Multimedia Summarization for Trending Topics in Microblogs

Jingwen Bian
National University of Singapore, Singapore
bian_jingwen@nus.edu.sg

Yang Yang
National University of Singapore, Singapore
yang.yang@nus.edu.sg

Tat-Seng Chua
National University of Singapore, Singapore
chuats@comp.nus.edu.sg

ABSTRACT

Microblogging services have revolutionized the way people exchange information. Confronted with the ever-increasing numbers of microblogs with multimedia contents and trending topics, it is desirable to provide visualized summarization to help users to quickly grasp the essence of topics. While existing works mostly focus on text-based methods only, summarization of multiple media types (e.g., text and image) are scarcely explored. In this paper, we propose a multimedia microblog summarization framework to automatically generate visualized summaries for trending topics. Specifically, a novel generative probabilistic model, termed multimodal-LDA (MMLDA), is proposed to discover subtopics from microblogs by exploring the correlations among different media types. Based on the information achieved from MMLDA, a multimedia summarizer is designed to separately identify representative textual and visual samples and then form a comprehensive visualized summary. We conduct extensive experiments on a real-world Sina Weibo microblog dataset to demonstrate the superiority of our proposed method against the state-of-the-art approaches.

Categories and Subject Descriptors

H.3.1 [**Information storage and retrieval**]: Content Analysis and Indexing—*Abstracting methods*

General Terms

Algorithms, Experimentation, Performance

Keywords

Microblog, Summarization, Trending Topic, Social Media

1. INTRODUCTION

Recent years have witnessed the emergence of microblogging services that change the way people live, work and communicate. For example, Sina Weibo[1], one of the largest microblogging platforms on the Web, has attracted more than 500 million registered users, and the average number of daily active users has reached 46 million by the end of 2012[2]. Users are allowed to share multimedia content on such platforms, such as news, images and video links. With the wide availability of information sources, rapid information propagation and ease of use, microblogging has quickly become one of the most important media for sharing, distributing and consuming interesting contents, such as the trending topics.

Currently, some microblogging platforms, such as Sina Weibo, offer users the list of (manually created) hot trending topics, together with a set of related microblogs in each trend. Such service offers a potentially useful way to help users to conveniently gain a quick and concise impression of the current hot topics. In addition, users may obtain further understanding of the topics by browsing the related mircoblogs. However, due to the tremendous volume of microblogs and the lack of effective summarization mechanism in existing trending topic services, users are often confronted with incomplete, irrelevant and duplicate information, which makes it difficult for users to capture the essence of a topic. Therefore, it would be of great benefit if an effective mechanism can be provided to automatically mine and summarize subtopics (i.e., divisions of a main topic) from microblogs related to a given topic.

It is natural to formulate microblog summarization as a multi-document summarization (MDS) task, which has been extensively studied in information retrieval. Over the years, numerous methods have been proposed. Notable approaches include the frequency-based methods, e.g., SumBasic [8] and MEAD [9], and semantic-based methods such as LSA [4] or LDA [5]. In addition, other methods such as the graph-based methods (e.g., LexPageRank [3]) and machine learning-based methods [11] have also been proposed. However, the previous MDS techniques are designed for well-organized texts, such as the news articles. There exist two challenges that make it greatly difficult to directly utilize the traditional MDS techniques to summarize microblogs: 1) short text length (e.g., a maximum 140 characters for Twitter) and 2) noisy and/or irrelevant microblog contents. Recently, several attempts [1, 2, 6, 10] have been made to summarize microblog texts by considering the above limitations. For example, Sharifi et al. [10] summarized Twitter hot topics through finding the most commonly used phrase

CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

http://dx.doi.org/10.1145/2505515.2505652.

