Visual Sentiment Analysis on Social Media Data

Harshala Bhoir[1], K. Jayamalini[2]

[1]Department of Computer Engineering, Shree. L. R. Tiwari College of Engineering, Mira Road(E), Thane, Maharashtra, India

[2]Assistant Professor, Department of Computer Engineering, Shree. L. R. Tiwari College of Engineering, Mira Road(E), Thane, Maharashtra, India

ABSTRACT

Visual sentiment analysis is the way to automatically recognize positive and negative emotions from images, videos, graphics, stickers etc. To estimate the polarity of the sentiment evoked by images in terms of positive or negative sentiment, most of the state-of-the-art works exploit the text associated to a social post provided by the user. However, such textual data is typically noisy due to the subjectivity of the user which usually includes text useful to maximize the diffusion of the social post. Proposed system will extract and employ an Objective Text description of images automatically extracted from the visual content rather than the classic Subjective Text provided by the user. The proposed System will extract three views visual view, subjective text view and objective text view of social media image and will give sentiment polarity positive, negative or neutral based on hypothesis table.

Keywords : Sentiment analysis, CNN, ResNet , GoogLNet, VGG Net, Feature Extraction, Subjective ,Objective

I. INTRODUCTION

1.1 Visual Sentiment Analysis

Sentiment analysis is the automated process of understanding an opinion about a given subject from written or spoken language. In a world where we generate 2.5 quintillion bytes of data every day, sentiment analysis has become a key tool for making sense of that data.[7].With the popularity of social networks and mobile devices, there is a huge volume of images and videos captured by users to record all kinds of activities in their lives every day and everywhere. For example, people may share their travel experiences, their opinions towards some events and soon. Automatically analysing the sentiment from these multimedia contents is demanded by many practical applications, such as smart advertising, targeted marketing and political voting forecasts .Compared with text based sentiment analysis which infers emotional signals from short textual description, visual contents, such as colour contrast and tone ,could provide more vivid clues to reveal the sentiment behind. Figure1 shows image examples from Twitter. Apparently, the images in the

upper row manifest positive sentiment, while those in the lower row deliver negative emotion.[9]

Sentiment	Image	
Positive		
Negative		

Figure 1 : Examples of Twitter images with positive or negative sentiment.

1.2 Problem statement

Visual Sentiment Analysis on Social Media Data by extracting and employing an Objective Text description of images automatically extracted from the visual content rather than the classic Subjective Text provided by the users and visual view of image based on hypothesis table.

1.3 Project Objective

To do Visual Sentiment Analysis on Social Media Data by extracting and employing an Objective Text description of images automatically extracted from the visual content rather than the classic Subjective Text provided by the users and visual view of image based on hypothesis table.

1.4 Project Idea

Social media users continuously post images together with their opinions and share their emotions. This trend has supported the growing of new application areas, such as semantic based image selection from crowd-sourced collections [1], Social Event Analysis [1] and Sentiment Analysis on Visual Contents [9]. Visual Sentiment Analysis aims to infer the sentiment evoked by images in terms of positive or negative polarity. Early methods in this field focused only on visual features or have employed text to define a sentiment ground truth. More recent approaches

combine visual and text features by exploiting well-known semantic and sentiment lexicons [1]

In this paper proposed system will exploit the text automatically extracted from images to build an embedding space where the correlation among visual and textual features is maximized. Several previous works define models which learn a joint representation over multimodal inputs (i.e., text, image, video, and audio) to perform Image Classification, Visual Sentiment Analysis [10], Image Retrieval, and Event Classification by exploiting social media contents.

In the proposed system the text associated to images is typically obtained by considering the meta-data provided by the user (e.g., image title, tags and description). It also describes images in an "objective" way by using scene understanding methods [2,4,6].Automatic extraction of text by the system from image is called as objective text. The "objective" emphasizing the fact that it is different to the"subjective" text written by the user for an image of a post.

