International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

An Efficient Multimedia summarization system using natural language processing

Swapnali Mohan Vazarkar and Prof. Majusha Tatiya

Department of Computer Engineering, Indira College of Engineering and Management, Parandwadi, Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Microblogging offerings have revolutionized the way human beings trade facts. Confronted with the ever-growing numbers of microblogs with multimedia contents and trending topics, it's far proper to offer visualized summarization to assist users to quickly hold close the essence of topics. While existing works normally attention on text-based strategies best, summarization of a couple of media sorts (e.g., text and image) are scarcely explored. In proposed approach a multimedia microblog summarization framework to automatically generate visualized summaries for trending topics. Specifically, a novel generative probabilistic model, termed multimodalLDA (MMLDA), is proposed to find subtopics from microblogs by means of exploring the correlations amongst different media kinds based on the records accomplished from MMLDA, a multimedia summarizer is designed to one by one pick out representative textual and visual samples and then form a complete visualized summary.

Keywords: *Microblog, Summarization, Trending Topic, Social Media, MMLDA*

Introduction

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 Twitter, 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.

Motivation

1. A multimedia microblog summarization method to automatically generate short summaries for trending topics.
2. Automatically generate a fixed-length textual summary to represent the principle content of the microblog data.
3. The Microblog Topic Modeling is that textual descriptions of images often provide important information about semantic aspects (topics), and image features are often correlated with semantic topics.

Literature Survey

P. Sinha, et al [1] Proposed strategies to process quality, assorted variety and inclusion properties utilizing multidimensional substance and setting information. The proposed measurements which will assess the photograph rundowns dependent on their portrayal of the bigger corpus and the capacity to fulfill client's data needs. Focal points are: The ravenous calculation for synopsis performs superior to the baselines. Synopses help in successful sharing and perusing of the individual photographs. Impediments are: Computation is costly.

H. Lin et al [2] in multi-record synopsis, excess is an especially significant issue since literary units from various archives may pass on a similar data. A top notch (little and significant) rundown ought not exclusively be instructive about the rest of additionally be minimized (non-excess). Points of interest are: The best execution is accomplished. Submodular synopsis accomplishes better ROUGE-1 scores. Impediments are: The proposed framework over the top expensive to fathom.

M. S. Bernstein et al [3] Eddi is a novel interface for perusing Twitter streams that groups tweets by points inclining inside the client's own channel. A calculation for subject identification and a theme arranged UI for social data streams, for example, Twitter channels. (1) benchmark TweeTopic against other subject identification approaches, and (2) contrast Eddi with a normal ordered interface for expending Twitter channels. Favorable circumstances are: A basic, novel subject recognition calculation that utilizes thing phrase location and a web crawler as an outer information base. Eddi is more pleasant and more proficient to peruse than the customary sequential Twitter interface. Burdens are: Users approached our customers temporarily, making it hard to extrapolate ends on how the instrument may be utilized longitudinally. Clients were seeing the historical backdrop of their feed as opposed to tweets they had never observed, making our assignment somewhat less sensible.

P. Goyal et al [4] proposes the clever thought of utilizing the setting touchy record ordering to improve the sentence extraction-based archive rundown task. Right now, a setting touchy archive ordering model dependent on the Bernoulli model of arbitrariness. Advantages are: The new context-based word indexing gives better performance than the baseline models. Disadvantages are: Need to calculate the lexical association over a large corpus.

D. Chakrabarti et al [5] in this paper we argue that for some profoundly organized and repeating occasions, for example, sports, it is smarter to utilize progressively complex methods to outline the applicable tweets. The issue of abridging occasion tweets and give an answer dependent on learning the fundamental concealed state portrayal of the occasion by means of Hidden Markov Models. Favorable circumstances are: The benefit of utilizing existing inquiry coordinating advancements and for basic oneshot occasions, for example, quakes it functions admirably. The HMM can learn contrasts in language models of sub-occasions totally consequently. Impediments are: The burden that SUMMHMM needs to represent tweet words that just happen in a portion of the occasions, yet not in others.