[1]http://www.weibo.com/
[2]http://en.wikipedia.org/wiki/Sina_Weibo/

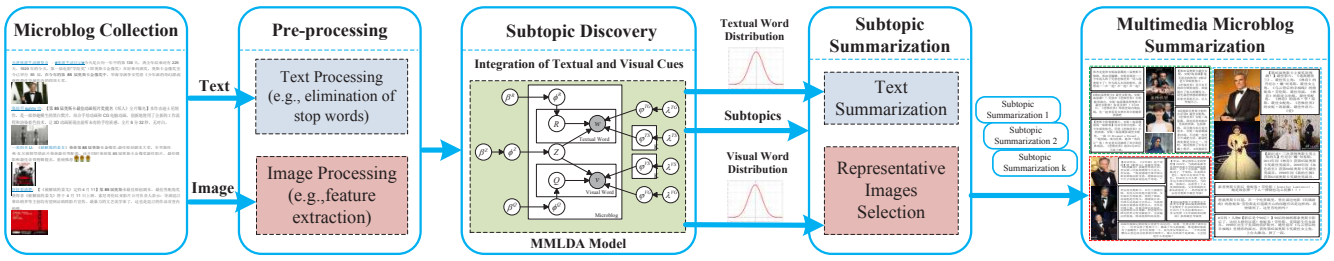

Figure 1: Flowchart of overall framework.

that encompasses the topic phrase. Inouye et al. [6] proposed a cluster summarizer based on the designed hybrid TF-IDF weighting. Nonetheless, the performance of these methods is still far from satisfactory because they did not address the problem of lack of expressive power caused by word restriction and noisy content.

Different from traditional documents that contain only textual objects, microblogs comprise of multiple media types, such as image and text. The high proportion of multimedia contents are precious resources for handling the above problems. It has been shown that integrating multiple types of media content may boost the performance of various real applications [14, 12]. The benefit of incorporating different media types into summarization is three-fold: 1) Images can supplement the textual content with additional information, especially in the circumstance of microblogging, where the text lacks sufficient expressive power as aforementioned. 2) Multimedia contents can facilitate subtopic discovery. Intuitively, given a trending topic, multimedia contents from different subtopics should have lower visual similarity while those within the same subtopic should have higher visual similarity. Thus, discriminative information embedded in visual information of multimedia contents can be exploited as critical cues for subtopic discovery. 3) Incorporating concrete multimedia exemplars into summarization can assist users to gain a more visualized understanding of interesting topics and/or subtopics.

However, it is non-trivial to integrate multimedia information to generate comprehensive summaries in the circumstance of microblogs. Without proper exploration of the intrinsic correlation among various media types, they may exert negative influence on each other. Based on the above analysis, in this paper we propose a novel framework to summarize multimedia microblogs for trending topics. Specifically, we first propose a novel generative probabilistic model, termed multimodal-LDA (MMLDA), to partition the microblogs related to the same topic into several subtopics. MMLDA model is capable of not merely capturing the intrinsic correlation between visual and textual information of microblogs, but also estimating the general distribution as well as subtopic-specific distribution under a trending topic. Based on the information achieved from MMLDA model, a summarizer is further elaborated to separately identify representative textual and visual samples and then form a comprehensive visualized summary. For text summarization, we specify three criteria, namely coverage, significance and diversity to measure the summarization quality. A greedy algorithm is developed to identify representative textual samples. For visual summarization, a two-step process is devised to automatically select the most representative images: 1) images within a subtopic are grouped by spectral clustering; and 2) images in each group are ranked by a manifold ranking algorithm and the top-ranked image is selected as representative. The flowchart of the proposed multimedia microblog summarization framework is illustrated in Figure 1. We can see that there are two main stages in the whole process: subtopic discovery and summary generation. We elaborate details of the two stages in the following sections.

The rest of the paper is organized as follows. Section 2 formally defines the problem. The details of the proposed MMLDA model for subtopic discovery are elaborated in Section 3. The process of generating multimedia summaries is introduced in Section 4. Experimental results are reported and discussed in Section 5, followed by the conclusion in Section 6.

2. PROBLEM DEFINITION

Suppose we are given a collection of microblogs $\mathcal{M} = \{M_1, ..., M_{|\mathcal{M}|}\}$ related to the same trending topic $\mathcal{T}$, where each microblog $M_i = \{T_i, I_i\}$ consists of two components: textual component T_i and visual component I_i. Note that I_i may be empty, which means M_i contains no visual samples. $|\cdot|$ denotes the cardinality of a set. The objective of our framework is to automatically generate a multimedia summary (e.g., both textual and visual) from the microblog collection $\mathcal{M}$ for revealing the multiple subtopics of topic $\mathcal{T}$. For topic $\mathcal{T}$, we define its topic-level summary as the union of all its subtopics' summaries, denoted as $\bigcup_{k=1}^{K} \mathcal{S}_k$, where K is the number of subtopics of $\mathcal{T}$. For each subtopic, a subtopic-level summary comprises both textual and visual exemplars idetified from $\mathcal{M}$.

