International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307

Deep Learning Networks For Visual Sentiment Analysis: CaffeNet and TensorFlow

Shafi S. Shaikh*[1], Pooja M. Tayade[2], Dr. S. N. Deshmukh[3]

[1,2] Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
[3]Professor, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

ABSTRACT

Predicting emotion, opinion, sentiment from the comments, tweets, blogs, or bunch of words written by user has importance in Machine Learning. In Natural Language Processing emotion of particular user or person is extracted from the text, sentence, and document which all being written by user. NLP uses written text word, n-grams, combination of words as features and tries to find out relation that may exist among them. But as technology is evolving day by day it is very easy to capture or click picture, selfie through cellphone, smartphone, tablets, phablets, camera and all digital devices. These pictures of self or group are uploading on social media at every second with a suitable caption. So it's becoming very easy to express through pictures, videos, images than to write thousand words. So to analyses sentiment of pictures on social media the NLP is way lack behind. It needs to be deal in objects rather than plain text. So it's been need to process image and extracts sentiment from these pictures. So his paper will deal with the image and recognizing face as object and try to find emotion or sentiment of that picture.

Keywords: NLP, VSA, Neural Network, Bottlenecks, TensorFlow, CaffeNet

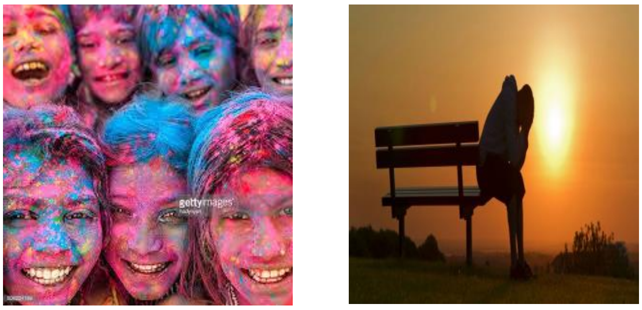
I. INTRODUCTION

Now a day's people are attracting more towards social networks as it provides user a virtual world to express themselves and to read, write upload any content as per their wish. There is no control over user about what to say? Whom to say? And how to say? There are billions of people who are using social network platforms such as Facebook, Twitter, YouTube, Flicker, Instagram etc. to share their views, thoughts and experience and to express their opinions virtually on all events and subjects. Recent growth of these social media has led to an explosion in amount, throughput and variety of multimedia content generated, each and very day billions of messages and posts are generated which causes Big Data. As technology is booming every person has his

own smartphone, tablet, laptop any many handheld devices and also the evolution in internet technologies like 3G/4G and coming soon 5G has brought all the world onto the fingertips and screen. This is making world into virtual village. It is very convenient for all users to capture or download image and upload photo or image with short text as post to express or share their opinion and lives. For example as shown in figure.1(a) suppose one user has posted content about his/her "Holi" day with texts and images showing that this tweet is related to 'Holi' day (Indian festival of color). Text containing words "Happy Holi !!!" clearly conveys the message with emotion he is in and with that text by looking at image contents " Happy Faces with holi colors" indicates person's happy, pleased state of mind and he is enjoying all this without any sort of doubt.

Now let's take another example as given in figure 1(b), if a person has posted a photo on social media with caption such as "Another Sunset….!!!". Now with the help of text based sentiment analysis when algorithm tries to evaluate the polarity of situation or statement only considers words.by looking at only text it's very hard to talk about the actual emotion that text carries. So NLP or Text based sentiment analysis will consider this text as “Neutral” and put a side in the context of emotion. Now by observing an image given below to the caption anyone can guess that a person who posted a picture is in Sad or Unpleased state of mind. As picture contains the object sun is going to settle down with reddish sky shoulders are dropped. This is nothing but a sign of loosing of hope or deep sadness. So now the sentiment of whole post clearly goes into Negative Category in terms of text.

