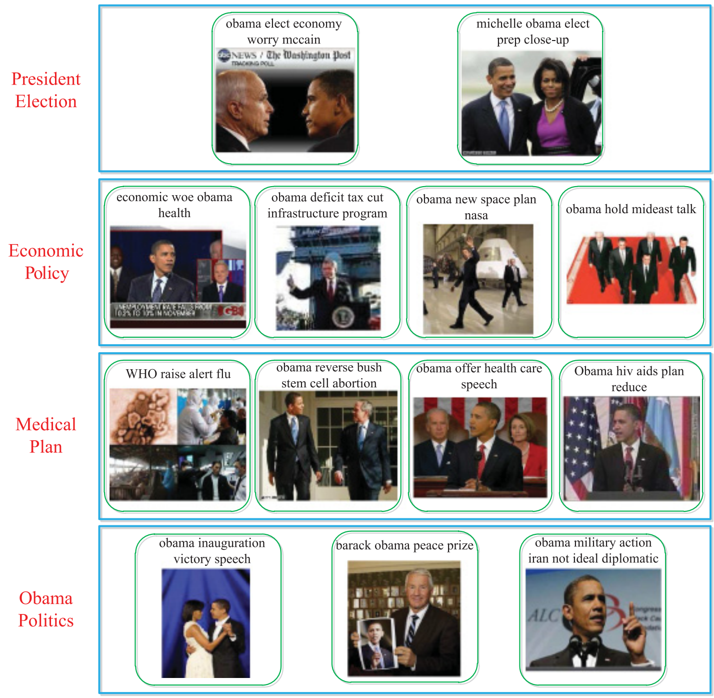

Fig. 1. The illustration of the topic structure discovered in the news data related to the query "Barack Obama."

of the Pew Research Center's Project for Excellence in Journalism,[2] 58.2% of Americans browse news online, overtaking newspaper readership (about 35%) by the end of 2011 [Li et al. 2013]. Unfortunately, it takes much time for users to find the news in which they are interested from such huge volumes of information. As a consequence, it urges the necessity to uncover compact presentations of possibly noisy and redundant news.

Today, several news search engines, such as Google News[3] and Yahoo! News,[4] constantly collect news documents from every corner of the world and attempt to provide updated news to users. However, these systems only return a list of related news articles for each query and do not provide a summary view to users, which prevents users from accessing the desired information quickly and comprehensively. Considering different concerns in various news articles and varying stages during news development, an ideal system for news search can present an informative and complete news story, including news origin, developing process, and its result about certain news subject. Here we present an illustrative example in Figure 1, which is the topic structure of related news documents to the query "Barack Obama." It is observed that there exist four topics: "President Election," "Economic Policy," "Medical Plan," and "Obama Politics." Therefore, how to search and summarize such news information is necessary and important for news search, which is the focus of this article.

On the other hand, most news Web pages contain multimodal information, such as image, video, and text. Cognitive psychology research indicates that human cognition is a process of "cross-media" information interaction—that is, cognition is the result stimulated by visual, auditory, and other sensory information simultaneously

[2]http://www.journalism.org/.
[3]http://news.google.com/.
[4]http://news.yahoo.com/.

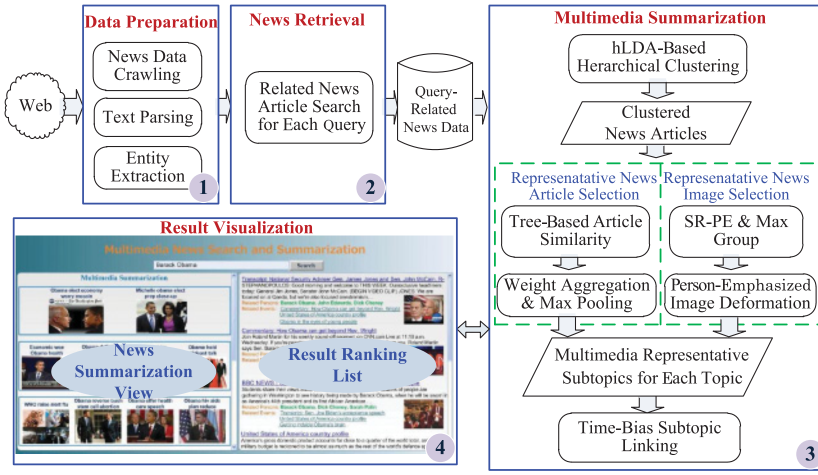

Fig. 2. The framework of our system.

[McGurk and MacDonald 1976; Cohen and Massaro 1990; Calvert 2001; Ghazanfar and Schroeder 2006]. There is a saying that "a picture is worth a thousand words." Thus, visual content should be exploited to complement the textual content of news to provide vivid and comprehensive information to users. Which images are selected and how to organize them in a compact form deserve our attention in this work.

As mentioned previously, it is desirable to generate a vivid and informative news story relevant to a specific query, which can allow users to quickly get a brief and comprehensive overview about their desired information. Toward this end, this work develops a multimedia news summarization framework focusing on the query-related news articles and their associated multimedia information. The proposed framework is illustrated in Figure 2. For each query, the related news articles can be found by many approaches for news retrieval. This work utilizes an existing correlation mining approach named MPMF [Li et al. 2010]. In the following, the returned corpus of related news data is summarized into a hierarchical presentation. First, we adopt hierarchical latent Dirichlet allocation (hLDA) [Blei et al. 2010] to uncover the hierarchical topic structure among those query-related news documents and propose an approach based on weighted aggregation and max pooling to identify one typical news event for each topic. One representative image is also selected for each topic to complement the textual contents, and the chosen images are deformed to emphasize newspersons in the news images. Next, a time-bias maximum spanning tree (MST) approach is proposed to link subtopics of one topic. Finally, a vivid and informative user interface is designed to display the multimodel news summary view as well as a ranking list of related news documents for each query. Experiments on a large-scale news dataset collected from four news Web sites (i.e., ABC News, BBC News, CNN News, and Google News) are used to evaluate the performance of the proposed solution.

Our main contributions are summarized as follows:

(1) We propose to search and summarize news in a multimodal form, which enables users to browse the related news with multimodal information, as well as access the desired information quickly and comprehensively.
(2) hLDA is adopted to discover the hierarchical topic structure from the query-related news articles, and an approach based on the weighted aggregation and max pooling to identify the typical news article for each topic is proposed.

(3) A time-bias MST method is developed to thread the subtopics within one topic to give a news summary on each topic in terms of temporal and spatial development.

The reminder of this article is organized as follows. Section 2 reviews related work. We introduce hLDA to discover the latent hierarchical topic structure in Section 3. Section 4 presents the details of multimedia news summarization. Section 5 shows the interface of our system. The utilized dataset is reported in Section 6, whereas experimental results and discussions are presented in Section 7. We conclude the article with future work in Section 8.

2. PREVIOUS WORK

Extensive research effort has been devoted to news categorization, automatic topic detection, summarization, and retrieval. In this article, we briefly review the related work on topic detection, summarization, and retrieval, respectively.

The hierarchical topic detection (HTD) task was proposed and evaluated in 2004, in which stories are classified in a hierarchy of topic clusters. The traditional topic detection task was replaced by a hierarchical structure [Trieschnigg and Kraaij 2004], and a hierarchical agglomerative clustering (HAC) algorithm achieved the highest performance in the evaluation [Trieschnigg and Kraaij 2005]. A notable technique under the bag-of-words assumption for document classification is latent Dirichlet allocation (LDA) [Blei et al. 2003]. Probabilistic latent semantic analysis [Hofmann 1999] is the predecessor of LDA. A hierarchical topic algorithm was proposed to address the issue of variable topics, named hierarchical latent Dirichlet allocation (hLDA) [Blei et al. 2010]. Many works based on LDA for document representations are proposed [Wei and Croft 2006; Cao et al. 2007; Phan et al. 2011; Lau et al. 2013].