3. SUBTOPIC DISCOVERY

In this section, we propose a novel generative probabilistic model, termed multimodal-LDA (MMLDA), to jointly model the relations between the visual and textual facets of microblogs in order to facilitate the subtopic discovery as well as the subsequent multimedia summary generation.

To model the relations among different media types, we introduce a shared latent variable Z to jointly model the semantics embedded in all media types. Different from traditional LDA which assigns a mixture of multiple latent topics to each document, the proposed MMLDA model associates only one semantic facet Z, i.e., subtopic, to each microblog. Note that microblog content is usually short and focused, hence it is reasonable to assume that each microblog only contains one subtopic. For each microblog, its textual words and visual words are generated from the same shared subtopic Z, where Z can be regarded as a semantic representation of the integration of visual and textual modalities in

terms of a shared distribution over factors that each visual or textual object can be assembled from.

We observe that all subtopics of the same topic may share some *general* words which indicate the common semantics related to the trending topic; while each individual subtopic uniquely possesses certain *specific* semantics, which distinguish itself from the other subtopics. Take the trending topic "Lushan Earthquake" as an example, "earthquake","Lushan" and "death" are more likely to be general words; while words like "hypocenter", "collapse" and "Premier" are more probable to appear in different subtopics. If the proportion of general contents is large, then they may dominate the result. In order to exclude the influence of general contents and discover discriminative cues for each subtopic, two new latent variables R and Q are introduced to generate the textual and visual words, respectively. For each textual (visual) word, R (Q) indicates whether it is generated from the general distribution or from the specific distribution corresponding to its subtopic. Next, we introduce the details of modeling and inference for MMLDA.

3.1 MMLDA Modeling

The graphical representation of MMLDA model is illustrated in Figure 2, and the detailed generation process is depicted as follows:

1. For the topic $\mathcal{T}$, draw $\varphi^{TG} \sim Dir(\lambda^{TG})$ and $\varphi^{VG} \sim Dir(\lambda^{VG})$ denote the *general* textual distribution and visual distribution, respectively. $Dir(\cdot)$ is the Dirichlet distribution. Then draw $\phi^Z \sim Dir(\beta^Z)$, which indicates the distribution of subtopics over the microblog collection corresponding to $\mathcal{T}$.

2. For each subtopic, draw $\varphi_k^{TS} \sim Dir(\lambda^{TS})$ and $\varphi_k^{VS} \sim Dir(\lambda^{VS})$, $k \in \{1, 2, ..., K\}$, correspond to the *specific* textual distribution and visual distribution.

3. For each microblog M_i, draw $Z_i \sim Multi(\phi^Z)$, corresponds to the subtopic assignment for M_i. $Multi(\cdot)$ denotes the Multinomial distribution. Then draw $\phi_i^R \sim Dir(\beta^R)$ indicates the general-specific textual word distribution of M_i. Similarly, draw $\phi_i^Q \sim Dir(\beta^Q)$ indicates that for visual words.

4. For each textual word position of M_i, draw a variable $R_{ij} \sim Multi(\phi_i^R)$:

   - If R_{ij} indicates *General*, then draw a word $w_{ij} \sim Multi(\varphi^{TG})$.
   - If R_{ij} indicates *Specific*, draw a word w_{ij} from the Z_i-th specific distribution $w_{ij} \sim Multi(\varphi_{Z_i}^{TS})$.