Sometimes when you feel happy, user can upload a sunshine picture or whenever you could feel saddened you will update a heavy downpour picture with sad emoticons, just suggesting that no need of words to express yourself. However, text based sentiment analysis methods cannot handle the information except text. Since images can provide an efficient way to express emotions and affects others. It is necessary to further study visual sentiment analysis, though it is very important aspect but comes with great challenges to overcome for better result. The advantage of having machine capable of understanding human feelings are numerous and would imply a revolution in various fields such as robotics, medicine, education or entertainment. There are number of approaches tends towards the sentiment discovery from image and to reduce the affective gap between low-level and high level features, has been carried out by the researchers and presented to world over the years. For visual content sentiment specifically for images but the performances are not so convincing and relative measures behind this purpose of visual sentiment analysis has been lacking.



(a)

“Happy Holi!!!”

(b)

“It’s Another Sunset!!!!”

Figure 1. Example of Twitter Posts With Text and Image

II. RELATED WORK

In today’s work sentiment analysis is generally divided into three categories: text based, image based and video based. Traditional SA focuses on text, document sentences and show immense progress. Whereas, image based SA is in its initial place until now. Very little work has carried out for image SA on the other hand, video base SA is just in the beginning.

A. Text Sentiment Analysis

In the text based SA, BoW or Corpora or dictionary dependent algorithms and machine learning methods are important factor. The approach in text SA, is first obtain the structure of words or sentences according to the dictionary and second, compute the weightage of particular word or distance between them to show that which sentiment it represents either positive, negative or neutral in most of the research. In Turney’s work, sentimental phrases are selected from reviews, and these are classified according to the average sentiment orientation. Most of the machine learning methods uses classification machine learning algorithms for prediction such as Naïve Bayes, MaxEnt and SVM. With the development of social applications text based SA were used in twitter [1-4].

B. Visual Sentiment Analysis

Affective Image Prediction With a growing number of images being used to express opinions in social networks [5], image sentiment analysis has attracted more and more attentions [6]. From the aspect of features used in this area, work can be roughly divide prior work into low-level based [7] [8] and mid-level based methods [9] [10]. Convolutional Neural Networks Several recent work has exploited deep convolutional neural networks for image sentiment prediction. Based on previous work, Chen et al. trained a deep CNN model on Caffe [11] [12].

III. METHODOLOGY

A. Emotion Distribution

The So aside from sentiment, which focuses on only positivity/negativity, what are probable mappings of ANPs [7] to emotions for each language? What emotions are most frequently occurring across languages? Given the set of keywords E(l) = {e(E(l) = {e(l)ij | i = 1…7, j = 1…ni} ijl) | i = 1… 24, j = 1… ni} describing each emotion i per language l, where ni is the number of keywords per emotion i, the set of ANPs belonging to language l, noted as $x \in X^{(l)}$, and the number of images tagged with both ANP x and emotion keyword eij , C(x) = {c(ij^{x}) | i = 1…7, j = 1… ni}, we define the probabilities of emotion for each ANP x in language l as:

$$\text{emo}^{i}(x) = \frac{\frac{1}{n_i}\sum_{j=1}^{n_i} c_{ij}^{(x)}}{\sum_{i=1}^{24}\frac{1}{n_i}\sum_{j=1}^{n_i} c_{ij}^{(x)}} \in [0,1] \text{eq.(1)}$$

So based on eq (1), let’s compute a normalized emotion score per language l and emotion i as:

$$\text{score}^{i}(l) = \frac{\sum_{x=1}^{|X^{(i)}|} \text{emo}^{i}(x) \;\; . \; count(x)}{\sum_{i=1}^{24}\sum_{x=1}^{|X^{(l)}|} emo^{i}(x) \;\; . \; count(x)} \in [0,1]$$

eq.(2)

So in this paper all algorithms are tending to derive the emotion score as per above rules and with the help of two neural networks. Figure (2) shows overall the architecture of visual sentiment prediction using convolutional neural network. CNN tries to fine tune at each and every step for better understanding of features and classification of image. CaffeNet and TensorFlow try to predict image class by self-improvement in learning. CNN algorithm process through several layers at each layer image features are extracted processed and compressed to give as output to next network layer. Last layer looks for likelihood of the features in the image and predicts the result. This process of learning is feed forward. Deep learning is the important aspect of TensorFlow and Caffe. So let’s implement these architectures and compare performances. Before that let’s brush-up with Caffe and TensorFlow.