Given any text document, automatic summarization attempts to abstract important information in text and present it in a condensed form sensitive to the user's or applications' needs [Mani and Maybury 1999]. Methods for text summarization can be generally classified into two categories: single-document summarization [Paice 1990] and multiple-document summarization [Mani and Bloedorn 1997]. Many approaches have been proposed for clustering and summarizing multiple documents [Radev 2000; Radev and McKeown 1999], which exploit meaningful relations among documents. It is a natural application of multidocument summarization algorithms to provide summarization for Web pages. Several systems [Neto et al. 2000; Radev and Fan 2000] have been developed to perform Web page clustering and generic summarization based on relevant results from a search engine. Techniques, particularly on news articles, have also been proposed (e.g., Radev et al. [2001], McKeown et al. [2002], and Radev et al. [2005]). NewsBlaster [McKeown et al. 2002] provides summaries that give a representation of a cluster. NewsInEssence [Radev et al. 2005] generates online document clusters and summaries for user-specified requirement. Other work generates structured representation for news events as a timeline summarization [Yan et al. 2011; Tuan et al. 2011; Tran 2013] to provide better information to users than traditional styles. Hierarchical summarization [Celikyilmaz and Hakkani-Yur 2010; Christensen et al. 2014] has been studied to generate a summary of multiple documents based on sentences. However, the proposed summarization scheme uncovers hierarchical topics and represents topics based on documents and images.

Some popular news services, such as Google News, present clusters of related articles, allowing readers to easily find all stories on a given topic. Google News provides the summarization, such as topic cluster, representative images, news articles, and news videos. However, the topics in Google News are predefined according to the location or categories such as sports, economy, and politics. The latest articles, images, and videos are simply selected. As claimed in Subašić and Berendt [2010], some alternative ways of tracking and browsing news collections have recently been developed. Google News Timeline[5] provides a preset time period overview of news using a timeline interface. Another Google system, named Fast Flip,[6] provides an interface for browsing news resembling hard-copy newspaper reading. The Yahoo! Correlator[7] associates a search item with all of its related "events." EMM NewsBrief[8] is a news summarization service. In Shahaf et al. [2012], a system named "metro maps" is designed to create structured summaries of information. However, our system automatically generates the summarization according to the search results and selects the representative images and news articles, which may not be the latest news.

Current Web pages always contain multimodal information. In Jiao et al. [2012], Web pages are visually summarized by selecting images from the internal and external images. The work in Li et al. [2010] jointly exploits textual news content and news image information to estimate inter- and intracorrelations among newspersons and news events. Multicorrelation probabilistic matrix factorization (MPMF) was proposed to reconstruct person-event, person, and event correlation matrices simultaneously by the shared latent person and event matrices. Based on the reconstructed correlations, the work can provide person-centric news search results. Li et al. [2011, 2013] focused on the relations between news documents and news geolocations, and news enrichment with Web images. The latent relations between news documents and news geolocation are uncovered by the proposed matrix factorization methods. Furthermore, they designed approaches to extract queries from news documents for image search and then selected suitable images to visualize news content. However, the preceding methods do not summarize the information and cannot provide a condensed view to readers. In addition, they ignore the topics hidden in the events.

Different from the preceding methods, in this article we consider the multimodal information of the current news Web pages to search and summarize news information. We present a vivid and informative interface to users that contains a multimodal news summary view and a ranking list of relevant news items.

3. HIERARCHICAL TOPIC STRUCTURE

Given the corpus of related news articles to one query, the latent topic structure is first discovered since the related news articles always cover several aspects. Most of the topic models, such as LDA [Blei et al. 2003], should be specified in a fixed number of topics. However, the number of topics hidden in the news data is usually uncertain. To address the issue of variable topics, the hLDA model [Blei et al. 2010], as a generalization of LDA, is adopted to discover multiple topics along a hierarchical structure in this article.

The hierarchical structure presents a top-down organization of news data, in which the lower node has the more specific (or abstract) interpretation. Therefore, it is appealing to show a compact news summary with such a hierarchically coarse-to-fine manner, which is not possible for a "flat" model.

The hLDA model represents topic distribution with a tree of fixed depth L. Each node is associated with a topic distribution over words, and a topic is a distribution across words. An article is generated by selecting a path from the root to a leaf, repeatedly sampling topics along the path and selecting words from the sampled topics. The tree structure is learned along with the topics using a nested Chinese restaurant process (nCRP) [Blei et al. 2010] method, which assigns probability distributions to an infinitely

[5]http://newstimeline.googlelabs.com.
[6]http://fastflip.googlelabs.com.
[7]http://correlator.sandbox.yahoo.net.
[8]http://emm.newsbrief.eu.

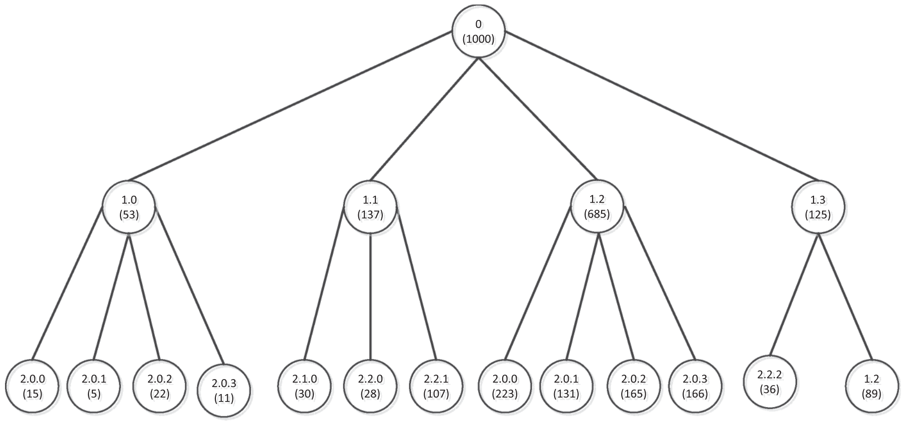

Fig. 3. The hLDA-based topic structure to the query "Barack Obama." The numbers in the brackets present the numbers of news articles that have nodes in their paths.

branching and infinitely deep tree. The basic process is specified as follows. A sequence of n news articles arrive. Similar to the case in which the first customer sits at the first table, the first article takes the initial path, starting with a single branch tree. The n_t-th subsequent article is assigned to a path drawn from the following distribution:

$$p(\text{an existing path } c|n_c, n_t) = \frac{n_c}{\gamma + n_t - 1}, \tag{1}$$

$$p(\text{a new path } c|n_c, n_t) = \frac{\gamma}{\gamma + n_t - 1}, \tag{2}$$

where n_c is the number of previous articles assigned to path c and n_t is the total number of articles observed in the tree. γ is a parameter to control how often an article chooses a path versus creates a new path.