5. The generation of visual words is similarly done as in step 4.

3.2 Model Inference

In our MMLDA model, the subtopic assignment Z as well as general-specific indicators R and Q are latent variables to be inferred from the observations, i.e., textual and visual words. We use Gibbs sampling to achieve the inference due to its efficiency and effectiveness in handling high-dimensional data. With Gibbs sampling, the latent variables can be obtained, as well as the specific textual/visual distributions φ^{TS} and φ^{VS}. For any textual word w, $\varphi_k^{TS}(w)$ measures the probability of w appearing in the k-th specific textual distribution, while $\varphi_k^{VS}(u)$ measures that of visual distribution. These specific distributions can be inferred as follows:

$$\varphi_k^{TS}(w) = \frac{N^w(Z=k, R=S) + \lambda^{TS}}{\sum_{t \in V^t}(N^t(Z=k, R=S) + \lambda^{TS})} \quad (1)$$

$$\varphi_k^{VS}(u) = \frac{N^u(Z=k, Q=S) + \lambda^{VS}}{\sum_{u \in V^v}(N^u(Z=k, Q=S) + \lambda^{VS})} \quad (2)$$

where V^t and V^v denote textual and visual vocabulary, respectively. $N^w(Z=k, R=S)$ represents the number of textual word w in the k-th specific textual distribution after Gibbs sampling process stops. The meaning of $N^u(Z=k, Q=S)$ is similar.

With MMLDA model, we manage to discover the subtopics of a given trending topic by exploring relations among textual and visual modalities. The obtained textual-visual distribution pair φ_k^{TS}-φ_k^{VS} depicts the discriminative multimedia cues for each subtopic. According to the subtopic assignment Z for each microblog, MMLDA partitions the microblog collection $\mathcal{M}$ into K subsets $\{\mathcal{S}_k\}|_{k=1}^K$ corresponding to K subtopics. Further, we separate each subset into textual part $\mathcal{S}_k^t$ and visual part $\mathcal{S}_k^v$. Next, we will employ the above results achieved with MMLDA for summarization.

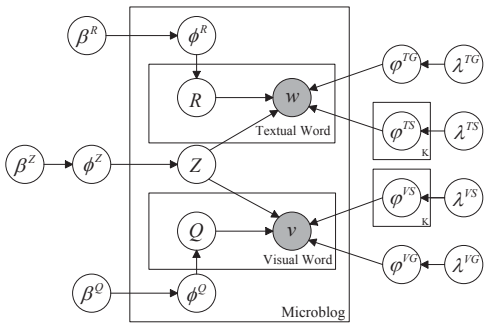

Figure 2: Graphical illustration of MMLDA model.

4. SUMMARY GENERATION

In this section, we first elaborate the processes of generating textual and visual summaries for each subtopic, by utilizing the reinforced textual/visual distributional information. Then, the textual and visual summaries are aggregated to form a comprehensive multimedia summary.

4.1 Text Summarization

We propose a text summarizer to automatically generate a summary for microblog text. Specifically, a greedy algorithm is developed to sequentially select representative samples based on a novel selection criterion, which takes three fundamental requirements into consideration: 1) **Coverage**, which implies that the summary should keep alignment with the original data collection, and reduce the information loss to the greatest extent; 2) **Significance**, which reveals that we should select microblogs that attract the greatest interests; and 3) **Diversity**, which indicates the summary should be concise and contain as little redundant information as possible.

In the following part, we introduce how the above three vital facets are embodied into the selection criterion. Without loss of generality, we consider the textual summary generation of the k-th subtopic from the subset $\mathcal{S}_k^t$. Denote $\mathcal{G}_k \subseteq \mathcal{S}_k^t$ as the summary set consists of the selected textual samples, and $\tilde{\mathcal{S}}_k^t = \mathcal{S}_k^t - \mathcal{G}_k$ is the remaining subset. In order to determine which sample is subsequently selected from $\tilde{\mathcal{S}}_k^t$, we calculate a selection score for each sample by considering coverage, significance and diversity as follows.