Proposed system will consider ResNet , GoogLNet and VGG Net CNN architecture to extract text from image. Also system will consider four image descriptors used in various Computer Vision tasks and extract 3 * 256 RGB histogram, a Bag of visual words image descriptor, SIFT descriptor and the SentiBank visual representation. All the obtained representations will reduce using Principal Component Analysis (PCA).

The proposed system will extract three views of given social media image i.e. Visual view, subjective text view and objective text view and will give sentiment polarity based on given hypothesis table shown in table 1 using rules described in section 3.4.

II. LITERATURE REVIEW

Several papers investigated the problem of joint modelling the representation of images and associated text or tags for different tasks, such as image retrieval, social images understanding, image annotation and visual sentiment analysis. In [3] the authors presented different learning architectures for sentiment analysis of social posts containing short text messages and an image (i.e., Tweets). They exploited a representation learning architecture that combines the input text with the polarity ground truth.

The approach proposed in [10], combines visual features with text-based features extracted from the text subjectively associated by the users to images (i.e., descriptions and tags). To represent contents for sentiment analysis estimation, the authors proposed three different type of features extracted considering pairs of images and the related subjective texts: a visual feature defined by combining different visual descriptors usually used for visual classification ,a feature obtained by using the traditional Bag of Words approach on the subjective text, and a sentiment feature obtained by selecting the words of the subjective text whose sentiment scores (positive or negative) reported in SentiWordNet are larger than a threshold, and applying the Bag of Words on this restricted vocabulary. The considered features are exploited to define an embedding space in which the correlation among the projected features is maximized. Then a sentiment classifier is trained on the features projected in the embedding space.

In [1] author extract and employ an Objective Text description of images automatically extracted from the visual content rather than the classic Subjective Text provided by the users. The proposed method defines a multimodal embedding space based on the contribute of both visual and textual features. The sentiment polarity is then inferred by a supervised Support Vector Machine trained on the representations of the obtained embedding space.

Experiments performed on a representative dataset of 47235 labelled samples demonstrate that the exploitation of the proposed Objective Text helps to outperform state-of-the-art for sentiment polarity estimation.

In [8] paper, an image dataset with sentiment tags is built for training .Authors conduct experiment by training 15000 scene images on three different CNNs models GoogLeNet, ResNet and VGG Net and providing that deep learning can perform rather well on sentiment prediction task.

III. PROPOSED SYSTEM

Section 3.1 highlights the main differences between subjective text and the proposed objective text, Section 3.2 describes features extraction process, Section 3.3 describes proposed System diagram and section 3.4 describes expected results.

3.1 Comparison of subjective and objective text

Analysing social pictures for Sentiment Analysis brings several advantages. Indeed, pictures published through social platforms are usually accompanied by additional information that can be considered. Most of the existing works exploit social subjective textual information associated to images either to define the ground truth (i.e., by performing textual Sentiment Analysis on the text) or as an additional data modality (i.e., views) .In the latter case, both the visual and the textual information are used as input to establish the sentiment polarity of a post. Although the text associated to social images is widely exploited in the state of the art methods, it can be a very noisy source because it is provided by the users. There is no guarantee that the subjective text accompanying an image is useful for the sentiment analysis task. In addition, the tags associated to social images are often selected by users with the purpose to maximize the visibility of such images by the platforms search engine. In Flickr, for instance, a good selection of tags

helps to augment the number of views of an image, hence its popularity in the social platform. Those tags are independent from the sentiment evoked by the images. The semantic of an image may be given by a single object category, while the user-provided tags may include a number of additional terms correlated with the object which could be related to a larger vocabulary. This information is hence not always useful for sentiment analysis.
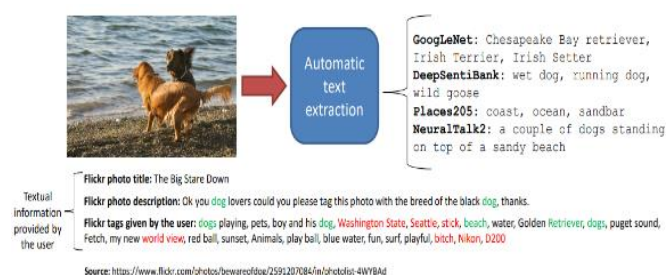