In hLDA, articles in a corpus are drawn from the following generative process:

(1) For each topic $t \in \mathcal{T}$ in the tree,
  (a) Draw a distribution $\beta_t \sim \text{Dirichlet}(\eta)$.
(2) For each article $d \in \mathcal{D} = \{d_1, d_2, \ldots, d_n\}$,
  (a) Draw a path $c_d \sim \text{nCRP}(\gamma)$.
  (b) Draw a distribution over levels in the tree, $\theta_d \sim \text{Dirichlet}(\alpha)$.
  (c) For each word a,
    i. Choose level $t_{d,a}|\theta_d \sim \text{Discrete}(\theta_d)$.
    ii. Choose word $w_{d,a}|\{t_{d,a}, c_d, \beta\} \sim \text{Discrete}(\beta_{c_d}[t_{d,a}])$, which is parameterized by the topic in position $t_{d,a}$ on the path c_d.

Intuitively, the parameters γ and η control the size of the tree. A model with larger γ and smaller η will tend to find a tree with more topics. In the hierarchical topic model, the internal nodes are not the summaries of their children but reflect the shared terminology of the articles assigned to the paths that contain them. We also present an example of the topic structure discovered by hLDA in Figure 3. The corpus of news articles consists of the top 1,000 related news articles to the query "Barack Obama."

4. MULTIMEDIA TOPIC REPRESENTATION

In this section, we will elaborate on how to obtain a multimedia representation for each latent topic based on the hierarchical topic structure.

4.1. Representative News Article Selection

In the hierarchical tree, each article is represented as a path, and each path is shared by many articles. The articles sharing the same path belong to the same topics, and thus they are more similar to each other. Motivated by Chen et al. [2009] and Celikyilmaz and Hakkani-Yur [2010], if the degree of an article is larger than the degree of any other article belonging to the same topic, it can better describe the topic. We adopt the degree as the score of the news article, which is calculated based on the news article similarity. In this work, the news article similarity is calculated by combining the text information and the hierarchical structure, which is defined as

$$\mathbf{S}_{\text{comb}} = \varepsilon \mathbf{S}^{\text{text}} + (1 - \varepsilon)\mathbf{S}^{\text{tree}}. \tag{3}$$

Here, $\mathbf{S}^{\text{text}}$ is the similarity matrix based on text information or mined by other methods, such as MPMF [Li et al. 2010], and $\mathbf{S}^{\text{tree}}$ is the similarity calculated based on the hierarchical tree. ε is a parameter to balance their importance. We propose a new algorithm to calculate the similarly $\mathbf{S}^{\text{tree}}$ based on the hierarchical tree and choose a representative news article for each topic using weighted aggregation and max pooling.

4.1.1. Tree-Based Article Similarity. Gibbs sampling is a particular Markov chain Monte Carlo algorithm to approximate the posterior for hLDA. Let $\mathbf{t}$ denote the level assignment for all words and $\mathbf{c}$ denote the path assignment for all articles conditioned on the observed words $\mathbf{w}$. Given the assignment of words $\mathbf{w}$ to levels $\mathbf{t}$ and assignments of articles to paths $\mathbf{c}$, the expected posterior probability of a word w at a given topic $\mathbf{t} = t$ of a path $\mathbf{c} = c$ is proportional to the number of times that w was generated by topic t:

$$p(w|\mathbf{t}, \mathbf{c}, \mathbf{w}, \eta) \propto \#[\mathbf{t} = t, \mathbf{c} = c, \mathbf{w} = w] + \eta. \tag{4}$$

Similarly, the posterior probability of a particular topic t in a given article d is

$$p(t|\mathbf{t}, \mathbf{c}, \alpha) \propto \#[\mathbf{t} = t, \mathbf{c} = c_d] + \alpha. \tag{5}$$

$\#[\cdot]$ is the count of elements of an array satisfying the given condition. The posterior should be normalized with total counts and their parameters.

$\mathbf{S}^{\text{tree}}(d, g)$ between news articles d and g is measured based on the distribution of words at each topic t on path c and distribution of topic t on path c. Based on Equation (4), we can obtain the distribution $p_{t,d} = p(\mathbf{w}_{t,d}|\mathbf{z}_d = t, c, \mathbf{w} = v_t)$, where $\mathbf{w}_{t,d}$ and $\mathbf{w} = v_t$ are the set of words in d that are generated from topic t and a vocabulary with words generated by the topic t, respectively. Similarly, $p_{t,g} = p(\mathbf{w}_{t,g}|\mathbf{z}_g = t, c, \mathbf{w} = v_t)$ is the probability of words $\mathbf{w}_g$ in g of the same topic t. We adopt the Jensen-Shannon divergence [Lin 1991] to measure the distance between $p_{t,d}$ and $p_{t,g}$:

$$div_{d,g,t} = \frac{1}{2}\left(KL\left(p_{t,d}\middle\|\frac{p_{t,d} + p_{t,g}}{2}\right) + KL\left(p_{t,g}\middle\|\frac{p_{t,d} + p_{t,g}}{2}\right)\right), \tag{6}$$

where $KL(d||g) = \sum_i d_i \log(d_i/g_i)$ is the Kullback-Liebler (KL) divergence. Then the divergence is transformed into a similarity measure [Manning and Schütze 1999]:

$$S_1^{\text{tree}}(d, g, t) = 10^{-div_{d,g,t}}. \tag{7}$$

We introduce the topic-based similarities based on article-topic mixing proportions. We calculate the topic proportion between d and g, represented by $p_{t_d} = p(z_d|t, c, \alpha)$ and $p_{t_g} = p(\mathbf{z}_g|t, c, \alpha)$ via Equation (5). $S_2^{\text{tree}}(d, g)$ between these two distributions is measured based on Equations (6) and (7). Two articles on the same path would have different words and hence have different posterior topic probabilities.

$\mathbf{S}_1^{\text{tree}}$ provides information about the similarity between two articles based on topic-word distribution. Similarly, $\mathbf{S}_2^{\text{tree}}$ provides information on the similarity between the

ALGORITHM 1: Tree-Based Article Similarity

Input: Tree $\mathcal{T}$ from hLDA, Articles: $\mathcal{D} = \{d_1, \ldots, d_n\}$
Output: The article similarity matrix based on the hierarchical tree: $\mathbf{S}^{\text{tree}}$

```
for each article d ∈ D do
    for each article g ∈ D do
        Find their sharing topic set T_{d,g};
        for each topic t ∈ T_{d,g} do
            Find the level l_t;
            Calculate S_1^tree(d, g, t) and S_2^tree(d, g, t);
        end
        Calculate S^tree(d, g) using Equation (8);
    end
end
return S^tree;
```

weights of the topics in each article. They jointly effect the article similarity and then are combined in one measure in all levels and all possible paths:

$$S^{\text{tree}}(d, g) = \frac{1}{|\mathcal{T}_{d,g}|} \sum_{t \in \mathcal{T}_{d,g}} S_1^{\text{tree}}(d, g, t) * S_2^{\text{tree}}(d, g, t) * l_t, \tag{8}$$

where $\mathcal{T}_{d,g}$ denotes the set of topics shared by articles d and g. $|\mathcal{T}|$ is the size of set $\mathcal{T}$, and l_t is the level of topic t. If $\mathcal{T}_{d,g} = \emptyset$, $S^{\text{tree}}(d, g) = 0$. We summarize the algorithm in Algorithm 1.

4.1.2. Weighted Aggregation. We have obtained the similarity graph based on the combined similarity $\mathbf{S}_{\text{comb}}$. Representative news articles for topics are selected from the bottom to top of the hierarchical tree. Let $\mathcal{D}_l$ and $\mathcal{D}_{c,l}$ denote the article set at the l-th level and the article set of the topic $t_{c,l}$ at the l-th level on path c, respectively, where $\mathcal{D}_l = \bigcup_c \mathcal{D}_{c,l}$.