Coverage. Intuitively, if a summary is able to well "cover" the information of its corresponding subtopic, then the word distributions over both of them should be close to each other. The word distribution over a summary $\mathcal{G}_k$, denoted as $\Theta_{\mathcal{G}_k}$, can be estimated as:

$$p(w|\Theta_{\mathcal{G}_k}) = \frac{tf(w, \Theta_{\mathcal{G}_k})}{\sum_{t \in V^t} tf(t, \Theta_{\mathcal{G}_k})}, \ \forall w \in V^t \tag{3}$$

where $tf(w, \Theta_{\mathcal{G}_k})$ denotes the term frequency of word w in $\mathcal{G}_k$. φ_k^{TS} is used as the word distribution over the corresponding subtopic, which is the distribution estimated in the learning process of MMLDA model (Eq. 1). We employ Kullback-Leibler (KL) divergence to measure the distance of two distributions D_1 and D_2 as:

$$D_{KL}(D_1 \parallel D_2) = \sum\nolimits_w p(w|D_1) log \frac{p(w|D_1)}{p(w|D_2)} \tag{4}$$

Given the current summary set $\mathcal{G}_k$, the new sample T_i to be selected should be the one that makes the new summary (i.e., $\mathcal{G}_k \cup \{T_i\}$) achieve the best coverage (i.e., minimize the distance between $\Theta_{\mathcal{G}_k \cup \{T_i\}}$ and φ_k^{TS}). Therefore, the coverage of each candidate T_i could be measured by the following equation:

$$\mathcal{U}_C(T_i) = D_{KL}(\Theta_{\mathcal{G}_k \cup \{T_i\}} \parallel \varphi_k^{TS}) \tag{5}$$

Significance. In general, the popularity of a microblog can be revealed from the repost number. A large repost number means that the microblog has gained a lot of attention and interest from other users, and can indirectly represent the quality of this microblog. Therefore, we use the repost number to measure the significance of a candidate:

$$\mathcal{U}_S(T_i) = log(RepostNum(T_i) + 1) \tag{6}$$

Diversity. We take the information redundancy into consideration in sample selection. Consider a candidate T_i, the redundancy it brings to the summary set can be measured by the similarity between this candidate and the previously generated summary, which is:

$$\mathcal{U}_D(T_i) = D_{KL}(\Theta_{T_i} \parallel \Theta_{\mathcal{G}_k}) \tag{7}$$

Overall Selection Score. The overall selection score is defined as a weighted linear combination of the above three scores. Define $\mathcal{F}(x) = 1/(1 + exp(-x))$ as a logistic increasing function for normalizing the scores to interval $[0, 1]$. Since we favor a small distance for $\mathcal{U}_C(T_i)$, and large values for the other two criteria, the overall selection score is computed as:

$$\mathcal{U}(T_i) = \omega_1 (1 - \mathcal{F}(\mathcal{U}_C(T_i))) + \omega_2 \mathcal{F}(\mathcal{U}_S(T_i)) + \omega_3 \mathcal{F}(\mathcal{U}_D(T_i))$$

where $\omega_1, \omega_2, \omega_3$ are trade-off parameters with $\sum_i \omega_i = 1$.

With the above selection score for all the microblog samples, we may derive the greedy algorithm for representative sample selection. In each iteration, we calculate the selection scores for all the remaining samples in $\tilde{\mathcal{S}}_k^t$, and then move the one with the largest score from $\tilde{\mathcal{S}}_k^t$ to $\mathcal{G}_k$:

$$T^* = \arg\max_{T_i \in \tilde{\mathcal{S}}_k^t} \mathcal{U}(T_i) \tag{8}$$

4.2 Representative Image Selection

Consider the visual subset $\mathcal{S}_k^v$, which contains all images related to the k-th subtopic. The objective of this step is to select representative images which can best depict the current subtopic. The selected images should provide enough visual description as well as diverse viewpoints. We develop a two-step approach to automatically select representative images satisfying the above two criteria: 1) apply spectral clustering to group the image set $\mathcal{S}_k^v$ into visually diverse clusters; and 2) rank images within each cluster with a manifold ranking algorithm and the top-ranked ones are identified as representative samples.

