2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)

Image Sentiment Analysis using Deep Learning

Namita Mittal
Department of Computer Science
MNIT, Jaipur, India
nmittal.cse@mnit.ac.in

Divya Sharma
Department of Computer Science
The IIS University, Jaipur, India
divya.sharma2414@gmail.com

Manju Lata Joshi
Department of Computer Science
International School of Informatics & Management, Jaipur, India
manjulata.joshi@icfia.org

Abstract— **Sentiments are feelings, emotions likes and dislikes or opinions which can be articulate through text, images or videos. Sentiment Analysis on web data is now becoming a budding research area of social analytics. Users express their sentiments on the web by exchanging texts and uploading images through a variety of social media like Instagram, Facebook, Twitter, WhatsApp etc. A lot of research work has been done for sentiment analysis of textual data; there has been limited work that focuses on analyzing the sentiment of image data. Image sentiment concepts are ANPs i.e. Adjective Noun Pairs automatically discovered tags of web images which are useful for detecting the emotions or sentiments conveyed by the image. The major challenge is to predict or identify the sentiments of unlabelled images. To overcome this challenge deep learning techniques are used for sentiment analysis, as deep learning models have the capability for effectively learning the image behavior or polarity.**

Image recognition, image prediction, image sentiment analysis, and image classification are some of the fields where Neural Network (NN) has performed well implying significant performance of deep learning in image sentiment analysis. This paper focuses on some of the noteworthy models of deep learning as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Region-based CNN (R-CNN) and Fast R-CNN along with the suitability of their applications in image sentiment analysis and their limitations. The study also discusses the challenges and perspectives of this rising field.

Keywords—image sentiment analysis, deep learning, deep neural network, convolutional neural network, Region based CNN and Fast R-CNN.

I. INTRODUCTION

Sentiments can be expressed using text, image or videos [1]. A plethora of research papers available for text sentiment analysis, but still image sentiment analysis is not much explored. With the increasing use of social media to express sentiments, it become one of the important area of research, therefore since last few years plenty of research has been done for the same to achieve optimum results.

Multiple techniques and algorithms have been proposed for image sentiment analysis broadly classified into two that is Machine Learning based algorithms and Lexicon based algorithms. Machine Learning based algorithms includes Support Vector Machine (SVM), Neural Network, Naïve Bayes, Bayesian Network, and Maximum Entropy while Lexicon based algorithms includes statistical and semantic-based techniques.

Deep learning is the subfield or a technique of machine learning that makes the computer intelligent enough, so the machine is capable to learn from experience and perceive the world of concepts. Computer fetch knowledge from the real world experience, without human help to make computer understand all the situation or to make decisions. [2].

Deep learning techniques have also been explored for image sentiment analysis and providing significant results too. The deep learning modal can acquire accuracy or sometimes may exceeds the human level intelligence and can give an efficient task performance. The term "Deep" in Deep Learning indicates to the number of hidden layers in the Neural Networks. Deep Learning models are trained by using any large set of labelled data and an architecture known as Neural Network Architecture that makes the feature or parameter learning directly from the given data without any human intervention or manual feature extraction . Deep learning plays a major role for image sentiment analysis for giving various techniques like Convolutional Neural Network (CNN), Region Neural Network (RNN), Deep Neural Network (DNN) and Deep Belief Network (DBN) etc to get optimum results. Deep Learning can be seen as a framework that produce accurate parameter learning for image classification. This paper studies and analyzes about different techniques of deep learning namely; DNN, CNN, R-CNN and Fast R-CNN. Further, in section II, the paper discusses the research work done so far for image sentiment analysis using above mentioned techniques and their outcomes. In section III paper will analyze the performance and limitations of techniques discussed in section II. Section IV concludes the paper.