For article $d \in \mathcal{D}_{c,l}$, we assign a score to d to represent its importance in the set $\mathcal{D}_{c,l}$:

$$\text{score}(d) = \sum_{g \in \mathcal{D}_{c,l}} S_l(d, g), \tag{9}$$

where $\mathbf{S}_l$ denotes the similarity matrix at level l, especially $\mathbf{S}_L = \mathbf{S}_{\text{comb}}$. A representative article denoted as $\text{td}_{c,l}$ is chosen to stand for the topic $t_{c,l}$ using max pooling:

$$\text{td}_{c,l} = \arg\max_{d \in \mathcal{D}_{c,l}} \text{score}(d). \tag{10}$$

We employ a sparse interpolation matrix $\mathbf{W}_l$ to build the relations between $\mathcal{D}_l$ and $\mathcal{D}_{l-1}$. $\mathbf{W}_l$ is defined as

$$W_l(d, g) = \begin{cases} 0 & \text{if } d, g \in \mathcal{D}_{l-1} \text{ and } d \neq g \\ 1 & \text{if } d, g \in \mathcal{D}_{l-1} \text{ and } d = g \\ \frac{S_l(d,g)}{\sum_{q \in \mathcal{D}_{l-1}} S_l(d,q)} & \text{if } d \in \mathcal{D}_l - \mathcal{D}_{l-1} \text{ and } g \in \mathcal{D}_{l-1}. \end{cases} \tag{11}$$

Thus, we can build the matrix $\mathbf{S}_{l-1}$ as

$$\mathbf{S}_{l-1} = \mathbf{W}_l^T \mathbf{S}_l \mathbf{W}_l. \tag{12}$$

$\mathbf{S}_{l-1}$ inherits the property of $\mathbf{S}_l$ by the interpolation matrix. This process is called *weighted aggregation*, which employs the property of spectral clustering [Li et al. 2015] to measure the similarity between two nodes [Chen et al. 2009].

ALGORITHM 2: Choosing Representative Articles via Weighted Aggregation and Max Pooling

Input: Tree $\mathcal{T}$ from hLDA, Articles: $\mathcal{D} = \{d_1, \ldots, d_n\}$, Article similarity matrix $\mathbf{S}_L = \mathbf{S}_{\text{comb}}$
Output: The representative article for topics $\mathcal{TD}$
for *level* $l \leftarrow L, \ldots, 2$ **do**
  for *each topic t corresponding path c at level l* **do**
    Score each article in $\mathcal{D}_{c,l}$ using Equation (9);
    Find the representative article $td_{c,l}$ for the topic t from $\mathcal{D}_{c,l}$ by max pooling (Equation (10));
  end
  Generate the sets $\mathcal{D}_{l-1}$ and $\mathcal{D}_{c,l-1}$;
  Calculate the matrix $\mathbf{W}_l$ based on Equation (11);
  Calculate $\mathbf{S}_{l-1}$ using Equation (12);
end
return $\text{TD} = \bigcup_{c,l} \text{td}_{c,l}$;

By repeating the preceding process, we choose one representative article for each topic. This algorithm is summarized in Algorithm 2.

4.2. Representative Image Selection

Current Web pages always contain multimodal information, such as text, image, and video. This work mainly focuses on news images. To give users a vivid and rich overview about their interested news, we summarize news images corresponding to a certain topic or subtopic, which is a complement to the preceding textual summarization.

Given the uncovered hierarchical topics, we can assemble news images included in certain news pages corresponding to each topic, since news images are associated with news documents. From the assembled image set corresponding to one topic, a representative image is selected to visualize the textual information of the topic. To provide informative visual information to the user, we also propose to deform the selected image to keep important objects (e.g., newspersons) as much as possible. The details will be introduced in the following.

An event is composed of a connected series of subevents with a common focus or purpose that happens in specific places during a given temporal period. The subevent associated with more images is more important and more representative for the event. As a consequence, we cluster images belonging to the same topic into several groups and then choose the most representative image from the biggest group to describe the topic. First, images belonging to the same topic are clustered into several groups. Since news images belonging to the same subevent are semantically similar and often visually similar, the near-duplicate (ND) detection algorithm is introduced to identify the semantically and visually similar group. For this purpose, the scale-rotation invariant pattern entropy (SR-PE) algorithm [Zhao and Ngo 2009] is used in this work[9] to detect the very similar images and segment the corpus into several groups. SR-PE can evaluate complex patterns composed of ND images under an unknown scale and rotation changes based on the bag-of-visual word (BoW) scheme, which can enable the task of news topic tracking. In this practical framework, three components (i.e., bag-of-words representation, local keypoint matching, and SR-PE evaluation) are jointly exploited for the rapid detection of NDs. Specially, ND pairs are first identified through the pattern entropy measure based on the observation that ND pairs share the duplication of regions and non-ND pairs often show random patterns. Other than the consideration of scaling and rotational effects, SR-PE utilizes mean shift to cluster

[9]Code is available at http://pami.xmu.edu.cn/∼wlzhao/sotu.htm.

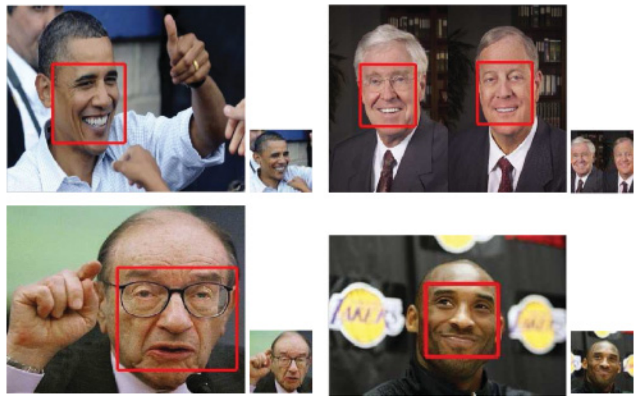

Fig. 4. The illustration of image downsizing. There are four groups: the original image with the detected face parts (the left one) and the downsized image (the right one).

visual ND images. As a result, the corpus of images is segmented into several groups $\mathcal{G}_r, r = 1, 2, \ldots, m$, where m is the number of groups in total. We can use the method in Hong et al. [2010] to mine the sequence of images. For simplicity, a simple strategy is introduced in this work to identify the representative image for each topic. The biggest group $r^* = \arg\max_r |\mathcal{G}_r|$ is chosen to visually represent this topic, where $|\mathcal{G}_r|$ denotes the number of images within $\mathcal{G}_r$. For the selected group $\mathcal{G}_{r^*}$, the pairwise similarity is calculated using the Gaussian kernel and then the graph is constructed. Based on the graph, the degree of each image is calculated by adding all similarities between this image and other images in the group. Finally, through the use of max pooling, the image with the biggest score is selected to be the representative image.