Figure 2 : Image From Flicker

Figure 2 shows an example image taken from the Flickr dataset. The textual information below the image is the subjective text provided by the user. As shown by this example, the text can be very noisy with respect to any task aimed to understand the sentiment that can be evoked by the picture. For example, the photo description is used to ask a question to the community. Furthermore, most of the provided tags include misleading text such as geographical information (i.e., Washington State, Seattle), information related to the camera (i.e., Nikon, D200), objects that are not present in the picture (i.e., boy, red ball, stick) or personal considerations of the user (i.e., my new word view). Another drawback of the text associated to social images is that two users can provide rather different information about the same picture, either in quality or in quantity. Finally, there is not guarantee that such text is present; this is key limit of all Visual Sentiment.

To overcome this limitations proposed system will exploit an objective aspect of the textual source that comes directly from the understanding of the visual content of the images. This text is obtained by employing three deep learning models trained to accomplish different visual inference tasks on the input image. In particular, the objective text associated to an image is obtained by considering the output text labels of GoogLeNet, ResNet and VGGNet.

3.2 Feature Extraction
3.2.1 Visual View

System considers four image descriptors used in various Computer Vision tasks. System will also extract 3 * 256 RGB histogram, a Bag of visual words image descriptor, SIFT descriptor and the SentiBank visual representation .Then, all the obtained representations will be reduce using Principal Component Analysis (PCA).

3.2.2 Text View

There are two textual views based on Subjective Text and objective text extracted from the images.
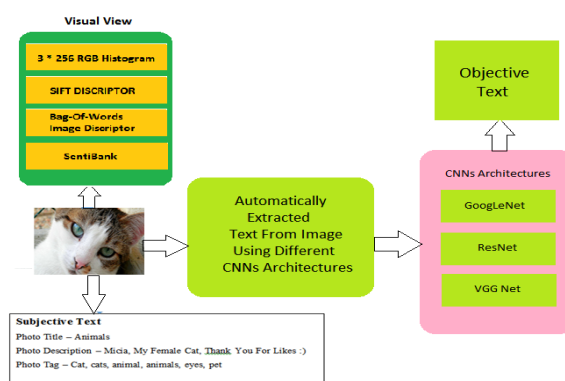

Figure 3: Feature Extraction Of System

3.2.2.1 Subjective Text view

This view reflects the subjective text information provided by the users such as photo title, description and tags. It consists of textual features, which will extract from text associated to images. System first constructs a vocabulary from a training dataset and represents the textual view of an image using a traditional bag-of-words approach, which counts how many times a word appears in text around the image. Since this representation is highly sparse,

system exploit SVD for large and sparse matrices to reduce the dimensions of the textual feature matrix appears in text around the image.

3.2.2.2 Objective Text View

Objective Text will be obtained through 3 deep learning CNNs architectures GoogLeNet, VGG Net and ResNet. As shown in Figure 3, each architecture will provide a description, in some sense objective, of the input image from a different point of view, as each architecture has been trained for a different task. This will allow to obtain a wide objective description of the image which takes into account different semantic aspects of the visual content. Redundant terms are not a drawback for the proposed approach, indeed the presence of more occurrences of similar or related terms enhances the weight of these correct terms in the representation extracted by proposed system, and reduces the effect of noisy results. For these reasons system will compute a Bag of Words representation followed by a SVD dimensionality reduction.