II. LITERATURE REVIEW

Though many researchers explored number of techniques for image sentiment analysis, machine learning based techniques are performing significantly well. Among various machine learning based techniques deep learning based techniques outperforming for image sentiment analysis. This section analyzes some important research works performed by researchers using deep learning techniques, along with their outcomes

A. Deep Neural Network (DNN)

DNN is used for visual sentiment analysis as well as textual sentiment analysis. Neural Network use multiple layers, initially image is fed in the input layer and then processed to give an output through output layer. In the middle of input and output layer multiple hidden layers are present for further processing of an input image, due to multiple hidden layers in a neural network it is known as Deep Neural Network. Computer only understand image in a matrix form with each pixel holding some value the value representing the pixel refers as activations. Each neuron is connected to other neurons; Activation of first layer only determines the activation of next layer. The goal is to join or bind the image pixels into edges, edges into the sub patterns and finally

978-1-5386-7325-6/18/$31.00 ©2018 IEEE
DOI 10.1109/WI.2018.00-11

combine the identified patterns into the image for analyzing sentiments.

The study [5] have proposed an approach for image or visual sentiment analysis in respect to deep coupled adjective and noun neural networks. The main aim of their work is to reduce the amount of inter-class variance, so the process discussed is to study the middle-level sentiment representation first along with deep neural network, then as a second step prediction is performed with optimization and at last trained the system with mutual supervision with learned adjectives and networks the Rectified Kullback-Leibler loss (*ReKL*). ReKN is used for eliminating unreasonable outcomes and train efficient sentiment representation. DCAN (Deep Coupled Adjectives and Noun Neural Networks) is used for visual sentiment analysis. The introduced system can analyze the feature of web noisy images along with ANP and get the result on Twitter and SentiBank dataset.

The paper [1] have presented multiple techniques for sentiment analysis using deep learning which covers both image sentiment as well as text sentiments. The paper includes models like CNN which is best suited for visual sentiment analysis. Recursive Neural Network falls under supervised learning technique, Deep Belief Network is used for unlabelled data and to overcome the limitation of unlabelled data, it consists of several layers that are hidden. Their study conveys that deep learning networks are better as compared to SVM and normal neural networks as deep neural networks have more hidden layers.

B. Convolutional Neural Network (CNN)

CNN is a feed-forward neural network majorly used for image processing, image classification or image prediction; So CNN's one of the most important application is the analysis of Images or visuals. A sequence of functions is performed for image Sentiment analysis using CNN. This comprises of convolutional layer followed by nonlinear layer followed by Pooling Layer and fully connected layer.

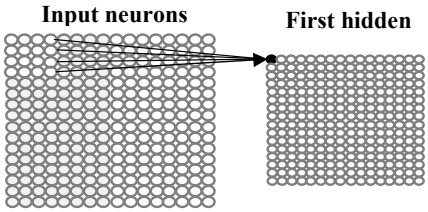

Fig. 1. Basic CNN model with one layer [17]

The very first layer for CNN image classification is convolutional layer as the image is entered, assume the analysis or reading of an image starts from the top left corner, then a small segment or matrix of the image is selected known as filters. There would be multiple convolutional networks as the image will pass one convolutional layer and the output of a layer would be the input for the next layer. The Nonlinear Layer is the second function in the pipeline, after the convolution operation. It consists of a function known as activation function which gives the CNN a nonlinearity behavior. Pooling Layer follows nonlinear layer reduces the workload by reducing the features of an image or the image volume will be reduced if the given image is of large size. For example, if any of the features are already identified in the modal in the previous convolution operation, then it would not be processed further for identification. This is referred to as down sampling or sub-sampling. After pooling layer the output is still awaited, and then the concept of *the fully connected layer* is introduced. It takes the output data from the convolutional networks. It flattened the matrix into the vector and then gives it into the fully connected layer.

The study [3] has introduced a concept of visual sentiment analysis with Deep CNN, they evaluated the data on two datasets namely Twitter and Tumbler. The data is collected based on the photos tags or hashtags and lexicons for identifying the emotions behind any sentence from word level. To get the best and true feeling or sentiment of the image a survey have conducted for classify the images namely strongly positive, weakly negative, neutral, etc. and takes two baseline techniques that is low-level visual features and also SentiBank, The study conducted experiments for baseline approach for low-level features conclude that SentiBank is more powerful for efficient sentiment analysis.

