J. Vis. Commun. Image R. 60 (2019) 312–318

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Correlation identification in multimodal weibo via back propagation neural network with genetic algorithm

Maofu Liu [a], Weili Guan [b], Jie Yan [a], Huijun Hu [a,*]

[a] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[b] Hewlett Packard Enterprise Singapore, Singapore 117417, Singapore

ARTICLE INFO

Article history:
Received 13 January 2019
Revised 29 January 2019
Accepted 14 February 2019
Available online 22 February 2019

Keywords:
Multimodal weibo
Correlation classification model
Back propagation
Genetic algorithm
GA-BP

ABSTRACT

The rapid development of social media services has spawned abundant user generated contents (UGC), such as Sina Weibo, which is one of the biggest Chinese microblogging platforms. In order to enhance the quality and popularity of the posted weibo (the microblog), Weibo users usually embed some social information and images or micro-videos, namely the multimodal weibo, and we assume that there is a close correlation among the multimodal weibo data, especially between the visual data (image/micro-video) and its corresponding text, for a multimodal weibo of high quality. Hence, we try to evaluate the quality of multimodal weibo via analyzing the correlation in the multimodal weibo. This paper constructs the classification model based on back propagation (BP) neural network with genetic algorithm (GA), to automatically identify the correlation within the multimodal weibo, and investigates three kinds of features from multimodal weibo to uncover their contributions to the correlation. The experimental results verify the superiority of the GA-BP based classification model over the traditional BP neural network.

1. Introduction

With the rapid development of social networking services (SNS), especially Twitter, Sina Weibo and Tencent Weibo (both are famous Chinese social media platforms), the user generated content (UGC) has become the most commonly used social media around the world. With over 445 million monthly active users as of Q3 2018, Sina Weibo (Weibo for short) enjoys numerous weibos daily. Via mining and analyzing the useful information and knowledge contained within these weibos, we are able to predict users behaviors and settle some social problems. For example, we can analyze the Weibo users attitudes towards a social event and make the right response to prevent some aggressive behaviors. To achieve these goals, collecting the high-quality weibos is the first task.

With the improved bandwidth and the emergence of smartphones, Weibo users have the ability to post text with social information, image and micro-video, namely the multimodal weibo in this paper. It has become the mainstream of social media, and the microblogging services have then been flooded by the multimodal weibos. In this paper, we assume that, in the multimodal weibo of high quality, the text is closely semantically correlated with the embedded social information, image and micro-video.

The image usually refers to the static visual picture, and meanwhile, the video is the display of dynamic visual media, more specifically, the video represents a continuous series of interrelated images. So, in this paper, we construct the classification model to automatically identify the correlation between the image and its corresponding text in multimodal weibo, and the correlation between the micro-video and its corresponding text can be settled in the same manner.

Generally speaking, the image and its corresponding text posted by the traditional media are highly correlated. The top two in Fig. 1 are examples from Netease[1] and the images are relevant to their texts. In the first image-text pair, the text says "老 爷 爷" (an old man) and "乌 龟" (tortoise), and the image depicts an old man walking with a tortoise. In the second image-text pair, the text contains the word "京 剧" (Beijing opera), and the image is the photo of an opera performer. The most possible reason for the information matching is that the image is filtered and edited by the editors of

This paper has been recommended for acceptance by Liqiang Nie.
* Corresponding author.
E-mail addresses: liumaofu@wust.edu.cn (M. Liu), huhuijun@wust.edu.cn (H. Hu).

[1] http://yuetu.163.com.

https://doi.org/10.1016/j.jvcir.2019.02.015
1047-3203/

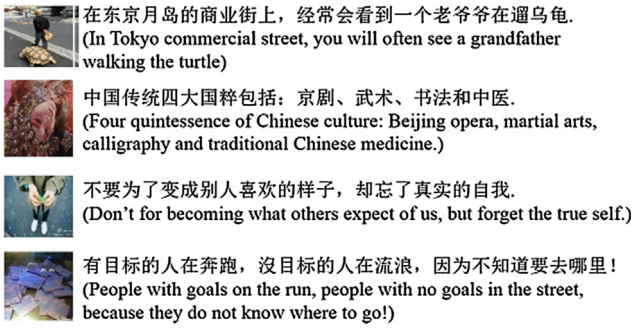

Fig. 1. The top two are examples from Netease and the bottom two are from Weibo.

the website strictly by tightening the correlation between the image and its corresponding text. However, the microblogging website is a free social network platform and the registered users post weibo freely. This leads to the random use of images and makes it difficult to guarantee the correlation between the image and its corresponding text in multimodal weibo, for example, the bottom two from Weibo[2] in Fig. 1.

