Face Mood Detection using Deep Learning Approach

Mrs. S.A.Shete[1] Mr. Vinayak Kulkarni[2] Mr. Abhishek Kadam[3] Mr. Shubham Alane[4] Mr. Ajit Waghmare[5]

[1]Lecturer [2,3,4,5]Student

[1,2,3,4,5]Department of Information Technology

[1,2,3,4,5]AISSMS Polytechnic, Pune-01, India

Abstract— The objective of emotion recognition is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. In this work we focus on identifying human emotion from facial expressions. Facial emotion recognition is one of the useful task and can be used as a base for many real-time applications. It can be used as a part of many interesting and useful applications like Monitoring security, treating patients in medical field, marketing research, E-learning etc;. We humans can easily identify the emotion of other humans without any effort. Automatic detection of emotion of a human face is important due to its use in real-time applications. The recent advance in GPU has taken many applications like face recognition, hand written digit recognition and object recognition to the next level. Especially the pretrained CNN based features better represent the images. Pretrained CNN features represent the most discriminative features and hence allows for better performance. Feature representation plays a major role on the performance of any machine learning algorithm. After observing unbelievable performance with deep learning models, we propose to use the deep convolutional features to better represent the given image instead of using the traditional handcrafted features. The downside of the deep learning models is that they require large datasets to obtain better performance. To leverage the use of deep learning models without the requirement of large datasets is to use pre-trained models. For feature extraction pre-trained Convolutional Neural Networks model (VGG16) is used and the concept of Deep Neural Networks model is used for classification. To show the performance of the proposed model, various face image dataset is used for the experimental studies. Based on the experimental results we claim that these unsupervised features better represent the images compared the handcrafted features.

Keywords: Facial Emotion Recognition, CNN, RBF Kernels, Extended Cohn-Kanade, Multi-Class SVM

I. Introduction

The objective of face emotion recognition (FER) is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. Psychological characteristics such as heartbeat and blood Pressure, speech, hand gestures, body movements, Facial expressions identify emotions of a person. The use of machines in society has increased widely in the last decades. Nowadays, machines are used in many different industries. As their exposure with human's increase, the interaction also has to become smoother and more natural. In order to achieve this, machines have to be provided with a capability that let them understand the surrounding environment. Specially, the intentions of a human being. When machines are referred, this term comprises to computers and robots. A distinction between both is that robots involve interaction abilities into a more advanced extent since their design involves some degree of autonomy. When machines are able to appreciate their surroundings, some sort of machine perception has been developed. Humans use their senses to gain insights about their environment. Therefore, machine perception aims to mimic human senses in order to interact with their environment. Nowadays, machines have several ways to capture their environment state trough cameras and sensors. Hence, using this information with suitable algorithms allow to generate machine perception. In the last years, the use of Deep Learning algorithms has been proven to be very successful in this regard. For instance, Jeremy Howard showed on his Brussels 2014 TEDx's talk how computers trained using deep learning techniques were able to achieve some amazing tasks. These tasks include the ability to learn Chinese language, to recognize objects in images and to help on medical diagnosis. Affective computing claims that emotion detection is necessary for machines to better serve their purpose. For example, the use of robots in areas such as elderly care or as porters in hospitals demand a deep understanding of the environment. Facial emotions deliver information about the subject's inner state. If a machine is able to obtain a sequence of facial images, then the use of deep learning techniques would help machines to be aware of their interlocutor's mood. In this context, deep learning has the potential to become a key factor to build better interaction between humans and machines, while providing machines with some kind of self-awareness about its human peers, and how to improve its communication with natural intelligence.

II. Literature Survey

According to [1] Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning by Stuti Jindal and Sanjay Singh 2015 International Conference on Information Processing (ICIP). In this work, an image sentiment prediction framework is built with Convolutional Neural Networks (CNN). Specifically, this framework is pretrained on a large scale data for object recognition to further perform transfer learning. Extensive experiments were conducted on manually labeled Flickr image dataset. To make use of such labeled data, we employ a progressive strategy of domain specific fine tuning of the deep network. The results show that the proposed CNN training can achieve better performance in image sentiment analysis than competing networks.

According to [2] Visual Sentiment Analysis for Social Images Using Transfer Learning Approach Jyoti Islam, Yanqing Zhang 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom). In this

paper, we propose a novel visual sentiment analysis framework using transfer learning approach to predict sentiment. We use hyperparameters learned from a very deep convolutional neural network to initialize our network model to prevent overfitting. We conduct extensive experiments on a Twitter image dataset and prove that our model achieves better performance than the current state-of-the-art.

According to [3] Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning by Malhar Anjaria, Ram Mahana Reddy Guddeti 978-1-4799-3635-9/14 2014 IEEE. In this paper, we introduce the novel approach of exploiting the user influence factor in order to predict the outcome of an election result. We also propose a hybrid approach of extracting opinion using direct and indirect features of Twitter data based on Support Vector Machines (SVM), Naive Bayes, Maximum Entropy and Artificial Neural Networks based supervised classifiers.

