International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

ICNSCET20- International Conference on New Scientific Creations in Engineering and Technology

Multimodal Sentimental Analysis for Tweets

T.Nandhini[1],S.Nivetha[2],R.Pavithra[3], Dr.S.T.Veena

Computer Science and Engineering
Kamaraj College of Engineering and Technology

Abstract—***Social media users are increasingly using both images and text to express their opinions and share their experiences, instead of only using text in the conventional social media. Consequently, the conventional text-based sentiment analysis has evolved into more complicated studies of multimodal sentiment analysis. To tackle the challenge of how to effectively exploit the information from both visual content and textual content from image-text posts. The proposed approach explores the correlation between the image and the text, followed by a multimodal sentiment analysis method. To be more specific, visual features extracted by the convolutional neural network algorithm is used to represent visual concepts, to develop a machine learning sentiment analysis approach. Extensive experiments are conducted to demonstrate the superior performance of the proposed approach. The reviews are classified as positive or negative sentiments by fusion of results from each mode.***

Keywords—Multimodal sentiment analysis, Textual sentiment, Visual sentiment, Social media, Hashtag

I. INTRODUCTION

With the rapid growth of the social media, users tend to share their opinions in social media platforms such as Twitter, Facebook and SinaWeibo. These user-generated content is moving toward a diversification of content and formats, where people tend to post text embedded images, namely *image-text post.* The posts are more informative since they contain visual content in addition to texts, unlike the conventional text-only posts.

Sentiment analysis aims to automatically uncover the underlying attitude of the posts. Due to the rich sentiment cues that can be found in images, sentiment analysis of visual content can contribute more towards extracting user sentiments and understand user behavior, stock market forecasting and voting for politicians.

The major challenge of sentiment analysis for social media lies in effective feature extraction and representation for both text content and visual content. This challenge has drawn attention in the field of computer vision and especially retrieval and emotional semantic image retrieval, which applies computer vision technology to eliminate the affective gap between low-level features and the emotional content of an image. In these conventional approaches, low-level visual features, such as color histogram, are directly used into sentiment analysis with textual features. This has caused a great loss of emotional information from image, and consequently, there still exists a great semantic gap between low-level features and emotional content in the images.

To tackle the challenge of analyzing both text content and visual content in image-text posts, a text-image consistency driven multimodal sentiment analysis approach is proposed in this paper.

The proposed approach is motivated by these two observations and this contributions of this paper are two-fold. First, to effectively exploit the information from both visual content and textual content from image-text posts, the proposed approach explores the correlation between the image and the text, where the mid-level visual features extracted by the conventional Senti-Bank approach are used to represent visual concepts, with the integration of other features, including textual, visual and social features. Second, the proposed approach performs a multimodal adaptive sentiment analysis by fusion of results from each mode.

II RELATED WORKS

Sentiment analysis, sometimes known as opinion mining, aims to judge emotional orientation (e.g., positive, negative) based on user-generated content. Traditional sentiment analysis concentrates on textual sentiment analysis. However, research on visual sentiment analysis is relatively much less done. In recent years, much research has been done on visual sentiment analysis due to the exponential growth in Internet use. In this section, we will briefly discuss the related work in areas of textual sentiment analysis, visual sentiment analysis and multimodal sentiment analysis in social media.

2.1. Textual sentiment analysis

A brief review on existing textual sentiment analysis approaches is provided in this section. In the existing body of research, most of the sentiment information come from Web, blog and twitters, where the text posts are studied for sentiment analysis. In most of these works, textual features are directly extracted from original texts, and then used in sentiment analysis. To further reduce the influence of noise and improve the precision of classification, text preprocessing is needed in textual sentiment analysis.

These special treatments of preprocessing decrease the accuracy of sentiment analysis; therefore, when identifying social data especially subjective sentences, most methods enter emotional words, assisting with various vocabulary and word frequency information into the machine learning classifiers. Text data can be collected from Twitter using twitter dataset. It can be preprocessed and the Root words can be identified by Data Preprocessing and the data can be classified as positive or negative by data filtering. The Probability distribution will be the output for the text tweets.

However, only relying on sentiment words may also cause a large deviation, especially for such comprehensive sentences as double negative sentences. For example, the “bad words“ of many negative emotions in horror movies unnecessarily denote the negative emotions of the reviewers. To extract deeper level semantic features, Liu et al. studied the characteristics of Weibo text features, including text length, noun density, verb density and named entity density four-dimensional text features, which contribute to association analysis and sentiment analysis.