Clustering step. Define the k-Nearest-Neighbors similarity graph of $\mathcal{S}_k^v$ as:

$$W_{ij} = \begin{cases} \exp(-\frac{d(x_i, x_j)^2}{\sigma^2}), & \text{if } x_i \in \mathcal{N}_k(x_j) \text{ or } x_j \in \mathcal{N}_k(x_i) \\ 0 & , \text{ otherwise.} \end{cases}$$

where x_i is the bag-of-visual-word representation of image $I_i \in \mathcal{S}_k^v$, $d(\cdot, \cdot)$ is the Euclidean distance, and σ is the bandwidth parameter. $\mathcal{N}_k(x_i)$ denotes the set of k nearest neighbors of x_i in $\mathcal{S}_k^v$. We then apply normalized cut to the image set to achieve visual diversity across the clusters.

Ranking step. In order to discover images with best representative ability within each cluster, we adopt manifold ranking algorithm to rank the images. Let D denote the degree matrix of W and $\mathbf{r}$ as the vector of ranking score, manifold ranking defines an iterative update process as follows:

$$\mathbf{r^{t+1}} = \gamma \mathbf{D^{-1/2} W D^{-1/2} r^t} + (\mathbf{1} - \gamma)\mathbf{h} \tag{9}$$

where $\mathbf{h}$ represents the vector of initial ranking scores, which is an all-one vector in standard manifold ranking setting. However, in our scenario, we expect $\mathbf{h}$ to possess the prior knowledge of the importance of each image. Recall that with MMLDA model, we have achieved the discriminative visual information for this subtopic, which is φ_k^{VS}. Intuitively, if an image is more consistent with φ_k^{VS}, it would have better descriptive ability for the whole subtopic image set, and should gain more emphasis. Therefore, instead of all-one vector which takes equal weighting for all images, we express $\mathbf{h}$ as prior knowledge measured by the KL divergency of an image I_i and φ_k^{VS}, i.e., $h_i = 1 - \mathcal{F}(D_{KL}(I_i \parallel \varphi_k^{VS}))$. By integrating the prior knowledge in the ranking scheme, the descriptive ability for the cluster as well as for the subtopic image set are both taken into consideration.

Finally, the image with the largest ranking value in $\mathbf{r}$ is selected from each cluster to construct the visual summarization set.

5. EXPERIMENTS

5.1 Dataset and Experimental Settings

We constructed the evaluation dataset by selecting 10 trending topics[3] from Sina Weibo. For each trending topic,

[3]Detailed topics are listed in Appendix.

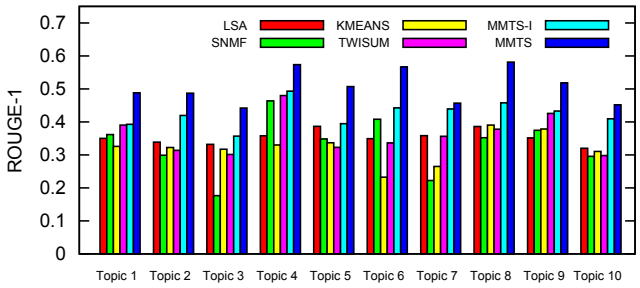

Figure 3: Detailed performance (ROUGE-1) of MMTS, MMTS-I and the comparing methods over all 10 topics.

we crawled the related microblogs in the life cycle of this topic. Due to limited information appended to repost action, only the original microblogs were included in our dataset, resulting in 127,118 microblogs and 48,656 images. In order to evaluate the quality of the generated summaries, five volunteers were invited to manually generate a textual summary for each topic as gold standard. Each manually generated summary consists of 50 microblogs selected from the microblog dataset.

In text pre-processing procedure, we first segmented Chinese words using IKAnalyzer[4], then removed the stop words, low-frequency words with document frequency of less than 5, and mentions (@somebody) from the textual vocabulary. Texts containing less than 3 words were also eliminated. For visual feature extraction, Scale-invariant feature transform (SIFT) descriptors were firstly extracted from each image. Then we trained a codebook of 1,000 visual words with descriptors sampled from images of all topics. With the trained codebook, each descriptor was quantized into a visual word. Each image was further represented as a 1,000-dimensional ℓ_2-normalized bag-of-visual-words feature.