The concept of DeepSentiBank was introduced by [10] for visual Sentiment classification with DCNN. The experiment is based on a single server system which consists of 16- core dual Intel E5-2650L processor with 64 GB Memory. The Deep CNN modal is used for training the system using Caffe. On the basis of the experiment done in their study CNN's based approach has great accuracy as compared to SVM.

The study mentioned at [7] has introduces a model, which was trained for large-scale image data, used Flickr photos for analysis as it consists of a huge amount of images with a large number of variations. It consists of progressive training as well as transfer learning for the labeled dataset of images. Their study concludes that using convolutional neural network one can achieve high accuracy and high performance for analysis of image sentiment.

The research in [9] have designed a new architecture using the CNN and introduced a new training technique to overcome the large-scale training data which is of a noisy nature. The experiment was performed on large image data set of Flicker images by SentiBank for training CNN and implemented the modal on the Caffe which is publically available. All the experiments are examined on Linux X86_64 systems with 32GB RAM. The system has taken 90% of the image data randomly from the Flicker dataset for training and 10% is used as a testing data. They also perform fine-tuning of the images [13].

The study in [6] presented the sentiment analysis technique for image sentiment as well as text sentiment. They present a method that consists of CNN architecture for text sentiment analysis and image sentiment analysis to perform multimedia sentiment analysis. The experiment is based on twitter datasets, the data contains both positive as well as negative tweets. The model proposed compares the text between CNN with the following Naïve Bayes, Logistic Regression and SVM, for analyzing textual sentiments and CNN is compared with the following that is low-level features, sentribute and SentiBank for image sentiment analysis. In this experiment, it is concluded that the multi

CNN has performed significantly better than other techniques.

The study in [4] proposed a system using the CNN to fetch certain parameters from the image and classify the image according to their behavior and parameters in an appropriate class. Also created different neural networks to train the modal for self-observation of the patterns by itself and tested the performance and the accuracy level of the modal on CPU and GPU and stated that CNN is one of the good choices for image classification [7].

The study in [12] have explored about the possibilities of emotions for an image by the help of deep learning, 5 emotions are identified in the paper and categories are Love, Happiness, Violence, Fear, and Sadness. Data was collected from Flicker for these categories and perform experiments using various classification methods as SVM, Fine-tuning, etc for the same. The paper proposed few methods (based on deep learning) for image sentiment analysis and conclude that ResNet-50 method on Flickr data performed best amongst all.

A rigorous empirical study by [15] shows comparisons among a number of fine-tuned CNN architectures for visual sentiment analysis. The study analyzed that deep architectures can learn features useful for identifying image sentiments in social networks and state of the art models, experimented on datasets of Twitter were presented. Their work also demonstrates that selection of pre-training a model initialization can make the difference when the data set is small, otherwise cost increases.

C. R-CNN (Region Convolutional Neural Network)

R-CNN was introduced in [14] is a method for searching selective regions for detecting objects from the given image for analyzing the sentiment of an image by the help of objects around the image.

R-CNN extract 2000 regions from the input image and generate a box boundary around the regions and fed the regions into CNN that is why it known as Region Convolutional Neural Network. CNN is used to extract features or parameters from the regions inputted and then the features extracted are submitted into SVM to classify the object present in the region. The major challenge for R-CNN is that takes more time and more memory as it trains the network to classify and analyze 2000 different regions for a single image.

In the work proposed by [8], they suggested a model based on the mid-level features of the images that combines the techniques of SentiBank, RCNN and SentiStrength. Results of experiments conducted on the Flickr image dataset show that their approach achieves better sentiment classification accuracy.

The work done in [16] proposed a framework to advantage regions, where they first used off the shelf objectness tool to generate the candidates and apply a potential candidate selection method to remove redundant and noisy objects. Further, they computed the sentiment scores using CNN and the effective regions are discovered. Then finally they combine both the scores as objectness score as well as sentiment scored to find effective regions automatically. Their framework only required image level label which significantly reduces the annotation burden required for training. The experiment conducted on 8 benchmark datasets shows that the proposed algorithm outperforms as compared to state of art approaches.