2.2. Visual sentiment analysis

A brief review on existing visual sentiment analysis approaches is provided in this section. Initially,the study of visual sentiment analysis was based on image aesthetic quality assessment and emotional semantic image retrieval proposed to classify the aesthetic quality of images by using the support vector machine classifier with common image features including *Bag-Of-Visual Words* (BOVW) proposed the feature representation approach by combining common low-level features, aesthetic features and mid-level features. Hayashi and Hagiwara (1998) proposed a highaccuracy rate based on the back-propagation neural network to construct the mapping relationship between visual features and impression keywords. These methods partly eliminate the gap between low-level features and the emotional content, but how to
represent the features of an image is still a problem in visual sentiment analysis.

Image data can be collected from Twitter using Twitter image dataset. It can be resized and can be classified as positive or negative from Convolutional neural network. The accuracy will be the output for the image tweets.

SentiBank in different languages and proposed a large-scale multilingual sentiment concept ontology.

2.3. Multimodal sentiment analysis

Sentiment recognition expressed in social media from multimodal signals, including visual, audio and textual information has been studied and multimodal sentiment analysis is an emerging area.

Social media user generated text is always posted with an accompanying image or short video, and this adds one more channel of information in user sentiment expression extracted textual posts with related images extracted from SinaWeibo and conducted sentiment analysis by combing the prediction results of using n-gram textual features and mid-level
features proposed extracting deep semantic features of images by identifying objects and scenes as salient, and then fusing features, with sentiments.

The reviews are classified as positive or negative sentiments by fusion of results from each mode.

III.PROPOSED SYSTEM

3. Proposed image-text consistency driven multimodal sentiment analysis approach

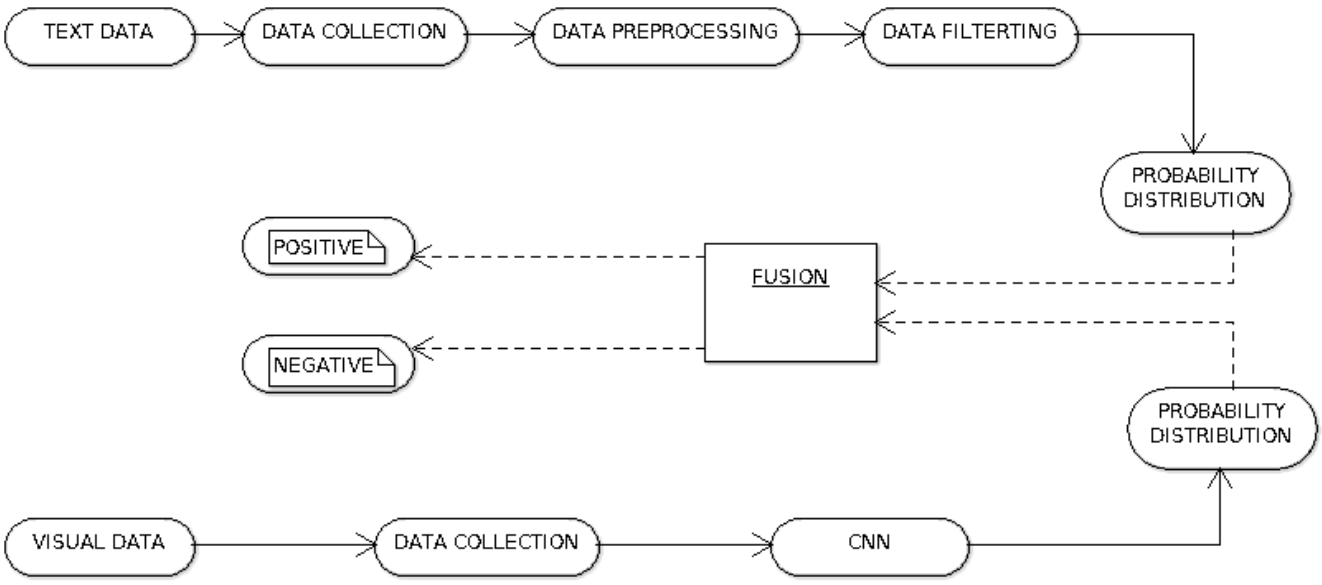

Fig 3

In this section, the proposed image-text consistency driven multimodal sentiment analysis approach is presented.

The proposed approach, as illustrated in Fig. 3, consists of four critical components, which are briefly described as follows.