For concentration parameters in MMLDA model, as stated in [5], the more specific a distribution is meant to be, the smaller its parameter. Accordingly, we set $\lambda^{TG}=0.1$, $\lambda^{TS}=0.01$, $\lambda^{VG}=1$, $\lambda^{VS}=0.1$, $\beta^{R}=0.1$, $\beta^{Q}=0.1$, and $\beta^{Z}=1$. For the final representation image selection procedure, the parameter γ is set to 0.85. According to our observation, the number of subtopics is limited for most topics. Therefore, we empirically set the subtopic number K to 10. While in our experiments, it is observed that some subtopics contain very small number of microblogs. We argue that such subtopics are probably composed of noisy microblogs and should be removed. In practice, we empirically choose a threshold $\epsilon=0.03$ and remove all the subtopics whose sizes are smaller than $\epsilon \cdot |\mathcal{M}|$. The total number of the selected microblogs is chosen to be 50, which is the same as the number of microblogs in the gold standards. The 50 microblogs quota are assigned to the remaining subtopics according to the proportion of microblog number in each subtopic.

5.2 Summarization Performance

We evaluate the effectiveness of our proposed framework as compared to several summarization approaches. For evaluation metric, we employ ROUGE evaluation toolkit [7] which automatically determines the quality of a summary as compared to human generated golden standards. In particular, F-measure scores of ROUGE-1, ROUGE-2, ROUGE-W (with W set to 1.2) and ROUGE-SU are reported. For fairness of evaluation, we select 50 microblogs for all the comparing approaches to form the summaries.

[4]http://code.google.com/p/ik-analyzer/

Table 1: Comparison among different summarization approaches. Average results of the 10 topics are reported for all evaluation measurement.

System	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-SU
LSA	0.3530	0.1547	0.0364	0.1386
SNMF	0.3302	0.1582	0.0291	0.1363
KMEANS	0.3209	0.0976	0.0283	0.0942
TWISUM	0.3602	0.1371	0.0353	0.1323
MMTS-I	**0.4239**	**0.2390**	**0.0523**	**0.1740**
MMTS	**0.5071**	**0.3039**	**0.0688**	**0.2327**

Table 2: Effects of coverage, significance and diversity criteria in subtopic discovery.

	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-SU
MMTS-C	0.4759	0.2483	0.0591	0.2127
MMTS-S	0.4143	0.2158	0.0473	0.1605
MMTS-D	0.4584	0.2684	0.0582	0.1988
MMTS	**0.5071**	**0.3039**	**0.0688**	**0.2327**

We compare our proposal with the following summarization approaches: 1) LSA [4], which conducts SVD on sample by term matrix first and select samples with highest entry value starting from most significant left eigenvector. 2) SNMF [13], which constructs the sample-sample similarity matrix first, clusters all samples with Symmetric Nonnegative Matrix Factorization (SNMF) and extracts centering sentences from the clusters. 3) KMEANS [9], which performs K-means clustering over the dataset and selects samples nearest to cluster centers. 4) TWISUM [6], which is a twitter summarization algorithm based on the proposed hybrid TF-IDF measurement. For our proposed approach, two specific methods are evaluated for comparison: 1) MMTS, the whole framework that uses both text and visual contents in constructing MMLDA model. 2) MMTS-I, which adopts the framework of MMTS without utilizing the visual information, i.e., in the subtopic discovery stage, when applying MMLDA model, all microblog samples are assumed to be comprised of texts only.

The overall comparison of proposed MMTS and MMTS-I with the other approaches are shown in Table 1. In addition, detailed ROUGE-1 performance for each topic is shown in Figure 3. As seen from the results, the proposed MMTS outperforms other methods for all topics as well as all evaluation measurements. The good performance of MMTS benefits from the following three aspects:

First of all, visual information is utilized in MMTS, the impact of which can be demonstrated by comparing the results of MMTS and MMTS-I. The latter approach differs from MMTS only with the lack of visual component. The performance illustrates the degradation of summarization ability when visual information is not used.

Secondly, MMTS discovers subtopics before the summarization procedure. As a result, all important branches for

Figure 4: A multimedia microblog summarization example on Topic 1.

a topic are covered in the final summarization. Although some comparing methods also consider the coverage of the summarization for the dataset, the coverage is only considered at the topic-level rather than the subtopic-level. In case a subtopic contains a small number of microblogs, there is a high probability that the microblogs related to this subtopic will be ignored by the comparing methods. The high performance of MMTS-I as compared to all the baseline methods demonstrates the effectiveness of subtopic discovery for enhancing the summarization performance.