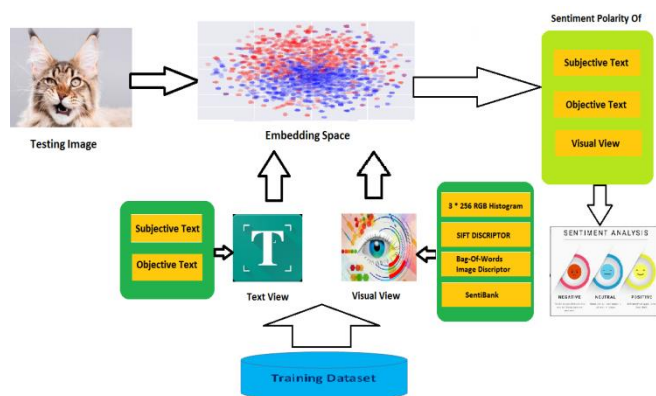### 3.3 Proposed System Diagram

Figure 4 : Overview Of proposed system

This section presents a visual sentiment analysis method that uses Visual view And Text views. An overview of the proposed system is shown in Figure 4. As shown, System will first extract features from each view (See 3.2). Then embedding space is used to train an visual sentiment analysis.

The system has 4 visual views 3 * 256 RGB histogram, SIFT descriptor , a Bag of visual words image descriptor and the SentiBank visual representation. The system has two text views such as subjective text view (provided by user), Objective text view (Extracted from image using different CNNs architectures)

After feature extraction from all views finally visual sentiment analysis is done to give sentiment polarity "positive" or "negative" or "Neutral" based on hypothesis table as shown in table 1 using four rules describe in section 3.4 .

IV. Expected Result

TABLE 1 HYPOTHESIS TABLE FOR SENTIMENT POLARITY

SR. NO	Sentiment Polarity			Proposed System sentiment Polarity
	Subjective Text View	Objective Text View	Visual View	
1	POSITIVE	POSITIVE	POSITIVE	POSITIVE
2	POSITIVE	POSITIVE	NEGATIVE	POSITIVE
3	POSITIVE	POSITIVE	NEUTRAL	POSITIVE
4	POSITIVE	NEGATIVE	POSITIVE	POSITIVE
5	POSITIVE	NEGATIVE	NEGATIVE	NEGATIVE
6	POSITIVE	NEGATIVE	NEUTRAL	NEGATIVE
7	POSITIVE	NEUTRAL	POSITIVE	POSITIVE
8	POSITIVE	NEUTRAL	NEGATIVE	NEUTRAL
9	POSITIVE	NEUTRAL	NEUTRAL	NEUTRAL
10	NEGATIVE	POSITIVE	POSITIVE	POSITIVE
11	NEGATIVE	POSITIVE	NEGATIVE	NEGATIVE
12	NEGATIVE	POSITIVE	NEUTRAL	POSITIVE
13	NEGATIVE	NEGATIVE	POSITIVE	NEGATIVE

14	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE
15	NEGATIVE	NEGATIVE	NEUTRAL	NEGATIVE
16	NEGATIVE	NEUTRAL	POSITIVE	NEUTRAL
17	NEGATIVE	NEUTRAL	NEGATIVE	NEGATIVE
18	NEGATIVE	NEUTRAL	NEUTRAL	NEUTRAL
19	NEUTRAL	POSITIVE	POSITIVE	POSITIVE
20	NEUTRAL	POSITIVE	NEGATIVE	POSITIVE
21	NEUTRAL	POSITIVE	NEUTRAL	NEUTRAL
22	NEUTRAL	NEGATIVE	POSITIVE	NEGATIVE
23	NEUTRAL	NEGATIVE	NEGATIVE	NEGATIVE
24	NEUTRAL	NEGATIVE	NEUTRAL	NEUTRAL
25	NEUTRAL	NEUTRAL	POSITIVE	NEUTRAL
26	NEUTRAL	NEUTRAL	NEGATIVE	NEUTRAL
27	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL

The proposed system will extract three views of given social media image i.e. visual view, subjective text view and objective text view and will give sentiment polarity based on given hypothesis table shown in table 1 using following rules.