- ***Data Collection:***

In this collection stage, datas from textual content and visual content are collected by using datasets. The dataset used to collect the textual datas and visual datas are 'Sentibank dataset'

- ***Data Preprocessing:***

In the preprocessing stage, some natural language processing methods e.g. stop-word removal, tokenization, stemming are used to process text data.

- ***Data Filteration***:

Feature extraction: Three main types of features, i.e. textual feature, visual feature, and image-text similarity, are extracted from the preprocessed image-text post.

- ***Fusion***:

A machine learning model is trained using three main types of features from image-textposts to make a binary decision on whether the image content and the text content are consistent with each other.

- ***Sentiment classifier:*** Performs sentiment classification on the input image-text posts and results will be displayed as positive or negative.

3.1.Preprocessing

Data preprocessing is a critical step in the proposed approach, especially for user generated data from social media platforms,
where the data is raw and unstructured. The proposed approach consists the following steps in preprocessing.

Information extraction: Raw dataset is composed of different kinds of information, including time, source, image width, etc. Based on the system goal and common sense, messages titles and descriptions are selected as text data for further analysis.

Special symbol removal: In social media platforms, users often post their messages and direct them to others or comments on others' posts by using symbol @. The information after this symbol is relative to the privacy of users and useless in sentiment analysis, so the words after @ need to be deleted.

Stop word removal: In natural language processing, some words, also known as "stop words" are filtered out. As such, common stop words were deleted.

Tokenization: Tokenization is the process of breaking a text corpus most commonly into words, but also into phrases and
meaningful elements, all of which are known as tokens. The tokens become the basic units for further text processing.

Stemming: Stemming and lemmatization aim to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. For example, the word *love* in the data set considering the other posts may appear similar *loves*, *loving*, and *lovable*, we can put them all into root love, to reduce the amount of data, and the only word, also won't lose a lot of information.

3.2Feature extraction

3.2.1. Texture feature extraction

Text feature extraction plays a crucial role in text analysis, directly influencing the accuracy of sentiment classification. *Word2vec* has garnered a lot of interest in the text mining community.The model derives a supervised learning task from the corpus itself using either the continuous bag-of-words model or continuous skip-gram model. On the other hand, it is considered unsupervised in the sense that one can provide any large corpus of one's own choice.

In this paper, the public pre-trained words and phrase vectors *Twitter datset- 300* is used in the proposed approach. It contains 300-dimensional vectors for 3 million words and phrases. It is sufficient enough to contain all our twitter data.

set corpus. The vector for each word is a semantic description of how that word is used in context, so two words that are used similarly in text will get similar vector representations. It has been shown that the word vectors capture many linguistic regularities,
for example vector operations vector('Paris') - vector('France') + vector('Italy') result in a vector that is very close to vector('Rome'), and vector ('king') - vector('man')+vector('woman') is close to vector('queen'). Once mapping words into vector space, one can then use vector math to find words that have similar semantics. After changing each word into 300-dimensional vectors, Twitter text can be seen as a representation of a word serialization. To avoid dimensional disasters and the variant length of the input text, the proposed approach uses the word addition (Maas et al., 2011) method. In this way, we calculate the summation of each dimension to get 300-dimension vectors as a representation of a text, which can be expressed as
$[v ,v , v , ,v] [v ,v , ,v]/n$

3.2.2. Social feature extraction

The social features can reflect the social characteristics of image-text posts to some extent and generally can be found or derived from the text or its surrounding information. The number of likes, the number of comments, the number of forwards and topic could be considered as social features. Importantly, social media can benefit from knowing how influential a topic will be so that they can determine the amount of coverage they are willing to give to a specific news. A combination of these three measurements will help to
gauge the value of a topic:

Lifespan: to determine if the topic is time specific or long term and how long it actually lasted.

Emotion transition: determine whether the emotions evolved over time.

Reach: quantify how many different users got involved in the discussion.

As such, we consider topics only as social features. The proposed approach extracts topic from the image-text twitter, to represent the social information. The social feature extraction is represented as 300-dimensional vectors by word2vec too.