Thirdly, three criteria (coverage, significance and diversity) are specified in MMTS, which are able to further facilitate the summary generation. We conduct experiment to evaluate the effectiveness of each individual component by removing each of the three criteria from our framework. The result is shown in Table 2. MMTS-C (MMTS-S or MMTS-D) denotes the method without taking coverage (significance or diversity) into consideration. The three trade-off parameters are set to $\omega_1 = 0.33$, $\omega_2 = 0.33$, $\omega_3 = 0.33$ for MMTS. While for the three comparing methods, we set ω to 0.5 for the two considered criteria, and 0 for the removed criterion. As can be seen, the performance of removing any criterion becomes worse, which illustrates that all components are necessary for our framework. Specifically, removing significance degrades the results the most, which demonstrates that users would more favor hot microblogs with large repost numbers.

An example of our summarization result is shown in Figure 4. This is a summary on Topic 1. Due to space limitation, only four subtopics are shown. For each subtopic, we list the top 3 images and top 3 texts. This example demonstrates the ability of our proposed framework in 1) well organizing the messy microblogs into structured subtopics; 2) generating high quality textual summary at subtopic level; and 3) selecting images relevant to subtopic that can best represent the textual contents.

6. CONCLUSION

In this paper, we proposed a multimedia microblog summarization method to automatically generate visualized summaries for trending topics. Specifically, a novel multimodal-LDA (MMLDA) model was proposed to discover various subtopics as well as the subtopic content distribution from microblogs, which explores the correlation among different media types. Based on MMLDA, a summarizer is elaborated to generate both textual and visual summaries. We conducted extensive experiments on a real world Sina Weibo microblog dataset to show the superiority of our proposed method as compared to the state-of-the-art approaches.

7. ACKNOWLEDGMENTS

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

8. REFERENCES

[1] D. Chakrabarti and K. Punera. Event summarization using tweets. In *ICWSM*, 2011.
[2] Y. Chang, X. Wang, Q. Mei, and Y. Liu. Towards twitter context summarization with user influence models. In *WSDM*, 2013.
[3] G. Erkan and D. R. Radev. Lexpagerank: Prestige in multi-document text summarization. In *EMNLP*, 2004.
[4] Y. Gong and X. Liu. Generic text summarization using relevance measure and latent semantic analysis. In *SIGIR*, 2001.
[5] A. Haghighi and L. Vanderwende. Exploring content models for multi-document summarization. In *NAACL HLT*, 2009.
[6] D. Inouye and J. K. Kalita. Comparing twitter summarization algorithms for multiple post summaries. In *SocialCom*, 2011.
[7] C.-Y. Lin. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: ACL-04 Workshop*, 2004.
[8] A. Nenkova and L. Vanderwende. The impact of frequency on summarization. *Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101*, 2005.
[9] D. R. Radev, H. Jing, M. Styś, and D. Tam. Centroid-based summarization of multiple documents. *Information Processing & Management*, 2004.
[10] B. Sharifi, M.-A. Hutton, and J. Kalita. Summarizing microblogs automatically. In *NAACL HLT*, 2010.
[11] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen. Document summarization using conditional random fields. In *ACL*, 2007.
[12] J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen. Inter-media hashing for large-scale retrieval from heterogeneous data sources. In *SIGMOD*, 2013.
[13] D. Wang, T. Li, S. Zhu, and C. Ding. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *SIGIR*, 2008.
[14] Y. Yang, Y. Yang, and H. T. Shen. Effective transfer tagging from image to video. *TOMCCAP*, 2013.

APPENDIX

Ten topics used in this work:

“第85届奥斯卡金像奖”，“Happy白色情人节”，“阳春三月，北京下雪”，“广药加多宝拼抢红罐归属权”，“柴静：《看见》”，“莫言荣获诺贝尔文学奖”，“美国春晚：2013超级碗”，“陕西房姐龚爱爱被爆有多处房产”，“聚美优品被指售假货”，“疯狂黄金周，丽江上万旅客打地铺”。