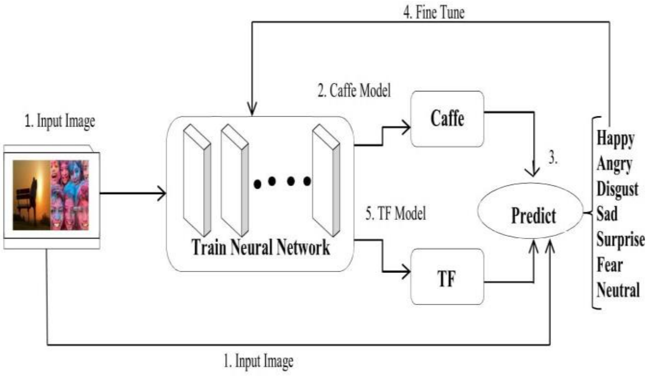


Figure 2. Overall Architecture of The Proposed Visual Sentiment Prediction Framework

B. CaffeNet

The adopted CaffeNet [11] architecture contains more than 60 million parameters, a figure too high for training the network from scratch with the limited amount of data available in the dataset. Given the good results achieved by previous works about transfer learning[8-10], it’s good to explore the possibility of fine-tuning an already existing model. Fine-tuning consists in initializing the weights in each layer except the last one with those values learned from another model. The last layer is replaced by a new one, usually containing the same number of units as classes in the dataset, and randomly initializing their weights before “resuming” training but with inputs from the target dataset. The advantage of this approach compared to fully re-training a network from a random initialization on all the network weights is that it essentially starts the gradient descent learning from a point much closer to an optimum, reducing both the number of iterations needed before convergence and decreasing the likelihood of over fitting when the target dataset is small.

C. TensorFlow

TensorFlow[14] is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

By default, this script runs 4,000 training steps. Each step chooses 100 images at random from the training set, finds their bottlenecks [13] from the cache, and feeds them into the final layer to get predictions. Those predictions are then compared against the actual labels to update the final layer's weights through a backpropagation process. The first phase analyzes all the images on disk and calculates the bottleneck values for each of them. These layers are pre-trained and are already very valuable at finding and summarizing information that will help classify most images.

IV. EXPERIMENTATION

A. Dataset

For the implementation of visual sentient analysis dataset of images is the important key. Various social sites Twitter, Facebook, Flicker, Instagram, Getty, Shutter etc. user post images publically and privately. For building dataset in this paper all publically available images are collected and these images are classified into their particular class as per user domain of sentiment. These top seven sentiments are grouped based on text related to that image or caption which are loosely bounded to all images. Google image search is also used to gather particular dataset. Also there are many datasets available publically on Github. So one such particular dataset from img_dataset[15] was also added.

There are total of 38853 images are available to train neural model and test particular model. Each image represents particular class of emotion as shown in Figure 3 to Figure 9.

TABLE I. TRAINING AND TESTING DATASET

Class	Angry	Happy	Sad	Surprise	Disgust	Fear	Neutral	Total
#of images	5422	9512	6475	4629	0547	5720	6548	38853

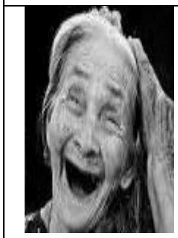
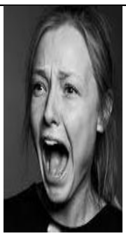
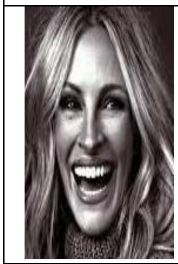
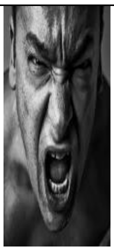
B. Training and Validation