3.2.3. Visual feature extraction

For low-level image features, the proposed approach extracts the low-level image features and combine them. Together, they form the basic image features.
Middle-level features, *Adjective Noun Pair*s (ANPs) are used as mid-level feature representation in sentiment analysis bySentiBank Visual learning of adjectives is understandably difficult due to its abstract nature and high variability.
Therefore, we use adjective nouns combinations to be the main semantic concept elements of the image. The advantage of using ANPs, as compared to nouns or adjectives only, is the feasibility of turning a neutral noun into a

strong sentiment ANP. Such combined concepts also make the concepts more detectable, compared to adjectives only. The above described ANP structure shares certain similarity with the recent trend in computer vision and multimedia concept detection. To extract the high level semantic features from images, we use five pre-trained CNN-based models to extract the top 10 tags from images individually. These models are VGG16, VGG19 Inception V3 andResnet which are the state-of-the-art image classification models, and these models are pre-trained by image net which is a large visual database designed for use in visual object recognition research. Over 14 million images have been hand-annotated by ImageNet and to make sure that accuracy and diversity of tags, we combine the results from each model and also use the word2vec model into 300-dimensional vectors for further use.

3.2.3*Proposed image-text consistency classifier*

In the proposed image-text consistency classifier, the text features (basic text features and topic features) and image features (image basic features and image tag features), as well as and text-image similarity features are concatenated and incorporated into a support vector machine classifier to train a machine learning model (*Support Vector Machine* (SVM) is used in the proposed approach) to decide whether the image content and the text content is correlated to each other or not. The output is *yes* or *no*. If the output is *yes*, that means the text feature and image feature has correlation, then we put both of them into the sentiment classifier. By doing this, we can enhance sentiment prediction accuracy compared with only using text or image feature. If the output of SVM is *no*, this means that the correlation between text and image is not strong, and they may even have opposite sentiments.

3.4. Sentiment classifier

In the final sentiment prediction module of the proposed approach, two SVM-based models are trained for two different conditions: related image-text data and unrelated image-text data. First, for the related image-text data, we input four types of features (basic text feature, social feature, OCR feature from image and ANPs feature from image) from the training data into the SVM-based model and the four types of features from related testing data will be used to predict the sentiment in related image-text post. On the other hand, for the unrelated image-text data, we only put ANPs features from image of training data into another SVM-based model and ANPs feature from images of unrelated testing data will be used to predict the sentiment in unrelated posts.

IV. Experimental Results

4.1 Dataset description

The dataset used in this is paper is the Sentibank data of Visual Sentiment Ontology. It contains 603 images in total, covering a diverse set of over 21 topics, and there is a corresponding emotional value ground truth. Based on the datasets characteristics, six categories of labels are created, including

- *Noun* means both text and image shares the same object like *newphew*.
- *Name* means both text and image appeared the same name like *Obama* in words.
- *Word* means both text and image contains the same word, phrase or sentence, like *I am falling in love with you.*
- *Verb* means there is a verb in text and image shows a corresponding activity.
- *Scene* means the text contains a scene and in the image portraying a similar message as that of the text
- *YorN*means there is some relationship for the image with text or not. Multiple labeling can also occur in the dataset at the same time.

4.2. Performance metric

The following performance metric are used in experiments in this paper, including *precision, recall, f1-score, accuracy*. These four metric are defined as

$$Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN},$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall},$$

where *TP* is True Positive and *FP* is False Positive.

4.3. Implementation

The detailed implementation of the proposed approach is presented in this section. First, after checking each post in the original dataset of 603 posts, we find that 36 posts did not have text. So we only used the remaining 567 posts for the study. We randomly choose 400 samples as training data and the left 157 samples as test data in later model building and performance evaluation.
Next, from these 567 text data, we extracted 567 textual features that are descriptions written by users and 397 social features
that are represented topics in the post, since some posts don't have topics. After tokenization, removing stopwords and stemming, each textual feature and social feature was represented by several words respectively. As for these 567 image data, we extracted 4 low-level image features, including color histogram, GIST, LBP, BoW, and combined them as our image basic features, which are represented by 2000-dimensional vectors. Moreover, we used 5 pre-trained CNN-based models to extract the top 10 tags from images individually, such that each image has 50 tags. We also extracted middle-level visual features by conventional SentiBank method, which used 1200 ANPs. We got similarity coefficient by using 50 image tags and textual data. Then, we built an image and text consistency classifier using the SVM approach, based on social features, textual features, similarity coefficients, tag features and visual features. Finally, two SVM sentiment classifiers using RBF kernel were built. The first model is used for related text-image posts based on social features, textual features and ANPs features to the model, while the second model is used for unrelated text-image posts based on ANPs features only.

4.4. Experimental results

The proposed approach is evaluated with conventional approaches, the performance evaluation is presented in Table 2.
A more detailed performance comparison between the proposed approach with the state-of-the-art approach (Borth et al., 2013) is presented in Table 3.