D. Fast R-CNN

To overcome the limitations of R-CNN that is to reduce the processing time and less memory usage a fast object detection algorithm was proposed [14, 18] which is popularly known as Fast R-CNN, it is similar to R-CNN. In

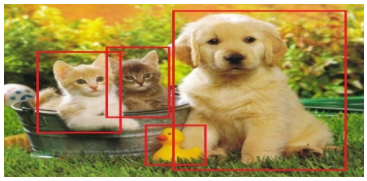

Fig. 2. Extracted regions from the input

Fast R-CNN, the image is not initially divided into different regions but the image is first inputted to CNN where it generate convolutional feature map. By the help of convolutional feature map regions are identified and box boundary is formed, and to give input to the fully connected layer it is reshaped it by using ROI (Region of Interest) pooling layer so that the fully connected layer must have the same size images, at last, softmax layer is used to predict the image. Therefore, in the Fast R-CNN, we need not to input 2000 regions to CNN each time. We just need to perform convolution operation only once for an image.

The research in the study [9] introduced a modal for image sentiment analysis by combining features of multiple techniques like SentiBank, RCNN, and SentiStrength, the experiment is performed on Flicker images dataset. The results show that for sentiments object detection is based upon SIFT feature, but this is not much efficient than used R-CNN for better and efficient output and R-CNN generates mean average precision (mAP) as 53.3%.

III. ANALYSIS

In this paper different deep learning techniques are discussed for image sentiment analysis including DNN, CNN, R-CNN, and Fast R-CNN.

This work analyzes that CNN is far efficient and accurate then DNN, R-CNN and Fast R-CNN for image sentiment analysis. As Neural Network and DNN with a large number of hidden layers will also increase cost, which is not efficient. CNN is performing better and generate the optimum result for analysis and increase almost 20 percent accuracy and with less number of features requirements [11]. The time to train the model will also reduce with the help of CNN. Fast R-CNN is more efficient then R-CNN as it occupies less space and takes less time to process as in fast R-CNN no need to input the 2000 regions every time.

The data set available and being used by the researchers for image sentiment analysis are; SUN database2 which is first large-scale scene attribute database containing more than 800 categories and 14,340 images as well as discriminative attributes labeled by crowd source human studies. The other datasets are Flickr dataset, Twitter testing dataset (consists of image tweets), standard Twitter dataset, SentiBank

Twitter Dataset6, MOUD dataset , Multimodel Opinion- Level Sentiment Intensity (MOSI) data set etc.

TABLE I. COMPARATIVE ANALYSIS OF DEEP LEARNING APPROACHES BASED IMAGE SENTIMENT ANALYSIS

Researchers Name and Year	*Model Used*	*Purpose*	*Data Set Used*	*Results*
J. Wang, J. Fu, Y. Xu, and T. Mei (2016)	DCAN	To reduce the amount of inter-class variance	Twitter and SentiBank	The system can efficiently get trained on mid-level sentiment from noisy web images
Q. T. Ain, M. Ali, and et al. (2017)	DNN, CNN and RNN	To get a review for sentiment analysis techniques	-----	Deep Learning is an optimum choice for sentiment analysis
C. Xu, S. Cetintas, K. C. Lee, and L. J. Li. (2014)	DCNN	Visual Sentiment Prediction	Twitter And Tumbler	SentiBank is more powerful than Low-Level Visual Feature
T. Chen, D. Borth, and S. F. Chang (2014)	DCNN	Visual Sentiment Concept Classification	Flicker images	CNN based approach is more accurate than SVM
S. Jindal, and S. Singh (2015)	DCNN	Image Sentiment Analysis with fine tuning	Flickr	Using CNN one can achieve high accuracy
Q. You, J. Luo, and J. Yang (2015)	CNN and Progressive CNN	Robost Image Sentiment Analysis	Flickr image from SentiBank	PCNN shows better and efficient result as compared to CNN
G. Cai, and B. Xia (2015)	CNN	Multimedia Sentiment Analysis	Flickr image from SentiBank & from Twitter	Multi CNN have performed better in all
V. Bharadi, A. I. Mukadam, and et al. (2018)	CNN	Image classification	Images by Digital camera or from databases	CNN is better choice for image classification
J. Mandhyani, L. Khatri, and et al. (2017)	CNN	Image Sentiment Analysis and categorizing emotions.	Images from SentiBank	R-CNN is better and generates 53.3% (mAP)
J. Yang, D. She, Rosin, and L. Wang (2018)	R-CNN	Visual Sentiment Prediction	IAPS, ArtPhoto, Twitter, Flickr, Instagram	The proposed method outperforms the methods on the popular affective datasets

