

# A Survey on “A heterogeneous data Summarization System to Generate Automatic Summary of Data using Tweeter API using Tweets, Images”

Priyanka Ravindra Wadagave<sup>1</sup>, Dr. Bindu R. Garg<sup>2</sup>

<sup>1</sup>M.Tech student, <sup>2</sup>Associate Professor, Bharati Vidyapeeth(deemed to be) University, College of Engineering, Pune.

**Abstract-** Social media is ever-developing field of microblogs with enormous quantities of energized news and trending topics, even though it's quite rare to understand the essence of topic from given microblogs. In this paper, proposes a multimedia microblog summarization framework to automatically generate visualized summaries for trending topics. Depending upon the results given by MMLDA, a multimodal summarizer is intended to create a total perspective on rundown by getting printed and visual examples from microblogs.

**Keywords-** Microblog, Summarization, Trending Topic, Social Media, OCR, MMLDA

## I. INTRODUCTION

With the great variety and availability of information from large number of sources, high rate of information propagation and ease of use, microblogging has rapidly become famous media for sharing, distributing and consuming exciting contents of trending topics. Various microblogging platforms, such as Twitter provides users the list of hot trending topics which are manually created, it also provides a set of related microblogs in each trend. Such service offers a conceivably useful way to help users to conveniently and easily gain a quick and concise understanding extract of the current hot topics. For additional information gain, users may get further information of topics by browsing the related microblogs. However, due to the large volume of microblogs and information with incomplete, inapplicable and incorrect features it is difficult for users to extract the subject of topic. Therefore, it would be really beneficial if an effective method can be provided to automatically mine and summarize subtopics (i.e., divisions of a main topic) from microblogs related to a given topic.

## II. RELATED WORK

In this section, we briefly review the related work on text summarization along with machine learning algorithm

[1] Proposed methods to compute quality, diversity and coverage properties using multidimensional content and context data.

[2] In multi-document summarization, redundancy is a particularly important issue since textual units from different documents might convey the same information.

[3]. This paper an algorithm for topic detection and a topic-oriented user interface for social information streams such as Twitter feeds.

[4] Proposes the novel approach of using the context sensitive document indexing in order to improve the summarization task based on sentence extraction-based document

[5] In this paper we argue that for some highly structured and recurring events, such as sports, it is better to use more sophisticated techniques to summarize the relevant tweets

[6] to generate visualized summaries from the microblog circulation of multiple media kinds this paper proposes a framework based on multimedia social occasion summarization.

[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

[8] For extractive summarization this paper proposes a singular matrix factorization technique.

[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

[10] In paper, take a look at the hassle of learning to summarize pix via textual content and visualize text utilizing pics,

Sr No	Paper Name	Author	Limitations	Proposed work
1	Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences	Shiva Twinandillaa , SatriyoAdhya , BayuSurarsob , RetnoKusumaningrum a *	It's difficult to summarize document which consist of this entire format	combined K-Means algorithm and LDA significance sentence . So it can generate the documents summerize based on topic.
2	Multimedia news summarization in search	Z. Li, J. Tang, X. Wang, J. Liu, and H. Lu,	Need to apply on news video.	peculiar 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.
3	Learning to summarize web image and text mutually	Li, J. Ma, and S. Gao	to improve the Mutual-Summarization overall performance	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 Mutual-Summarization.
4	A low-rank approximation approach to learning joint embeddings of news stories and images for timeline summarization	W. Y. Wang, Y. Mehdad, D. R. Radev, and A. Stent,	Only work on summarizing synchronous multi-modal content.	The paper proposes a singular matrix factorization technique for extractive summarization, leveraging the success of collaborative filtering.
5	Multimedia summarization for social events in microblog stream	J. Bian, Y. Yang, H. Zhang, and T.-S. Chua,	Need to extend the pass-media framework for routinely detecting social occasions and retrieving related candidate microblogs. Personalisedmicroblog summarization based person profile.	The paper proposes a multimedia social occasion summarization framework to automatically generate visualized summaries from the microblog circulation of multiple media kinds.
6	A context-based word indexing model for document summarization	P. Goyal, L. Behera, and T. M. MCGinnity	Need to calculate the lexical association over a large corpus.	this paper, proposes a context sensitive document indexing model based on the Bernoulli model of randomness
7	Event summarization using tweets	D. Chakrabarti and K. Punera	The disadvantage that SUMMHMM has to account for tweet words that only occur in some of the events, but not in others.	The problem of summarizing event-tweets and give a solution based on learning the underlying hidden state representation of the event via Hidden Markov
8	Learning to summarize web image and text mutually	P. Li, J. Ma, and S. Gao	Need to improve the Mutual-Summarization overall performance.	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 Mutual-Summarization.