As seen from these tables, the proposed approach achieves overall better performance than conventional approaches, since it is able to adaptively determine the correlation between visual information and the textual information of the image-text posts, and provide a better semantic feature for more accurate sentiment classification.

Table 2
The performance evaluation of various sentiment analysis approaches.

Approach	Precision	Recall	F-score	Accuracy
Visual model	0.76	0.72	0.74	0.68
Textual model	0.83	0.64	0.72	0.69
CCR (You, Luo et al., 2016)	0.85	0.76	0.80	0.80
SentiBank (Borth et al., 2013)	0.87	0.84	0.83	0.84
T-LSTM (You, Cao et al., 2016)	1.00	0.81	0.89	0.88
Proposed approach	0.88	0.88	0.88	0.87

Table 3
The performance comparison between the proposed approach and the state-of-the-art approach (Borth et al., 2013).

Category	Method	Precision	Recall	F-score
Positive	SentiBank (Borth et al., 2013)	**0.98**	0.67	0.80
Sentiment	Proposed approach	0.91	**0.83**	**0.87**
Negative	SentiBank (Borth et al., 2013)	0.76	**0.99**	0.86
Sentiment	Proposed approach	**0.86**	0.92	**0.89**

V . CONCLUSION

An image-text consistency driven multimodal sentiment analysis approach has been proposed in this paper for social media. The proposed approach exploits a image-text consistency approach to decide whether the image content and the text content are consistent with each other, and then adaptively further merge the textual features and the visual features used in the conventional SentiBank to provide more accurate sentiment analysis for image-text posts.

References

[1] ZiyuanZhaoa, HuiyingZhua, ZehaoXuea, Information Processing and Management, “An image-text consistency driven multimodal sentiment analysis approach for social media” (2019)

[2]. Agarwal, Xie,Vovsha,Rambow, Sentiment analysis of twitter data. In: Proc. ACL 2011 Workshop on Languages in Social Media, pp. 30–38 (2017)

[3] .Barbosa, L.Feng, Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of COLING, pp. 36–44 (2019)

[4]Gimpel,K.Schneider,N.O'Connor,B.Das,D.Mills,D.Eisenstein,Part-of-speech tagging for twitter: Annotation, features, and experiments. Tech. rep., DTIC Document (2018)

[5] .A.Bhayani, R.Huang, L.Twitter sentiment classification usingdistant supervision. CS224N Project Report, Stanford (2019)

[6].Guerra, P.Veloso, A. Meira, W.Almeida, From bias to opinion: A transfer-learning approach to real-time sentiment analysis. (2017)Google Scholar.

[7] Kouloumpis, E.Wilson, T.Moore, Twitter sentiment analysis: The good the bad and the omg! (2017)

[8] Liu, M., Zhang, L., Liu, Y., Hu, H., & Fang, W. (2017). Recognizing semantic correlation in image-text weibo via feature space mapping.*Computer Vision and Image Understanding*, *163*, 58–66.

[9] Zhao, S., Gao, Y., Ding, G., & Chua, T. (2018). Real-time multimedia social event detection in Microblog.*IEEE Transactions on Cybernetics*, *48*, 3218–3231.

[10] Zhao, S., Yao, H., Gao, Y., Ding, G., & Chua, T. S. (2016). Predicting personalized image emotion perceptions in social networks. *IEEE Transactions on Affective*

[11] Radha Krishnan, B., Vijayan, V., Parameshwaran Pillai, T. and Sathish, T., 2019. Influence of surface roughness in turning process—an analysis using artificial neural network. *Transactions of the Canadian Society for Mechanical Engineering*, 43(4), pp.509-514.

[12] Krishnan, B.R., Ramesh, M., Giridharan, R., Sanjeevi, R. and Srinivasan, D., Design and Analysis of Modified Idler in Drag Chain Conveyor. *International Journal of Mechanical Engineering and Technology*, 9(1), pp.378-387.

[13 Krishnan, B.R., Vijayan, V. and Senthilkumar, G., 2018. Performance analysis of surface roughness modelling using soft computing approaches. *Applied Mathematics & Information Sciences*, 12(6), pp.1209-1217.

[14] KRISHNAN, B.R. and PRASATH, K.A., 2013. Six Sigma concept and DMAIC implementation. *International Journal of Business, Management & Research (IJBMR)*, 3(2), pp.111-114.

