International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 14th July 2017. Link :
http://www.ijiemr.org/downloads.php?vol=Volume-6&issue=ISSUE-5

Title: FB Microblog Filtering Method,The Microblog Content, Consisting.

Volume 06, Issue 05, Page No: 1800 – 1805.

Paper Authors

***UPPALAPATI VAHINI, N. JAGAJEEVAN.**

* Dept of CSE, Chalapathi Institute of Engineering And Technology.

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

FB MICROBLOG FILTERING METHOD,THE MICROBLOG CONTENT, CONSISTING

***UPPALAPATI VAHINI,** N. JAGAJEEVAN**

*PG Scholar, Dept of CSE, Chalapathi Institute of Engineering and Technology.

**Associate Professor, Dept of CSE, Chalapathi Institute of Engineering and Technology.

garlapati.vahini@gmail.com jeevan1024@gmail.com

ABSTRACT

In recent years, we have witnessed the boom of social media platforms, through which people have been generating a lot of social media data. This data touches almost every aspect of life and may have significant societal and marketing values for a variety of corporations and organizations. Thus, the development of effective techniques for gathering and analyzing social media content has attracted much research attention. As social media data tends to be heterogeneous, conversational and fast evolving in content, a recent work reported a multi-faceted approach to gather comprehensive brand-related data by crawling data using evolving keywords, key users, similar image content and known locations. Although such approach has been found to be effective in gathering representative data, it also brings in a lot of noise. This paper aims to develop an accurate classifier to filter out noise by taking into account the multimedia content and social nature of brand related data. In particular, we develop a microblog filtering method based on a discriminative social-aware multiview embedding. Besides the conventional content-based features, such as textual, low-level visual features, and high-level visual semantic features, that form the three key views of microblogs, we also incorporate the brand and social relations among the microblogs to learn a discriminative and social-aware embedding. With such a learned embedding, an off-the-shelf classifier, such as SVM, can then be trained and applied to microblog filtering. We verify the efficacy of our method on noise filtering in the brand data gathering task on the Brand-Social-Net dataset. Our approach is able to achieve significantly better filtering performance and improve the quality of brand data gathering.

Index Terms—Microblog filtering, Multiview embedding, Social graph.

I. INTRODUCTION

THe popularity of microblogging platforms, such as Twitter1 and Sina Weibo2 , has encouraged users to generate and share a huge amount of social media data known as user-generated contents (UGCs). These UGCs offer real-time information resources for a wide range of topics and are beneficial to a broad range of users and applications [21], [30]. Therefore, extensive research efforts have been dedicated to social media analysis, such as social event summarization [5], social network analysis [3], [9], [10], social TV [16], and sensing the topics [1]. Among the social contents, brand related microblogs, which

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

spread much faster than traditional media content, have significant marketing values for enterprises and governmental organizations [12], [32]. For example, positive or negative comments of a brand may impact the decision of users on whether to purchase the product, especially when the comments come from their friends. As another example, sensitive news or comments can often spread very fast across the entire social network and become viral and uncontrollable. As a result, it is essential for companies and organizations to know what people are talking about them in social network and be able to take preventive actions if necessary.

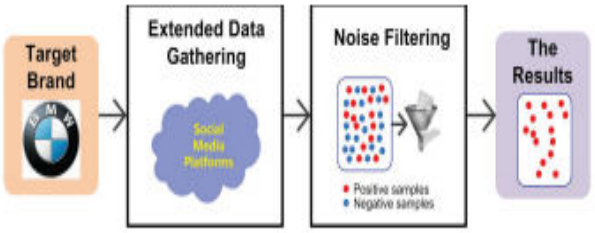

Fig. 1. A typical framework for brand data gathering.

In social media gathering, brand data gathering is an important and difficult task, which has attracted much research attention ([11], [20], [33]) in recent years. Figure 1 shows an example framework for this task, which consists of two main stages: extended data gathering and noise filtering. The objective of extended data gathering is to gather a comprehensive set of brand related data from live social media streams using a variety of techniques with the aim to improve data recall; while the goal of the noise filtering step is to remove the noise from the gathered data to achieve high precision. The overall task is challenging due to several factors. First, the short and conversational nature of microblogs leads to rapid changes in content and vocabularies in the related posts. Under such circumstances, the use of a fixed set of keywords is not adequate to track representative data and analyze social exposure for the target brand. Second, the content of microblogs tends to be heterogeneous. For example, it often consists of text, images, and/or videos. Our recent study [11] shows that about 40% of microblogs contain visual content, while only a small portion of such microblogs (about 30%) contains meaningful text annotations. It is not easy to identify the category of the microblog using just text or image content individually. Considering the rapid changed discussion background in microblog platform, similar content may convey different meanings when the circumstance varies. Concern ing all these properties of microblogs, microblog filtering is much challenger than traditional image classification. Figure 2 shows some example microblogs containing text content, visual content only, and both of the text content and visual content together.

Fig. 2. Example microblogs. (a) A microblog containing text only. (b) A microblog containing images only. (c) A microblog containing both text and images.