We should downsize the representative images with little information loss. Specifically, an image patch-based method is adopted to summarize each image, and the summary image should satisfy two properties: it should contain as much visual information from the original image as possible and should introduce as few artifacts as possible that were not in the original image. As we know, news images always contain the persons related to the news, and people are always more important than other objects in the news. We should guarantee to change the human facial parts as little as possible. Consequently, we first make use of the face detection method [Nilsson et al. 2007] to detect the facial parts. Then, we introduce "importance weights" as

$$\omega_v = \begin{cases} 1.0 & \text{if } area(v_f) > 0.5 * area(v) \\ 0.4 & \text{otherwise,} \end{cases} \tag{13}$$

where $area(v)$ denotes the area of patch v and v_f denotes the part of patch v corresponding to the face. In the experiments, the area is measured by the number of pixels. We introduce the importance weights into the objective function in Simakov et al. [2008]:

$$\min_T \frac{\sum_{v \subset O} \omega_v \cdot \min_{u \subset T} D(v, u)}{\sum_{v \subset O} \omega_v} + \frac{\sum_{u \subset T} \omega_{\hat{v}} \cdot \min_{v \subset O} D(u, v)}{\sum_{u \subset T} \omega_{\hat{v}}}. \tag{14}$$

Here, v and u are patches in the original image O and the target image T, respectively. $\hat{v} = \arg\min_{v \subset O} D(u, v)$ and $D(,)$ is a distance measure. We implement the gradual resizing procedure in a coarse-to-fine manner within a Gaussian pyramid, which can escape local minima and speeds up convergence. We present several examples of image downsizing in Figure 4, which demonstrates that the utilized approach can achieve the expected goal.

4.3. Time-Bias Subtopic Linking

We have discovered the hierarchical tree and chosen one representative article and one news image for each topic. How to thread the topics as the presentation of their parent topic is a challenging and necessary problem. We formulate this problem as finding a spanning tree of a graph while considering the time coherence. A spanning tree is a subgraph of the original one, which is a tree that connects all of the vertices together. Since the edge represents the similarity between subtopics, a spanning tree with larger weights may have more possibility of being proper. The best choice among allspanning trees may be the one with the biggest weights, i.e., the MST [Pemmaraju and Skiena 2003], which is a spanning tree with maximum weight. It can be computed by negating the weights for each edge using the faster algorithms [Chazelle 2000] with the complexity of $O(E)$, where E is the number of edges in the graph.

The timestamp is an important aspect for news. We should fuse the time information in the spanning process of MST. In this article, we propose a naive time-bias MST method. We rewrite the format of time as "YYYYMMDD," such as from "Sep. 12 2010" to "20100912," which is denoted as *date*. Then, as for news articles d and g, we measure their similarity based on date, which is defined as

$$S_{\text{date}}(d, g) = 1 - \frac{|\text{date}(d) - \text{date}(g)|}{\sum_k |\text{date}(d) - \text{date}(k)|}. \tag{15}$$

Then, we linearly combine the date similarity and the similarity obtained by weighted aggregation in Section 4.1.2 as

$$S_{\text{MST}}(d, g) = \xi S_l(d, g) + (1 - \xi) S_{\text{date}}(d, g), \text{ for } d, g \in \mathcal{D}_{c,l-1}, \tag{16}$$

where $\mathcal{D}_{c,l-1}$ is the set of representative articles of the topic at level l on path c, and d and g are the representative articles of its subtopics. In our experiments, we empirically set $\xi = 0.4$ to rely more on the time information.

Based on the weighted graph, we can abstract the spanning tree by incorporating the time information. For each topic, we utilize a time-bias MST to link and place all of its subtopics properly.

5. USER INTERFACE OVERVIEW

In this section, we introduce the designed interface for multimedia news search and summarization. In our system, we present a summary view apart from a concise result list as shown by many news Web services. Our interface contains two parts: the left part and the right part. We put the result list with the top three related newspersons and top three related news articles for each item in the right part, which are mined by MPMF [Li et al. 2010]. To provide a brief and rich view to users, we present the hierarchical topic structure related to a reader's query in the left part, as shown in Figure 5. In the both parts, we present multimedia information: text and image information. We elaborate on the summary view as follows.

As mentioned earlier, we adopt the hLDA model to explore the hierarchical topics. In our experiments, we set the depth of the tree to 3. The first level is the root node, the second level corresponds to the topic, and the third level corresponds to the subtopic in this article. In the left view of the interface, we summarize the search results with the preceding obtained hierarchical structure, where each blue rectangle frames the information belonging to a topic. Then, we employ the proposed time-bias MST to link the subtopics from left to right. For each subtopic, we present the keywords of its representative article and the chosen representative image. Here, we present an illustrative example of multimedia news summarization in search in Figure 1. We adopt hLDA to explore the hidden topic structure among a modest number (up to 1,000)

Fig. 5. The user interface of our designed system.

Table I. Details of Our Web News Dataset

Web Site	ABC	BBC	CNN	Google	Total
Articles (#)	47,163	11,073	41,649	35,423	135,308
Images (#)	31,789	5,309	16,636	15,410	69,144

of the top related news articles to the query "Barack Obama" [Lau et al. 2013]. The representative news article for each topic is identified by the approaches of tree-based article similarity, weighted aggregation, and max pooling. For each topic, the typical image is also chosen to complement the textual content to describe this topic. Finally, we utilize the time-bias MST to thread the subtopics of each topic.

Through the multimedia news summarization, we present the text summary and image summary in a hierarchical structure. Online news readers can browse news conveniently and access the desired information quickly and comprehensively.

6. EXPERIMENT PROTOCOL

In this section, we introduce our news dataset collected from multiple Web sites for the experimental verification. The objective of our experiments is to examine the effectiveness of our methods, which aim to provide a friendly, vivid, and compact interface to online news readers. We first explain the dataset and the experimental settings in this section and then present experimental results by empirical justification in Section 7.

6.1. Datasets and Experimental Settings

We build a large-scale multimedia news dataset collected from four news Web sites, including ABCNews.com,[10] BBC.co.uk,[11] CNN.com,[12] and Google News. There are 135,308 news articles and 69,144 news images in total, whose distribution over these four Web sites is shown in Table I. Considering that the person's name has the relative definitiveness and the meaning presented by one news event is usually broad, we mainly employ the person's name as a query in our experiments for the convenience of experiment evaluation. It is worth noting that for our proposed approaches, there is no limit for the query, which can be persons' names and news events. We chose 180

[10]http://abcnews.go.com.
[11]http://www.bbc.co.uk/news.
[12]http://edition.cnn.com.

Table II. Graded Score Criterion

Degree of Satisfaction	Score
Totally Dissatisfied	1
A Little Bit Satisfied	2
Satisfied	3
More Satisfied	4
Very Satisfied	5

Table III. Query List Used in the Experiments

Andy Murray	Andy Roddick	Angela Merkel	Barack Obama
Bill Gates	Bobby Dall	Bret Michaels	C. C. Deville
Carl Lewis	Carlo Ancelotti	Charlie Sheen	Chen Kaige
Cheryl Cole	Colin Powell	Dara Torres	Debbie Rowe
Dwight Howard	Dwyane Wade	Elena Dementieva	Eli Roth
George Soros	George W. Bush	Gerhard Schroeder	Ian Crocker
James Blake	Janet Jackson	Jason Lezak	Jerry Springer
John Sculley	Ju Dou	Justine Henin	Karl Malone
Kill Bill	Landon Donovan	Larry Ellison	Lebron James
Li Kaifu	Ma Huateng	Mariah Carey	Martin Bashir
Michael Dell	Michael Johnson	Michael Phelps	Michelle Obama
Michelle Yeoh	Novak Djokovic	Paul Allen	Pete Sampras
Quincy Jones	Ray Ozzie	Rikki Rockett	Robert Rodriguez
Scottie Pippen	Serena Williams	Sharon Osbourne	Simon Cowell
Steve Ballmer	Steve Jobs	Steve Nash	Steve Wozniak
Toni Kukoc	Tracy Mcgrady	Tyson Gay	Usain Bolt
Victoria Beckham	Virginie Razzano	Wayne Rooney	Whitney Houston
William Hague	Wim Duisenberg		

personalities from multiple domains, such as politics, sports, and business, to conduct our experiments.