J. Bian et al [6] the paper proposes a multimedia social occasion summarization framework to automatically generate visualized summaries from the microblog circulation of multiple media kinds. Specifically, the proposed framework contains three ranges: 1) A noise removal approach is first devised to cast off probably noisy pictures. 2) A novel move-media probabilistic version, termed Cross-Media-LDA (CMLDA), is proposed to together find out subevents from microblogs of more than one media sorts. 3) Finally, based at the move-media know-how of all of the observed subevents. Advantages are: Eliminates the doubtlessly noisy pics from raw microblog photograph collection. Generate the multimedia summary for social activities utilising the go-media distribution information of all the located subevents. Disadvantages are: Need to extend the pass-media framework for routinely detecting social occasions and retrieving related candidate microblogs. Need to personalised microblog summarization based totally on person profile.

Z. Li et al [7] in paper, proposes a singular Robust Structured Subspace Learning (RSSL) algorithm with the aid of integrating image knowledge and function gaining knowledge of into a joint studying framework. The scholarly subspace is went with as a transitional zone to decrease the semantic empty between the lowdegree seen abilities and the high-organize semantics. Favorable circumstances are: The proposed RSSL empowers to successfully investigate a strong based subspace from records. The proposed system can diminish the commotion incited vulnerability.

W. Y. Wang et al [8] the paper proposes a singular matrix factorization technique for extractive summarization, leveraging the success of collaborative filtering. First to consider illustration learning of a joint embedding for textual content and snap shots in timeline summarization. Advantages are: It is straightforward for builders to set up the device in realworld packages. Scalable method for studying lowdimensional embedding's of information tales and snap shots. Disadvantages are: Only work on summarizing synchronous multi-modal content.

Z. Li, et al [9] In paper, expand a novel method of multimedia news summarization for looking consequences on the Internet, which uncovers the underlying subjects among question-associated news information and threads the news occasions inner every topic to generate a query-related brief evaluate. HLDA is adopted to find out the hierarchical subject matter shape from the question-related information articles, and an approach based totally on the weighted aggregation and max pooling to identify the typical news article for each subject matter is proposed. A time-bias MST technique is evolved to thread the subtopics inside one topic to give a news precis on every subject matter in phrases of temporal and spatial improvement. Advantages are: Proposed gadget can gift shiny and complete data without problems. Readers can fast recognize the information that they require thru the multimedia summarization in this system.

P. Li, et al [10] In paper, take a look at the hassle of learning to summarize pix via textual content and visualize text utilizing pics, that is known as

MutualSummarization. Thus separates the web image-text statistics space into 3 subspaces, namely pure photo area (PIS), natural textual content area (PTS) and photograph-text joint space (ITJS). Advantages are: In photograph summarization method, map photos from PIS to ITJS via picture class model and describe these snap shots making use of several excessive stage semantic sentences. In textual content visualization system, map text from PTS to ITJS thru text categorization version and then provide a visual show utilizing photographs with high confidences in ITJS. Disadvantages are: Need to improve the MutualSummarization overall performance.

Fei Wu, et al [11] proposes an approach to learn the hidden aspects in the topics via non-parametric Bayesian model for multimedia summarization, namely aspect learning for multimedia summarization via nonparametric Bayesian (ALSNB). More specifically, we introduce the priors of beta-Bernoulli process and Dirichlet process into the traditional dictionary learning. As a result, the proposed approach is able to adaptively identify the particular aspects of individual topic.

Gap Identification:

The section II represents the overview of literature survey on given papers: [1], [5], [7], [8], [9], [10] and [11] for study all the details about how to generate summary using textual data. The paper [3] and [6] represents the microblog summarization. The paper [8] and [9] elaborates the news contents summary for the people gets knowledge about news and images from multimedia websites. All referring [2], [4] and [11] these papers for improving the accuracy of asynchronous collections such as text and image in proposed methodology section.

Proposed Methodology

Traditional documents that contain only textual objects, microblogs constitute of multiple media types, such as image and text. In this paper proposes a novel framework to summarize multimedia microblogs for trending topics. Specifically, first proposes a novel generative probabilistic model, called multimodal-LDA (MMLDA), to partition the microblogs relevant to the same topic into different 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. For text summarization, specifies three criteria, namely coverage, significance and diversity to measure the summarization quality. 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 Fig. 1 shows the architecture of multimedia microblog summarization system. 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.