9	Summarization of personal photologs using multidimensional content and context	. P. Sinha, S. Mehrotra, and R. Jain	Computation expensive.	The proposed metrics which will evaluate the photo summaries based on their representation of the larger corpus and the ability to satisfy user’s information needs
10	Automatic Summarization of Lecture Slides for Enhanced Student Preview –Technical Report and User Study	Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, and Hiroaki Ogata	Not produce the good result for students because of span attention and it takes more time.	In this paper, a novel method of summarization is provided in order to increase student’s understanding of lecture slides to enhance preview .

III. OPEN ISSUES

Text summarization is performed for the purposes of saving users time by reducing the amount of content to read. 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. Some of the issues are given below:

- The lack of effective summarization mechanism.
- Users’ needs to deal with information related drawbacks for eg. Incomplete, irrelevant information due to services provided by existing trending topics.

- It makes strenuous for users to extract the essence of a topic.

IV. SYSTEM OVERVIEW

Traditional documents that contain only textual objects, microblogs constitute of multiple media types, such as image and text. In this paper proposes a offbeat approach to summarize multimedia microblogs related to trending topics. For text summarization, specifies three criteria, namely coverage, significance and diversity to measure the summarization quality. The Fig. 1 shows the architecture of multimedia microblog summarization system. 1) first step is to generate a textual and visual summaries and 2) second step is to generate aggregated summary using output of first step.

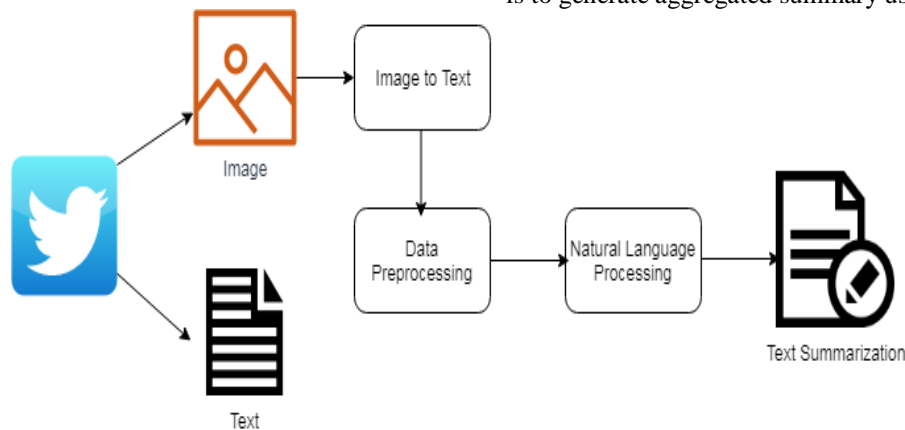


Fig.1: System Architecture

V. CONCLUSION

In this paper, proposes a multimedia summarization method for microblog to automatically generate visualized summaries for trending topics. A summarizer is elaborated to generate both textual and visual summaries. Summarizer will arrange the microblogs into structured subtopics. Also generates high quality textual summary at subtopic level.

VI. 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. 23rd 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 tweets,” 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 low-rank 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.
- [11]. Jingwen Bian, Yang yang, Tat seng Chua “Multimedia summarization for trending topics in microblogs,” 2017
- [12]. Shiva Twinandillaa , Satriyo Adhya , Bayu Surarsob , Retno Kusumaningruma, “Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences,” 2018
- [13]. 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...