IV. CONCLUSION

Image Sentiment Analysis is one of the imperative research areas for study, as now people are more used to of visual data to converse. Inspired by the arising issue of image sentiment analysis and its promising solution through deep learning techniques, in this paper we addressed some of the significant studies executed in past for image sentiment analysis using deep learning techniques. Further, as the research work in this area is being performed it is expected that soon researchers will propose more efficient techniques to get optimum results.

REFERENCES

[1] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman. "Sentiment analysis using deep learning techniques: a review" Int J AdvComputSciAppl, vol. 8 issue 6, 424, 2017.

[2] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. " Deep learning " vol. 1.Cambridge: MIT Press. 2016.

[3] C. Xu, S. Cetintas, K. C. Lee, and L. J. Li. "Visual sentiment prediction with deep convolutional neural networks". arXiv preprint arXiv:1411.5731, 2014.

[4] V. Bharadi, A. I. Mukadam, M. N. Panchbhai and N. N. Rode, "Image Classification Using Deep Learning". 2018

[5] J. Wang, J. Fu, Y. Xu, and T. Mei, "Beyond Object Recognition: Visual Sentiment Analysis with Deep Coupled Adjective and Noun Neural Networks". In IJCAI pp. 3484-3490., 2016.

[6] G. Cai, and B. Xia, "Convolutional neural networks for multimedia sentiment analysis". In Natural Language Processing and Chinese Computing pp. 159-167. Springer, Cham, 2015.

[7] S. Jindal, and S. Singh, "Image sentiment analysis using deep convolutional neural networks with domain-specific fine tuning". In Information Processing (ICIP), 2015 International Conference on pp. 447-451, IEEE, 2015.

[8] J. Mandhyani, L. Khatri, V. Ludhrani, V., Nagdev, and S. Sahu, "Image Sentiment Analysis". International Journal of Engineering Science, 4566

[9] Q. You, J. Luo, H. Jin, and J. Yang, "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks". In AAAI pp. 381-388, 2015.

[10] T. Chen, D. Borth, T. Darrell, and S. F. Chang, "Deepsentibank: Visual sentiment concept classification with deep convolutional neural network". arXiv preprint arXiv:1410.8586, 2014

[11] C. Wang, and Y Xi, Convolutional Neural Network for Image Classification". *Johns Hopkins University Baltimore, MD*, *21218*.

[12] V. Gajrala and A. Gupta, " Emotion detection and sentiment analysis of Images". Georgia Institute of Technology, 2015.

[13] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation". IEEE transactions on pattern analysis and machine intelligence, 38(1), 142-158, 2016.

[14] R. Girshick, Fast R-CNN. In: IEEE International Conference on Computer Vision (ICCV), 2015.

[15] V. Campos, B. Jou, and X. Giro-i-Nieto, "From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction. Image and Vision Computing", 65, 15-22, 2017.

[16] J. Yang, D. She, M. Sun, M. M. Cheng, P. Rosin, and L. Wang. "Visual sentiment prediction based on automatic discovery of affective regions". IEEE Transactions on Multimedia, 2018.

[17] R. Dalai, and K. K. Senapati, "Comparison of Various RCNN techniques for Classification of Object from Image". International Research Journal of Engineering and Technology (IRJET), vol. 4 issue 07, 2017.

[18] S.Ren, K. He, R. Dirshick and J. Sun. "Faster R-CNN: Towards real time object detection with region proposal networks". In advances in neural information processing systems (pp. 91-99), 2015.