A. Architecture

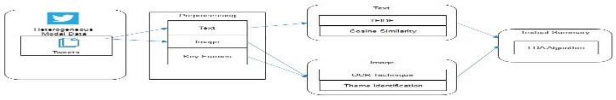

Fig. 1 System Architecture

Advantages

- It provides to automatically mine and summarize subtopics (i.e., divisions of a main topic) from microblogs related to a given topic.
- Microblogs comprise of multiple media types, such as image and text.
- Multimedia contents can facilitate subtopic discovery.
- Well organizing the messy microblogs into structured subtopics.
- Generating high quality textual summary at subtopic level.
- Selecting images relevant to subtopic that can best represent the textual contents.

B. Mathematical Model:

Consider the textual summary generation of the k-th subtopic from the subset $S_k{}^t$. Denote $G_k \subseteq S_k{}^t$ as the summary set consists of the selected textual samples, and $S_k{}^t = S_k{}^t - G_k$ is the remaining subset. In order to determine which sample is subsequently selected from $S_k{}^t$, we calculate a selection score for each sample by considering coverage, significance and diversity as follows.

Coverage

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

$$u_C(T_i) = D_{KL}(\Theta_{G_k \cup \{T_i\}} || \varphi_K{}^{TS}) \qquad (1)$$

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:

$u_S(T_i) = log(RepostNum(T_i) + 1)$ (2)

Diversity: We take the information redundancy into consideration in sample selection. Consider a candidate Ti, the redundancy it brings to the summary set can be measured by the similarity between this candidate and the previously generated summary, which is:
$u_D(T_i) = D_{KL}(\Theta_{Ti}||\Theta_{Gk})$ (3)

C. Algorithms

1. MMLDA Algorithm
Steps:

1. For the topic 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 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 Mi. Multi(·) denotes the Multinomial distribution. Then draw $\phi_i{}^R \sim Dir(\beta^R)$ indicates the generalspecific 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 $W_{ij} \sim Multi(\varphi_Z{}^{TS}{}_i)$
5. The generation of visual words is similarly done as in step 4.

2. Optical character recognition (OCR) Algorithm
Step 1: Image Preprocessing
Step 2: Edge Detection
Step 3: Detection of Text Regions
Step 4: Enhancement & Segmentation of Text Regions

3. Hidden Markov Model (HMM) algorithm for speech recognition:
A HMM is characterized by 3 matrices viz., A, B and PI.
A - Transition Probability matrix $(N \times N)$
B - Observation symbol Probability Distribution matrix
$(N \times M)$
PI - Initial State Distribution matrix $(N \times 1)$
Where, N =Number of states in the HMM
M = Number of Observation symbols
After can apply HMM for speech recognition by using following steps:

1. Recursive procedures like Forward and Backward Procedures exist which can compute P (O|L), probability of observation sequence.

Forward Procedure:
i) Initialization:
$\alpha_1(i) = \pi_i b_i o_1, 1 \le i \le N$ (4) ii) Induction

$$\alpha t{+}1(j) = [\sum_{i=1}^{N} \alpha t(i)aij]\ bj(ot{+}1),$$
$1 \le t \le T - 1, 1 \le j \le N$ (5) iii) Termination
$$P(O|\lambda) \sum_{i=1}^{N} \alpha_T(i) \quad (6)$$

Backward Procedure: i) Initialization:
$\beta_T(i) = 1, 1 \le i \le N$ (7) ii) Induction
$$\beta_T(i) = \sum_{j=1}^{N} a_{ij}\, b_j(o_{t+1})\beta_{t+1}(j), \quad (8)$$
$T - 1 \le t \le 1, 1 \le i \le N$ iii) Termination
$$P(O|\lambda) \sum_{i=1}^{N} \alpha_T(i) \quad (9)$$

2. The state occupation probability $t(sj)$ is the probability of occupying state sj at time t given the sequence of observations $O_1, O_2, \dots, O_N$.
3. Baum-welch algorithm for parameter re-estimation.

D. Dataset

We use twitter realtime microblog dataset using apps.twitter.com developer website.