No well-defined ground-truth dataset can be used to evaluate the performance of multimedia news summarization. Thus, we invited a group of 30 anonymous participants from two counties from a wide range of ages to evaluate the news summarization and the interface. The group of participants consisted of doctoral students, researchers, and technical staff, who declared that they were proficient in English and always read news online. They were asked to freely choose queries and search news. As defined in Table II, the participants could present five types of graded relevance according to their satisfaction regarding the results.

In addition, some parameters were set in advance. For the parameters in hLDA, we set $\alpha = 50$ and $L = 3$ and will discuss the parameters η and γ. For each query, we chose the top 1,000 related news articles for multimedia news summarization. How to better set their values or find a better default value is an open problem and will be considered in the future.

7. EXPERIMENTAL ANALYSIS

In this section, we conduct extensive experiments to evaluate our system and compare it with other works. Our experiments evaluate several aspects of the system, including parameter analysis for our system and performance comparison of our approaches with other related approaches. Each participant was asked to randomly choose three persons' names from our dataset to search. There are 70 queries in total selected from the 180 queries, and we present these 70 queries in Table III. Queries are selected from the rest of the queries to tune the hyperparameters in our methods.

Table IV. Average Number of Topics per Query from hLDA for Different η Under $L = 3$ and $\gamma = 1.0$

η	0.5	1.0	2.0	5.0	10
Topics (#)	187	41	23	7	4

Table V. Average Number of Topics per Query from hLDA for Different γ Under $L = 3$ and $\eta = 2.0$

γ	0.1	0.5	1.0	5.0	10
Topics (#)	3	8	23	79	485

Table VI. Average Score of Participants' Satisfaction for Different Values of ε

ε	0	0.2	0.4	0.5	0.6	0.8	1.0
Average Score	3.63	3.65	3.73	3.82	3.77	3.69	3.58

7.1. Parameter Sensitiveness

In hLDA, there exist two important parameters, η and γ, which affect the number of topics. Here, their impact for our summarization task are studied. The corpus of related news articles for each query is used to discover topics and record the number of topics for each query. We average the total number of generated topics and show results for different values of η and γ in Tables IV and V. It can be observed that small values of γ and large values of β suppress the number of topics. hLDA generates fewer topics with larger η and smaller γ, and more topics with smaller η and larger γ. When $\eta = 2.0$ and $\gamma = 1.0$, a reasonable number of topics is generated, and we set $\eta = 2.0$ and $\gamma = 1.0$ in the rest of the experiments.

The parameter ε balances the importance of similarities based on the textual information and the hierarchical tree. In this experiment, we tune it within the range of $\{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1.0\}$. The performance in terms of participants' satisfaction by varying ε is presented in Table VI. When $\varepsilon = 0$ or $\varepsilon = 1$, users are not satisfied with the summarizations. The average score of participants' satisfaction achieves the best value when $\varepsilon = 0.5$. Thus, ε is set to 0.5 in our experiments.

7.2. Summarization Evaluations

In this section, we evaluate the performance of multimedia news summarization and compare our approaches to other related methods. To evaluate the performance of news summarization, participants were asked to give scores about the results according to Table II. Each of them was asked to select three queries and determine their satisfaction with the returned results.

7.2.1. Experiment 1: On Representative News Article Selection. To validate that our method is effective to discover the representative article for each topic, we compare our approach to the following methods:

(1) *LDA* [Blei et al. 2003]: LDA with different cluster numbers ($K = 5, 15, 50$). The article with the biggest score calculated by a similar formula as that in Equation (9) based on $\mathbf{S}^{\text{text}}$ is selected.
(2) *hLDA-t*: hLDA based on $\mathbf{S}^{\text{text}}$ to find representative news articles without weighted aggregation.
(3) *SUMMA* [Christensen et al. 2014]: A hierarchical summarization system that summarizes multiple documents based on sentences.

We compare these methods in terms of the conciseness and comprehensiveness of their obtained summarization. Each participant gave a graded score for each approach. Figure 6 presents the average scores. Aside from the preceding methods, other special

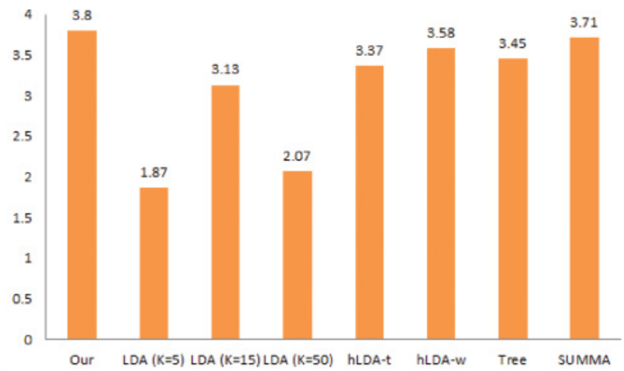

Fig. 6. The comparative results for representative news article selection. The average scores are shown for these approaches.

Table VII. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and LDA ($K = 15$) for Representative News Article Selection and ANOVA Test Results

Our approach vs. LDA ($K = 15$)		Factor of Ranking Scheme		Factor of Users	
Our method	LDA ($K = 15$)	F-statistic	p-value	F-statistic	p-value
$\mathbf{1.933 \pm 0.449}$	1.100 ± 0.305	42.647	3.758×10^{-6}	0.209	1.000

cases of our approach, i.e., hLDA-w ($\varepsilon = 1$ in Equation (3) in our method) and Tree (representative articles were selected without weighted aggregation in our method), are also evaluated. The performance of hLDA-based approaches and SUMMA are superior to LDA, which indicates advantages of the hierarchical topics. Since LDA assigns the same topic number to all queries, it is not reasonable and cannot achieve good average performance. In addition, our method remarkably outperforms hLDA-t, demonstrating the effectiveness of similarity based on the hierarchical tree and weighted aggregation. Users prefer our summarization results to SUMMA, as our method can provide a more comprehensive result. In addition, the improved performance of our method over Tree shows the importance and effectiveness of the proposed weighted aggregation. Without considering the tree-based similarity, hLDA-w is inferior to the proposed method. Thus, all aspects in our method for representative news article selection, i.e., text-based similarity, tree-based similarity, and weighted aggregation, are helpful for our system.

We also conduct experiments to compare our approach with LDA ($K = 15$), which achieves the best performance among the LDA-based methods, and hLDA-t in terms of the summary reasonableness. In each pairwise comparison, participants were allowed to freely search persons' names and compare the returned summarized results. They could choose "better," "much better," and "comparable" options for the comparison of two ranking schemes. We quantize the results as follows. We assign score 1 to the worst ranking scheme, and the other scheme is assigned a score 2, 3, or 1 if it is better, much better, or comparable to this one, respectively. Thus, there are 30 ratings for each comparison. The average rating scores and the standard deviation values are shown in Tables VII, VIII, and IX. Since there will be disagreements among the evaluators, we perform a two-way analysis of variance (ANOVA) test [King and Minium 1999] to statistically analyze the comparison; the results are also shown in the tables. From the results, it is observed that users obviously prefer our approach and the performance of our method statistically significantly outperforms others. The p-values show that the difference of the two schemes is significant, and the difference of users is insignificant.