Rule 1 : If any two views among three views have positive polarity then proposed system will give output as positive sentiment polarity as shown in serial numbers 1,2,3,4,7,10,19.

Rule 2 : If any two views among three views have negative polarity then proposed system will give output as negative sentiment polarity as shown in serial numbers 5,11,13,14,15,17,23.

Rule 3 : If any two views among three views have neutral polarity then proposed system will give output as neutral sentiment polarity as shown in serial numbers 9,18,21,24,25,26,27.

Rule 4 : If all three views of image having unique polarity i.e. one positive ,one negative and one neutral polarity then system will consider objective text view polarity as output polarity as shown in highlighted serial numbers 6,8,12,16,20,22.

V. Conclusion

The proposed system will addresses the challenge of image sentiment polarity estimation by proposing a novel source of text for this task. The aim is to deal with the issue related to the text provided by users which is commonly used in most of the previous works. It explains a study in which Objective Text extracted considering the visual content of images is compared with respect to the Subjective Text provided by users. Proposed system will first identifies several drawbacks brought by the Subjective Text due its intrinsic nature, and then it will demonstrates experimentally that the exploitation of Objective Text associated to images provides better results than the use of the Subjective Text provided by the user. The Objective Text exploited by proposed approach will not present the highlighted limitations and it will automatically extract from the image. The expected result will support the use of Objective text automatically extracted from images for the task of Visual Sentiment Analysis in lieu of the Subjective Text provided by users and will give sentiment polarity based on hypothesis table.

VI. REFERENCES

[1]. Alessandro Ortis Giovanni M. Farinella,Giovanni Torrisi,Sebastiano Battiato Visual Sentiment Analysis Based on Objective Text [Journal]. - Catania, Italy : IEEE, 2018. - Vols. 978-1-5386-7021-7/18/.

[2]. B. Zhou A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva Learning deep features for scene

recognition using places database [Journal]. - 3Universitat Oberta de Catalunya : [s.n.]. - Vols. Advances in Neural Information Processing Systems, 2014, pp. 487–495.

[3]. Bertini1 Claudio Baecchi1 ·Tiberio Uricchio1 · Marco A multimodal feature learning approach for sentiment [Journal]. - New York : Springer, 2015.

[4]. C. Szegedy W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, Going deeper with convolutions [Journal]. - [s.l.] : IEEE, 2015. - Vol. In proceedings of the IEEE Conference on Computer Vision and Pattern.

[5]. Eunjeong Ko Chanhee Yoon,Eun Yi Kim Discovering Visual Features for Recognizing User's [Journal]. - Konkuk University,South Korea : IEEE, 2016. - Vols. 978-1-4673-8796-5/16.

[6]. Fei-Fei A. Karpathy and L. Deep visual-semantic alignments for generating image descriptions [Journal]. - [s.l.] : IEEE. - Vols. JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015.

[7]. https://monkeylearn.com/sentiment-analysis/ [Online] // https://monkeylearn.com.

[8]. Junfeng Yao Yao Yu and Xiaoling Xue Sentiment Prediction In Scene Images Via Convolution Neural Networks [Journal]. - Beijing,China : IEEE, 2016. - Vols. 978-1-5090-4423-8/16.

[9]. Kaikai Songa Ting Yaob, Qiang Linga,∗, Tao Mei Boosting Image Sentiment Analysis with Visual Attention [Journal]. - China : ELSEWHERE, 2018.

[10]. Marie Katsurai Shin'ichi Satoh IMAGE SENTIMENT ANALYSIS USING LATENT CORRELATIONS AMONG VISUAL, [Journal]. - Tokyo, Japan : IEEE, 2016. - Vols. 978-1-4799-9988-0/16.

[11]. Varshney Mayank Amencherla and Lav R. Color-Based Visual Sentiment for Social [Journal]. - Urbana-Champaign : IEEE, 2017. - Vols. 978-1-5090-6026-9/17.

Cite this article as :