This phenomenon increases the difficulties in microblog data analysis, in which using text analysis alone is inadequate. Thus, a multi-faceted approach is needed for social media processing to overcome the limitation of textual keyword-based methods. Third, the social media data is often accompanied with social information, such as user friendship relations and geographic network, which may contain valuable information for brand data gathering and analysis. The multimodal data in social media has also attracted research attention from various of fields. In [22], Ntalianis and Doulamis proposed to jointly employ the multimedia content for human life summarization, where the content analysis is comprised of visual content and other associated metadata, such as date, events, likes, and comments. Targeting the task of sensing trending topics in Twitter, Aiello et al. [2] proposed to explore the temporal distributions of concepts to deal with heterogeneous streams.

For data gathering, [11] is the first attempt to use a multifaceted approach to collect brand-related microblogs in social media streams. In this method, a set of seeds are selected based upon the text-based search and logo detection. After that, the text, social context, and visual content of the seeds are used together to collect more microblogs. In this way, a comprehensive set of relevant microblogs can be obtained in comparison to most existing approaches that rely on text only. It is noted that the improvement on data coverage comes at the cost of bringing in a lot of noise. Therefore, an effective noise filtering step is needed which is the focus of this work. In this paper, we propose a microblog filtering method, which is able to filter out noisy data from the gathered microblogs for a given brand. The idea behind our brand filtering. framework is to map the social media data, microblogs in our case, into a latent subspace that is not only discriminative with respect to the target brand but also consistent with various types of information, including content features, such as textual and visual content, and social features, such as user relationship and spatial-temporal relatedness. Specifically, we propose a multiview embedding approach to jointly model the textual content and visual content represented using the lowlevel visual features and high-level visual semantic features. We further utilize the brand labels and social features as regularizers in order to take into account both brand and social relations. With the learned embedding in the latent space, an SVM classifier is trained and used for microblog filtering. Following the brand data gathering procedure in [11], we evaluate the proposed microblog filtering method on the Brand-Social-Net dataset. The evaluation is conducted on 3 million microblogs from Sina Weibo with 100 brands. We compare our method with several existing filtering methods and the experimental results demonstrate that our method is able to achieve significantly better performance. The work in [11] targets on gathering data in social media streams, where a multi-faceted approach is employed to investigate different clues in microblog platforms. Different from [11], this work mainly focuses on filtering out the noise data in microblog platform. [11] aims to improve the coverage of data gathering, while the current work is for the precision of data gathering.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Therefore, this work can be used as the next procedure after data gathering [11].

3. EXISTING SYSTEM

Relevant microblogs can be obtained in comparison to most existing approaches that rely on text only. It is noted that the improvement on data coverage comes at the cost of bringing in a lot of noise.

Therefore, an effective noise filtering step is needed which is the focus of this work. Compare our method with several existing filtering methods and the experimental results demonstrate that our method is able to achieve significantly better performance.

4. PROPOSED SYSTEM

Proposed to jointly employ the multimedia content for human life summarization, where the content analysis is comprised of visual content and other associated metadata, such as date, events, likes, and comments. Targeting the task of sensing trending topics in Twitter, Aiello et al. Proposed to explore the temporal distributions of concepts to deal with heterogeneous streams.

We evaluate the proposed microblog filtering method on the Brand-Social-Net dataset. Proposed microblog filtering method named discriminative social-aware multiview embedding. A text-based hashtag recommendation approach, where the text content was used to investigate the correlation between the microblog and the hashtags

6. ALGORITHM:

Polynomial Algorithm

Polynomial algorithm that is guaranteed to terminate within a number of steps which is a polynomial function of the size of the problem. See also computational complexity, exponential time, nondeterministic polynomial-time.

7. CONCLUSION

In this paper, we proposed a microblog filtering method, which can be used as the noise filtering step for the brand data gathering task. The key component of our method is a discriminative social-aware multiview embedding approach, which maps the microblog content, consisting of three (or more) views, into a latent space, while taking into account the brand information and social relations of microblogs.

We also discovered one interesting property of social information from the experiments that it makes more impact to microblog filtering for brands that have influential users with large social connections and followings. It is noted that data labeling requires a lot of manual work. Although we have fully annotated all the data for evaluation in this work, it is not affordable to label all the coming data. Confronting the changing circumstances in microblog platform, one possible solution is to actively keep monitoring the related data which can be used to automatically update the training dataset without additional human annotations.

There are still several open issues on microblog filtering. First, learning the brand-related visual context is an important issue, and the ability to extract target object information from the images and discover the associated high-level semantic feature may greatly

improve the microblog filtering results. Second, the deep structure of social context such as conversational graph structure of the microblogs should be exploited to boost the precision and recall of brand data prediction. Last but not least, exploring the multiple-topic information of microblogs via topic models may provide more semantic level information, and can be incorporated to enhance our multiview embedding framework. For our current work, there are also limitations. For the brand-similarity grand and the social-similarity graph, they are with high computational complexity. For our task, our objective is to filter the microblogs for a given target, such as a brand. The current framework is able to handle tens or hundreds of thousands microblogs each time but has the limitation on handling more data, such as 1 million data, which is a limitation of our work.