Table VIII. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and hLDA-t for Representative News Article Selection and ANOVA Test Results

Our approach vs. hLDA-t		Factor of Ranking Scheme		Factor of Users	
Our method	hLDA-t	F-statistic	p-value	F-statistic	p-value
$\mathbf{1.967 \pm 0.615}$	1.200 ± 0.407	18.686	1.656×10^{-4}	0.152	1.000

Table IX. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and SUMMA, and ANOVA Test Results

Our approach vs. SUMMA		Factor of Ranking Scheme		Factor of Users	
Our method	SUMMA	F-statistic	p-value	F-statistic	p-value
$\mathbf{1.933 \pm 0.640}$	1.233 ± 0.504	13.198	1.100×10^{-3}	0.191	1.000

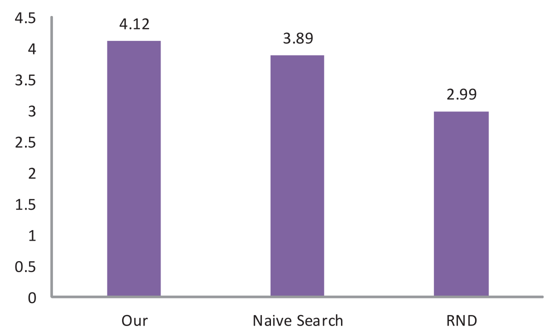

Fig. 7. The average scores of our approach, naive selection, and random selection for the representative image selection.

7.2.2. Experiment 2: On Representative Image Selection. We conducted experiments to evaluate the performance of our approach in choosing an image for each topic. Our method is compared to the following two methods:

(1) *NS (Naive Selection)*: The image corresponding to the selected representative news article is chosen for one topic.
(2) *RND (Random)*: A representative image is randomly selected from the corpus of images of one topic.

These methods are tested in terms of understandability of the selected image—that is, the selected image is helpful to understand the content of the topic. The comparative results are shown in Figure 7. The superiority of our approach is obviously observed.

Similar to Experiment 1, we also compare our method to the other two approaches and perform ANOVA by conducting the same process. Tables X and XI illustrate the mean and standard deviations of the rating scores as well as the ANOVA test results. From the numbers in the tables, we can see that the results of the representative image selection by our method is the best and the difference is statistically significant, which further confirms the effectiveness of our approach.

7.2.3. Experiment 3: On Subtopic Linking. The organization of topics is necessary for presenting the summarized results to users. The topic linking can be regarded as a ranking problem. To indicate the effectiveness of our proposed time-bias MST, it is compared to the method TimeLink, which links the subtopics of one topic purely based on the time information.

Participants were asked to give scores about the reasonableness of topic linking, and averaged scores of our approach and TimeLink are 4.06 and 3.67, respectively. This indicates the effectiveness of our subtopic linking approach.

Table X. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and Naive Selection for Representative News Image Selection and ANOVA Test Results

Our approach vs. NS		Factor of Ranking Scheme		Factor of Users	
Our method	NS	F-statistic	p-value	F-statistic	p-value
$\mathbf{1.867 \pm 0.629}$	1.200 ± 0.407	14.500	6.723×10^{-4}	0.220	0.9999

Table XI. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and RND for Representative News Image Selection and ANOVA Test Results

Our approach vs. RND		Factor of Ranking Scheme		Factor of Users	
Our method	RND	F-statistic	p-value	F-statistic	p-value
$\mathbf{2.400 \pm 0.455}$	1.133 ± 0.217	46.740	2.500×10^{-11}	0.250	0.999

Table XII. Average and Standard Deviation Values Converted from the User Study on the Comparison of Our Methods and TimeLink for Subtopic Linking, and ANOVA Test Results

Our approach vs. TimeLink		Factor of Ranking Scheme		Factor of Users	
Our method	TimeLink	F-statistic	p-value	F-statistic	p-value
$\mathbf{1.970 \pm 0.243}$	1.133 ± 0.195	33.263	3.02×10^{-6}	0.150	1.000

We conducted another user study with the same 30 participants and the process introduced in Experiment 1 to compare these two linking schemes. The mean and standard deviations of the rating scores and ANOVA test results are illustrated in Table XII. The results demonstrate the superiority of our approach over TimeLink. The ANOVA test shows that the superiority is statistically significant and the difference of the evaluators is not significant.

7.3. Interface Evaluations

In our news retrieval system, we developed a new style interface as shown in Figure 5. In this experiment, we evaluated the performance of our interface from a subjective view and compared it to other systems. While browsing results of the queries, each participant was asked to give scores according to Table II based on the following aspects:

—*Convenience*: The convenience of retrieval system is important for users to browse the results. Is it convenient to search and browse the news?
—*Efficiency*: Efficiency is a typical problem for the retrieval system. Users cannot tolerate much time for results to be returned. How long do the systems take to return results for each query?
—*Friendliness*: Users like Web pages that make them comfortable. Do the users enjoy the interface? Does the interface seem comfortable?
—*Diversity*: Do the systems show users many kinds of information? Can they present users' multiviews effectively?
—*Quick Overview*: News readers want to understand what they read as quickly as possible. Do the systems provide quick overviews to news readers?
—*Summary Effectiveness*: We should examine the summary's usefulness with respect to news retrieval. To what degree does it affect the effectiveness of news retrieval?

The average scores are presented in Figure 8. It can be observed that compared to other news service Web sites, although our system takes more time to return the results to readers, it can provide multimedia information, a more condensed view, and a convenient and friendly interface to users. In other words, users prefer the interface of our system, and our system can present vivid and comprehensive information conveniently. Through the multimedia information, readers are given a vivid overview about the returned results. Readers can quickly understand the information that they require via the multimedia summarization in our system. The runtime is also compared.

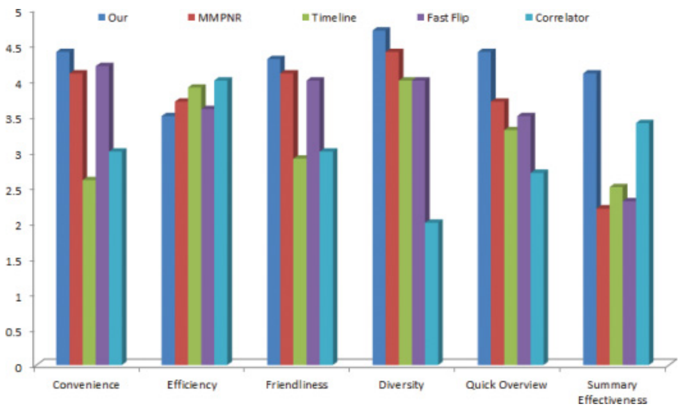

Fig. 8. The mean scores about the interfaces of several systems.

Unfortunately, it is not convenient to accurately obtain the average runtime for each query in the Timeline, Fast Flip, and Correlator systems. Thus, we only present the runtime of our system and MMPNR. For each query, the average times of our system and MMPNR are 2.24s and 1.96s, respectively. It indicates that our system is practical. In fact, our system can be speeded up by further optimizing the software.

8. CONCLUSION AND FUTURE WORK

In this article, query-related news multimedia information is summarized to provide a brief and vivid browsing view to online news readers. First, we adopt hLDA to explore the latent hierarchical topics. Next, we propose to measure similarities of articles based on the hierarchical tree, and we choose representative articles for topics based on weighted aggregation and max pooling approaches. Finally, the subtopics are threaded using the proposed time-bias MST, and the representative news images are selected for them. To display the news results, a vivid and informative interfaceis designed. The reasonable and comprehensive evaluations are performed to demonstrate the effectiveness of our approach. In the future, we will consider news videos, which can provide another aspect of information to news readers.