In this paper, we firstly collect pairs of the text and its embedded image from Weibo to form a multimodal weibo dataset. And then, we formulate the annotation specification to manually label a small part of the multimodal weibo dataset with tag 'correlated' or 'uncorrelated'. Finally, we construct a classification model to automatically identify the correlation between the image and its corresponding text in each multimodal weibo via features and compare the results of the BP model and the GA-BP model.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the features of multimodal weibo and the correlation classification model. Section 4 then presents the dataset and experiments. Finally, we conclude this paper in Section 5.

2. Related work

As one of the biggest Chinese microblogging platforms, Weibo is worth studying. However, the previous researches mainly focus on the text of weibo. With the rapid popularization of smartphones, it is easier for people to post a message with images or mciro-videos. As early as in 2011, Asur, Yu, and Huberman [1] reported that 56.43% of all microblog posts in Weibo were multimodal weibo. Zhao et al. [2] discovered that the multimodal weibo was forwarded more frequently and survived longer than the text-only posts. Therefore, thus far, among all the weibos, the multimodal weibo is the most common form.

The multimodal data, such as image or micro-video, conveys more information compared with the mere text. Analyzing the image or micro-video when ranking social media meets the general trend. Hence, some preliminary studies have described the multimodal social media. Chen et al. [3] studied a large corpus of image tweets. Liu et al. [4] constructed end-to-end deep learning model, packing three parallel long short-term memory (LSTM) models to capture the sequential structures and a convolutional neural network to learn the sparse concept-level representations of micro-videos, and they also applied the model to the micro-video categorization. Song et al. [5,6] developed a robust multiple social network learning model and applied it to predict the users' volunteerism tendency from user-generated contents collected from multiple social networks based on a conceptual volunteering decision model. Nie et al. [7] enriched the textual data with the appropriate multimodal data in community question answering, and they [8,9] also harvested the image search with multimodal data.

Uncovering the relationship between image/micro-video and text is the main research direction. Gao et al. [10] proposed an approach to simultaneously utilizing both visual and textual information to estimate the *relevance of user tagged images*. Liu et al. [11] built a structure-guided multi-modal dictionary learning model to learn the concept-level micro-video representation by jointly considering their venue structure and *modality relatedness*. Zhang et al. [12] proposed a visual translation embedding network (VTransE) for *visual relation* detection in multimodal social media. Mohammadi-Nejad, Hossein-Zadeh, and Soltanian-Zadeh [13] put forward a structured and sparse CCA (ssCCA) technique to discover the true *association* between the multimodal datasets. Chen, He, and Kan [14] found that the traditional image features were far from adequate in interpreting the necessary semantics latent in image tweets, and enriched the representation of images in image tweets by considering their *social context*.

In general, there are two kinds of methods in correlation identification in multimodal data, especially between image and text. One is to map the two types of data, i.e. visual and textual, into a common space [15]. The other tries to investigate the combination between the visual and textual feature sets. Hotelling et al. [16] proposed canonical correlation analysis (CCA) to find the linear combinations between two sets of variables, and Rasiwasia et al. [17] and Sharma et al. [18] adopted this kind of method. Some researchers directly combined the bimodal data, i.e. image and text, and applied a model to train them. Among these models, latent Dirichlet allocation (LDA) [19] is a commonly used one, which is a topic model for discovering the abstract topics occurring in a collection of documents. Chen et al. [20] developed a visual-emotional LDA (VELDA) model to capture the image-text correlation from multiple perspectives, including namely, visual and emotional, and they demonstrated that VELDA significantly outperforms the existing methods on cross-modality image retrieval.

The weibo has some unique characteristics, such as the maximum length of 140 words and the non-standard syntax. The first example[3] in Fig. 2 uses the word "北 鼻 (baby)", but these is no such phrase in the Chinese vocabulary, for " 北 鼻 (baby)" is an onomatopoeia of baby in Chinese. In the second example[4] in Fig. 2, both the text and image contain the word "白 鹿 原 (White Deer Plain)", but it is difficult to associate the image with text. Fortunately, some specific features contained in the plain text microblogs can be utilized for analysis. Sriram et al. [21] applied a naive Bayes learning model with a set of eight features, such as the author ID, the presence of shortened words, '@' signs, opinionated words, emphasized words, currency and percentage signs and time phrases, to perform hard classification into five categories. For the multimodal weibo, in this paper, we employed three kinds of features, i.e. the social, linguistic and similarity, as the inputs of GA-BP based classification model. The genetic algorithm is used to select the weight and threshold value of each level.

3. Correlation classification model

In this section, a classification model based on back propagation neural network with genetic algorithm is proposed to identify semantic correlation in multimodal weibo, incorporating with the three kinds of features, i.e. social, linguistic, and similarity.

[2] http://weibo.com.

[3] https://m.weibo.cn/detail/4330563718000786.

[4] https://m.weibo.cn/p/2304135873769280.