Result and Discussions

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. Some functions used in the algorithm are provided by list of jars like Twitter-core and Twitter-stream jars etc. Tweets are retrieved in a streaming way, and Twitter provides the Streaming API for developers and researchers to access public tweets in real time. The aim of this paper is to bridge the gap by carrying out a performance evaluation, which was from two different aspects of NLP and machine leaning algorithms. Some functions used in the algorithm are provided by list of jars like standford core NLP jar for keywords extraction using POS tagger method. This is very useful for implementation of MMLDA algorithm. The OCR algorithm is used to extract text from image using tesseract training dataset. The resultant outcome is to display the list of tweets with images and videos summary using MMLDA algorithm. The Micro-blog Summarization Framework is based on twitter posts. Fig. 2 shows that when summarizing micro-blog data which gives better accuracy. The image file applies the OCR algorithm for textual part extraction and generates summary.

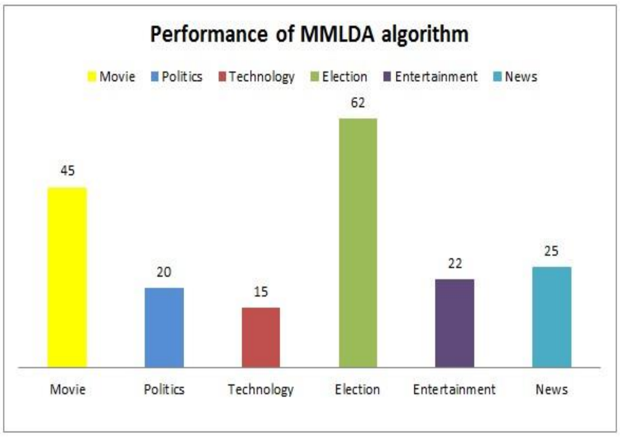

Fig. 2 Performance graph for Micro-blog Summarization framework

Conclusion

In this paper, we proposed a multimedia microblog summarization method to automatically generate visualized summaries for trending topics. Microblogs comprise of multiple media types, such as image and text and video. 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. Well organizing the messy microblogs into structured subtopics. Generating high quality textual summary at subtopic level.

References

[1]. P. Sinha, S. Mehrotra, and R. Jain, "Summarization of personal photologs using multidimensional content and context," in Proc. 1st ACM Int. Conf. Multimedia Retrieval, 2011, p. 4.

[2]. H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in Proc. Human Lang. Technol.: Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 912–920.

[3]. M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: Interactive topic-based browsing of social status streams," in Proc. 23nd Annu. ACM Symp. User Interface Softw. Technol., 2010, pp. 303–312.

[4]. P. Goyal, L. Behera, and T. M. Mcginnity, "A context-based word indexing model for document summarization," IEEE Transactions on Knowledge & Data Engineering, vol. 25, no. 8, pp. 1693–1705, 2013.

[5]. D. Chakrabarti and K. Punera, "Event summarization using tweet," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 66–73.

[6]. J. Bian, Y. Yang, H. Zhang, and T.-S. Chua, "Multimedia summarization for social events in microblog stream," IEEE Transactions on Multimedia, vol. 17, no. 2, pp. 216–228, 2015.

[7]. Z. Li, J. Liu, J. Tang, and H. Lu, "Robust structured subspace learning for data representation," IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 10, pp. 2085– 2098, 2015.

[8]. W. Y. Wang, Y. Mehdad, D. R. Radev, and A. Stent, "A lowrank approximation approach to learning joint embeddings of news stories and images for timeline summarization," in NAACL-HLT, 2016, pp. 58–68.

[9]. Z. Li, J. Tang, X. Wang, J. Liu, and H. Lu, "Multimedia news summarization in search," ACM Transactions on Intelligent Systems and Technology, vol. 7, no. 3, p. 33, 2016. [10] P. Li, J. Ma, and S. Gao, "Learning to summarize web image and text mutually," in Proceedings of the 2nd ACM International Conference on Multimedia Retrieval. ACM, 2012, p. 28.

[10]. [11] Fei Wu, Hanyin Fang, Xi Li, Siliang Tang, Weiming Lu, Yi Yang, Wenwu Zhu, Yueting Zhuang" Aspect learning for multimedia summarization via non-parametric Bayesian" IEEE 2016