REFERENCES

David M. Blei, Thomas L. Griffiths, and Michael I. Jordan. 2010. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchical. *Journal of the ACM* 57, 2, 1–30.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, 4–5, 993–1022.

Gemma A. Calvert. 2001. Cross-modal processing in the human brain: Insights from functional neuron imaging studies. *Cerebral Cortex* 11, 12, 1110–1123.

Juan Cao, Jintao Li, Yongdong Zhang, and Sheng Tang. 2007. LDA-based retrieval framework for semantic news video retrieval. In *Proceedings of the IEEE International Conference on Semantic Computing*. IEEE, Los Alamitos, CA, 155–160.

Asli Celikyilmaz and Dilek Hakkani-Yur. 2010. A hybrid hierarchical model for multi-document summarization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 815–824.

Bernard Chazelle. 2000. A minimum spanning tree algorithm with inverse-Ackermann type complexity. *Journal of the Association for Computing Machinery* 47, 6, 1028–1047.

Yuanhao Chen, Benyu Zhang, and Hongjiang Zhang. 2009. Weighted aggregation based clustering algorithm for blog tag taxonomy construction. *Journal of Chinese Computer Systems* 30, 7, 1293–1297.

Janara Christensen, Stephen Soderland, Gagan Bansal, and Mausam. 2014. Hierarchical summarization: Scaling up multi-document summarization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 902–912.

M. Cohen and D. Massaro. 1990. Synthesis of visible speech. *Behavior Research Methods: Instruments and Computers* 22, 2, 260–263.

Asif A. Ghazanfar and Charles E. Schroeder. 2006. Is neocortex essentially multisensory? *Trends in Cognitive Sciences* 10, 6, 278–285.

Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 50–57.

Richang Hong, Jinhui Tang, Hung-Khoon Tan, Chong-Wah Ngo, Shuicheng Yan, and Tat-Seng Chua. 2010. Beyond search: Event driven summarization for Web videos. *ACM Transactions on Multimedia Computing, Communications, and Applications* 7, 4, 35.

Binxing Jiao, Linjun Yang, Jizheng Xu, Qi Tian, and Feng Wu. 2012. Visually summarizing Web pages through internal and external images. *IEEE Transactions on Multimedia* 14, 6, 1673–1683.

Bruce M. King and Edward M. Minium. 1999. *Statistical Reasoning in Psychology and Education*. Wiley, New York, NY.

Jey Han Lau, Timothy Baldwin, and David Newman. 2013. On collocations and topic models. *ACM Transactions on Speech and Language Processing* 10, 3, 10.

Zechao Li, Jing Liu, Jinhui Tang, and Hanqing Lu. 2015. Robust structured subspace learning for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 10, 2085–2098.

Zechao Li, Jing Liu, Meng Wang, Changsheng Xu, and Hanqing Lu. 2013. Enhancing news organization for convenient retrieval and browsing. *ACM Transactions on Multimedia Computing, Communications, and Applications* 10, 1, 1.

Zechao Li, Jing Liu, Xiaobin Zhu, and Hanqing Lu. 2010. Multi-modal multi-correlation person-centric news retrieval. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, 179–188.

Zechao Li, Meng Wang, Jing Liu, Changsheng Xu, and Hanqing Lu. 2011. News contextualization with geographic and visual information. In *Proceedings of the ACM International Conference on Multimedia*. ACM, New York, NY, 133–142.

Jianhua Lin. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 37, 1, 145–151.

I. Mani and E. Bloedorn. 1997. Multi-document summarization by graph search and matching. In *Proceedings of the National Conference on Artificial Intelligence*. 622–628.

I. Mani and M. Maybury. 1999. *Advances in Automatic Text Summarization*. MIT Press, Cambridge, MA.

Chris Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Harry McGurk and John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264, 5588, 746–748.

Kathleen R. McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Judith L. Klavans, Ani Nenkova, Carl Sable, Barry Schiffman, and Sergey Sigelman. 2002. Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In *Proceedings of the Human Language Technology Conference*. 280–285.

Joel Larocca Neto, Alexandre D. Santos, Celso A. A. Kaestner, Neto Alexandre, D. Santos, Celso A. A, Kaestner Alex, Alex A. Freitas, and Catolica Parana. 2000. Document clustering and text summarization. In *Proceedings of the International Conference on Practical Applications of Knowledge Discovery and Data Mining*. 41–55.

Mikael Nilsson, Jörgen Nordberg, and Ingvar Claesson. 2007. Face detection using local SMQT features and split up SNoW classifier. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, Los Alamitos, CA, 589–592.

C. D. Paice. 1990. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management* 26, 1, 171–186.

Sriram Pemmaraju and Steven Skiena. 2003. *Computational Discrete Mathematics: Combinatorics and Graph Theory in Mathematica*. Cambridge University Press, Cambridge, England.

Xuan-Hieu Phan, Cam-Tu Nguyen, Dieu-Thu Le, Le-Minh Nguyen, Susumu Horiguchi, and Quang-Thuy Ha. 2011. A hidden topic-based framework toward building applications with short Web documents. *IEEE Transactions on Knowledge and Data Engineering* 23, 7, 961–976.

D. Radev. 2000. A common theory of information fusion from multiple sources step one: Cross-document structure. In *Proceedings of the ACL SIGDIAL Workshop on Discourse and Dialogue*. 74–83.

Dragomir Radev, Jahna Otterbacher, Adam Winkel, and Sasha Blair-Goldensohn. 2005. NewsInEssence: Summarizing online news topics. *Communications of the ACM* 48, 10, 95–98.

D. R. Radev, S. Blair-Goldensohn, Z. Zhang, and R. S. Raghavan. 2001. Interactive, domain-independent identification and summarization of topically related news articles. In *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*. 225–238.

D. R. Radev and W. Fan. 2000. Automatic summarization of search engine hit lists. In *Proceedings of the ACL Workshop on Recent Advances in Natural Language Processing and Information Retrieval*. 99–109.

D. R. Radev and K. R. McKeown. 1999. Generating natural language summaries from multiple on-line sources. *Computational Linguistics* 24, 3, 469–500.

Dafna Shahaf, Carlos Guestrin, and Eric Horvitz. 2012. Trains of thought: Generating information maps. In *Proceedings of the ACM International Conference on World Wide Web*. 899–908.

Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani. 2008. Summarizing visual data using bidirectional similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Los Alamitos, CA, 1–8.

Ilija Subašić and Bettina Berendt. 2010. Experience stories: A visual news search and summarization system. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*. 619–623.

Giang Binh Tran. 2013. Structured summarization for news events. In *Proceedings of the ACM International Conference on World Wide Web Companion*. ACM, New York, NY, 343–348.

D. Trieschnigg and W. Kraaij. 2004. TNO hierarchical topic detection report at TDT 2004. In *Proceedings of the 7th Topic Detection and Tracking Conference*.

D. Trieschnigg and W. Kraaij. 2005. Scalable hierarchical topic detection. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 655–656.

T. A. Tuan, S. Elbassuoni, N. Preda, and G. Weikum. 2011. Cate: Context-aware timeline for entity illustration. In *Proceedings of the ACM International Conference on World Wide Web*. ACM, New York, NY, 269–272.

Xing Wei and W. Bruce Croft. 2006. LDA-based document models for ad-hoc retrieval. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 178–185.

R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang. 2011. Evolutionary timeline summarization: A balanced optimization framework via iterative substitution. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 745–754.

Wan-Lei Zhao and Chong-Wah Ngo. 2009. Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *IEEE Transactions on Image Processing* 18, 2, 412–423.

Received October 2014; revised July 2015; accepted September 2015