In this paper two most deep learning algorithms are used Caffe and TensorFlow. Deep learning refers to a class of artificial neural networks (ANNs) composed of many processing layers. ANNs existed for many decades, but attempts at training deep architectures of ANNs for visual experiments not carried out until CNN comes in play. Convolutional neural networks are a special type of feed-forward networks. These models are designed to emulate the behavior of a visual cortex. CNNs perform very well on visual recognition tasks. CNNs have special layers called convolutional layers and pooling layers that allow the network to encode certain images properties. Caffe works combining the feature of CNN. TensorFlow uses a single dataflow graph to represent all computation and state in a machine learning algorithm, including the individual mathematical operations, the parameters and their update rules, and the input preprocessing. In a TensorFlow graph, each vertex represents a unit of local computation, and each edge represents the output from, or input to, a vertex. Let’s refer to the computation at vertices as operations, and the values that flow along edges as tensors.

C. Implementation

Dataset is divided into two parts 75% of data images are used as training purpose and 25% of data images are used to test the trained model. Images are classified for training purpose in classes of sentiments and each image represent the sentiment as class label (Happy, Sad, Fear, Angry, Surprise, Disgust, and Neutral). For validation of trained model input image is given without knowing its label. Result will give the score (between 0-1) of each sentiment and it will give the higher score to image to which class it belongs. Such validated example or results are shown in Table II. And Table III compares the result between different algorithms.

TABLE II
TRAINED TENSORFLOW ALGORITHM CLASSIFYING IMAGE AS PER SENTIMENT SCORES

Test Image	Sentiment Score	Test Image	Sentiment Score
	1= 0.928817 0=0.067379 4=0.003632 3=7.172e-06 6=2.42e-07 2= 0.00000 5= 0.00000		2=0.855352 1=0.123859 6=0.016086 3=0.002513 4=0.001424 5=0.000000 0=0.000000
	1= 0.940841 4=0.058374 0=0.000724 6=2.01e-05 3=1.478e-06 5= 0.000000 2=0.000000		3=0.885482 1=0.099059 0=0.008214 6=0.007016 4=5.352e-06 5=0.000000 0=0.000000

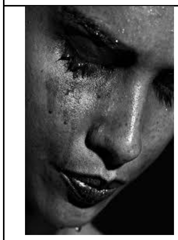
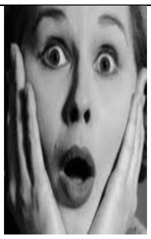
Test Image	Sentiment Score	Test Image	Sentiment Score
	5=0.942715 0=0.053164 3=0.003077 6=3.847e-05 1=8.002e-06 4=0.000000 2=0.000000		4=0.880981 1=0.118798 0=0.001345 3=3.706e-05 5=3.565e-07 2=0.000000 1=0.000000
INDEX: 0 = Neutral 1 = Happy 2 = Fear 3 = Angry 4 = Surprise 5 = Sad 6 = Disgust			

TABLE III. COMPARISON OF OVERALL PERFORMANCE BETWEEN DIFFERENT ALGORITHMS

Algorithms	Overall Validation
Low-level Features (Borth et al. 2013)	0.508
SentiBank (Borth et al. 2013)	0.514
CaffeNet (Proposed Method)	0.813
TensorFlow (Proposed Method)	0.897