8. REFERENCES

[1] Luca Maria Aiello, Georgios Petkos, Christian Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Goker, Ioannis Kompatsiaris, and Aldo Jaimes. Sensing trending topics in twitter. IEEE Transactions on Multimedia, 15(6):1268–1282, 2013.

[2] Luca Maria Aiello, Georgios Petkos, Christian Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Goker, Ioannis Kompatsiaris, and Aldo Jaimes. Sensing trending topics in twitter. IEEE Transactions on Multimedia, 15(6):1268–1282, 2013.

[3] Santhanakrishnan Anand, KP Subbalakshmi, and Rajarathnam Chandramouli. A quantitative model and analysis of information confusion in social networks. IEEE Transactions on Multimedia, 15(1):207–223, 2013.

[4] Rabia Batool, Asad Masood Khattak, Jahanzeb Maqbool, and Sungyoung Lee. Precise tweet classification and sentiment analysis. In Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on, pages 461–466. IEEE, 2013.

[5] Jingwen Bian, Yang Yang, Hanwang Zhang, and Tat-Seng Chua. Multimedia summarization for social events in microblog stream. IEEE Transactions on Multimedia, 17(2):216–228, 2015.

[6] Matthew B. Blaschko, Christoph H. Lampert, and Arthur Gretton. Semi-supervised Laplacian regularization of kernel canonical correlation analysis. In Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, 2008.

[7] Chun Chen, Feng Li, Beng Chin Ooi, and Sai Wu. Ti: an efficient indexing mechanism for real-time search on tweets. In Proceedings of Iinternational Conference on Management of Data, 2011.

[8] Ning Chen, Jun Zhu, Fuchun Sun, and Eric P Xing. Large-margin predictive latent subspace learning for multiview data analysis. IEEE TPAMI, 34(12):2365–2378, 2012.

[9] Quan Fang, Jitao Sang, Changsheng Xu, and M Shamim Hossain. Relational user attribute inference in social media. IEEE Transactions on Multimedia, 17(7):1031, 2015.

[10] Quan Fang, Jitao Sang, Changsheng Xu, and Yong Rui. Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. IEEE Transactions on Multimedia, 16(3):796–812, 2014.

[11] Yue Gao, Fanglin Wang, Huabo Luan, and Tat-Seng Chua. Brand data gathering from live social media streams. In Proceedings of the 4th ACM Conference on Multimedia Retrieval, 2014.
[12] Chunmei Gu and Shanshan Wang. Empirical study on social media marketing based on sina microblog. In International Conference on Business Computing and Global Informatization, pages 537–540, 2012.
[13] H. Hotelling. Relations between two sets of variates. Biometrika, 28(3/4), 1936.
[14] Xia Hu, Nan Sun, Chao Zhang, and Tat-Seng Chua. Exploiting internal and external semantics for the clustering of short texts using world knowledge. In Proceedings of the 18th ACM conference on Information and knowledge management, pages 919–928. ACM, 2009.
[15] Jeff Huang, Katherine M Thornton, and Efthimis N Efthimiadis. Conversational tagging in twitter. In Proceedings of the 21st ACM conference on Hypertext and hypermedia, pages 173–178. ACM, 2010.
[16] Yichao Jin, Yonggang Wen, Han Hu, and Marie-Jose Montpetit. Reducing operational costs in cloud social tv: an opportunity for cloud cloning. IEEE Transactions on Multimedia, 16(6):1739–1751, 2014.
[17] David G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.
[18] Kamran Massoudi, Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. Incorporating query expansion and quality indicators in searching microblog posts. Advances in Information Retrieval, pages 362–367, 2011.
[19] Gilad Mishne. Using blog properties to improve retrieval. In Proceedings of International Conference on Weblogs and Social Media, 2007.
[20] Rinkesh Nagmoti, Ankur Teredesai, Martine De Cock, et al. Ranking approaches for microblog search. In Proceedings of International Conference on Web Intelligence and Intelligent Agent Technology, 2010.
[21] Nasir Naveed, Thomas Gottron, Jer´ ome Kunegis, and Arifah Che Alhadi. ˆ Searching microblogs: coping with sparsity and document quality. In Proceedings of ACM International Conference on Information and Knowledge Management, pages 183–188, 2011.
[22] Klimis Ntalianis and Nikolaos Doulamis. An automatic eventcomplementing human life summarization scheme based on a social computing method over social media content. Multimedia Tools and Applications, pages 1–27, 2015.
[23] Brendan O'Connor, Michel Krieger, and David Ahn. Tweetmotif: Exploratory search and topic summarization for twitter. In Proceedings of International Conference on Weblogs and Social Media, 2010.
[24] Timo Ojala, Matti Pietikainen, and David Harwood. A comparative study ¨ of texture measures with classification based on featured distributions. Pattern Recognition, 29(1):51–59, 1996.
[25] Novi Quadrianto and Christoph H. Lampert. Learning multi-view neighborhood preserving projections. In ICML, 2011.