V. CONCLUSION

Sentiment analysis is an important task for various applications such as advertisements and recommendation, financial and educational sectors. While vast majority of previous works of sentiment analysis on social web were conducted on text, this paper proposes to focus on the analysis of images, one of the dominant media types of online micro blogging services. In this paper, TensorFlow and CaffeNet, novel sentiment analysis framework based upon convolutional neural network are introduced for visual sentiment prediction. Result shows that the image representations from the CNN trained on a large-scale dataset could be efficiently transferred for sentiment analysis. To evaluate the proposed method on real-world data, dataset is constructed from the photo posts publically on social media like Flicker. Validation introduces results for seven classes in the range of 0-1.Experiments on own image dataset demonstrate that TensorFlow proposed model outperform the state-of-the-art methods (Borth et al. 2013). Borth et al. 2013 implemented sentiment analysis algorithm on Twitter dataset with extracting low-level features and SentiBank collection of ANPs and images which gives overall performance of 0.508 and 0.514 respectively. But CaffeNet implementation can predict sentiment of image with accuracy nearly 0.813 on own image dataset. But when same dataset is taken and Tensor Flow comes in play goes to 90% of overall performance which is far better than any other algorithms with low computation time than other. It's very important that which dataset is you are using and how large is your dataset. As large as dataset overall performance is also going to be accurate.

There are several interesting future directions to explore. First given the amount of training dataset increase dataset and try to implement CNN with un-supervised intent. Images are of different nature, like one color in image may express happy sentiment and in another image it may looks sad. So to differ between features in images as per context is big concern. Then try to extend this work on video sentiment prediction.

Figure 3. Image Data Showing Happy Sentiment

Figure 4. Image Data Showing Angry Sentiment

Figure 6. Image Data Showing Disgust Sentiment

Figure 7. Image Data Showing Neutral Sentiment

Figure 8. Image Data Showing Sad Sentiment

Figure 9. Image Data Showing Surprise Sentiment

VI. REFERENCES

[1] P. Turney, “Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews”, In Isabelle P, ed. Proc. Of the ACL. Morristown: ACL, 2002, 417-424.

[2] S. M. Kim, E. Hovy, “Automatic detection of opinion bearing wordsand sentences”, In Carbonell JG, Siekmann, IJCNLP, 2005, ACL, 61-66.

[3] H. Yu, and V. Hatzivassiloglou, “Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences,” In Collins M, Steedman M, eds. Proc. of the EMNLP, 2003. Morristown: ACL, 129—136.

[4] M. Hu, B. Liu, “Mining and summarizing customer reviews”, In Kohavi R, ed. Proc. of the KDD, 2004. New York: ACM Press, 168—177.

[5] Jana Machajdik and Allan Hanbury, “Affective image classification using features inspired by psychology and art theory,” In Proceedings of the international conference on Multimedia. ACM, 2010, pp. 83–92.

[6] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang, “Can we understand van Gogh's Mood?: Learning to infer A_ects from Images in Social Networks”, In ACM MM, 2012, 857-860.

[7] D. Borth, R. Ji., “Large-Scale Visual Sentiment Ontology And Detectors Using Adjective Noun Pairs” , In Proc. of the 21st ACM International Conference on Multimedia, 2013, 223-232.

[8] D. Cao, R. Ji, D. Lin, S. Li, “A cross-media public sentiment analysis system for microblog”, In ACM Multimedia Systems Journal, Special Issue Paper, DOI: 10.1007/s00530-014- 0407-8 , 2014.

[9] Rongrong Ji, Donglin Cao, Dazhen Lin, “Cross-modality sentiment analysis for social multimedia”, In 2015 IEEE.

[10] Quanzeng You and Jiebo Luo, Hailin Jin, Jianchao Yang, “Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia”, In 2016 ACM.

[11] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev et al., “Caffe: Convolutional Architecture for Fast Feature Embedding”, In arXiv:1408.5093, 2014.

[12] Tao Chen, Damian Borth, Trevor Darrell, and Shih-Fu Chang, “Deepsentibank: Visual sentiment concept classifica-tion with deep convolutional neural networks”, In arXiv preprint arXiv:1410.8586, 2014.

[13] Christian Szegedy, Wei Liu, Yangqing Jia, Andrew Rabinovich et al., “Going Deeper with Convolution”, In 2015, IEEE Explore.

[14] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen et al., “TensorFlow: A System for Large-Scale Machine Learning”, In 2016, 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’16).

[15] Weblink - https://github.com/sjchoi86