According to [4] Inferring Sentiment from Web Images with Joint Inference on Visual and Social Cues: A Regulated Matrix Factorization Approach by Yilin Wang, Yuheng Hu, Subbarao Kambhampati, Baoxin Li Proceedings of the Ninth International AAAI Conference on Web and Social Media. In this paper, we study the problem of understanding human sentiments from large scale collection of Internet images based on both image features and contextual social network information (such as friend comments and user description). Despite the great strides in analyzing user sentiment based on text information, the analysis of sentiment behind the image content has largely been ignored. Thus, we extend the significant advances in text-based sentiment prediction tasks to the higher level challenge of predicting the underlying sentiments behind the images.

According to [5] Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks by Yuhai Yu1, Hongfei Lin, Jiana Meng, † and Zhehuan Zhao. In this paper, we utilize deep learning models in a convolutional neural network (CNN) to analyze the sentiment in Chinese microblogs from both textual and visual content. We first train a CNN on top of pre-trained word vectors for textual sentiment analysis and employ a deep convolutional neural network (DNN) with generalized dropout for visual sentiment analysis. We then evaluate our sentiment prediction framework on a dataset collected from a famous Chinese social media network (Sina Weibo) that includes text and related images and demonstrate state-of-the-art results on this Chinese sentiment analysis benchmark. 6. Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks by Quanzeng You and Jiebo Luo Hailin Jin and Jianchao Yang. In this paper we first design a suitable CNN architecture for image sentiment analysis. We obtain half a million training samples by using a baseline sentiment algorithm to label Flickr images. To make use of such noisy machine labeled data, we employ a progressive strategy to fine-tune the deep network. Furthermore, we improve the performance on Twitter images by inducing domain transfer with a small number of manually labeled Twitter images. We have conducted extensive experiments on manually labeled Twitter images. The results show that the proposed CNN can achieve better performance in image sentiment analysis than competing algorithms.

According to [6] How do your friends on social media disclose your emotions? by Yang Yang, Jia Jia, Shumei Zhang, BoyaWu, Qicong Chen, Juanzi Li, Chunxiao Xing, Jie Tang. In this paper, we formally formalize the problem and propose a novel emotion learning method by jointly modeling images posted by social users and comments added by their friends. One advantage of the model is that it can distinguish those comments that are closely related to the emotion expression for an image from the other irrelevant ones. Experiments on an open Flickr dataset show that the proposed model can significantly improve (+37.4% by F1) the accuracy for inferring user emotions. More interestingly, we found that half of the improvements are due to interactions between 1.0% of the closest friends. 8. DeViSE: A Deep Visual-Semantic Embedding Model: by Andrea Frome, Greg S. Corrado, Jonathon Shlens, Samy Bengio Jeffrey Dean, Marc'Aurelio Ranzato, Tomas Mikolov. In this paper we present a new deep visualsemantic embedding model trained to identify visual objects using both labeled image data as well as semantic information gleaned from unannotated text. We demonstrate that this model matches state-of-the-art performance on the 1000-class ImageNet object recognition challenge while making more semantically reasonable errors, and also show that the semantic information can be exploited to make predictions about tens of thousands of image labels not observed during training. Semantic knowledge improves such zero-shot predictions achieving hit rates of up to 18% across thousands of novel labels never seen by the visual model.

According to [9] Sentiment Analysis for Social Media Images by Yilin Wang and Baoxin Li, In this proposal, we study the problem of understanding human sentiments from large scale collection of Internet images based on both image features and contextual social network information (such as friend comments and user description). Despite the great strides in analyzing user sentiment based on text information, the analysis of sentiment behind the image content has largely been ignored. we extend the significant advances in text-based sentiment prediction tasks to the higher-level challenge of predicting the underlying sentiments behind the images. We show that neither visual features nor the textual features are by themselves sufficient for accurate sentiment labeling. Thus, we provide away of using both of them, and formulate sentiment prediction problem in two scenarios: supervised and unsupervised.

III. PROPOSED SYSTEM DESIGN

Initially the face region is extracted from the given face images using viola jones face detection algorithm. Then from the cropped face regions, deep features are extracted from the last fully connected layer of VGG16. Multi-class SVM based classifier is used for classification. Based on the experimental results we observed that the proposed method is simple and gives better performance compared to the other methods.

Recently, social media users based life clients are progressively utilizing images and recordings to express

their emotions and offer their encounters. Emotions identification of such huge scale visual substance can assist better with extricating client assessments toward occasions or subjects, for example, those in image tweets, so expectation features extracted from visual substance is correlative to printed sentiment investigation. It is definitely worth much more with regards to pass on human sentiments. The assumptions models that help this are bounteous: incredible dazzling images frequently contain rich information contains that help view effectively interface with those images. With the appearance of internet based life, an expanding number of individuals begin to utilize images to express their feelings, opinion, and fatigue via web-based networking media stages like Flickr and Instagram. Programmed induction of the sentiment and estimation data from such consistently developing, huge measures of client produced images is of expanding significance to numerous applications in medicinal services, human sciences, correspondence studies, advertising, and many sub-territories inside software engineering, for example, computer vision.

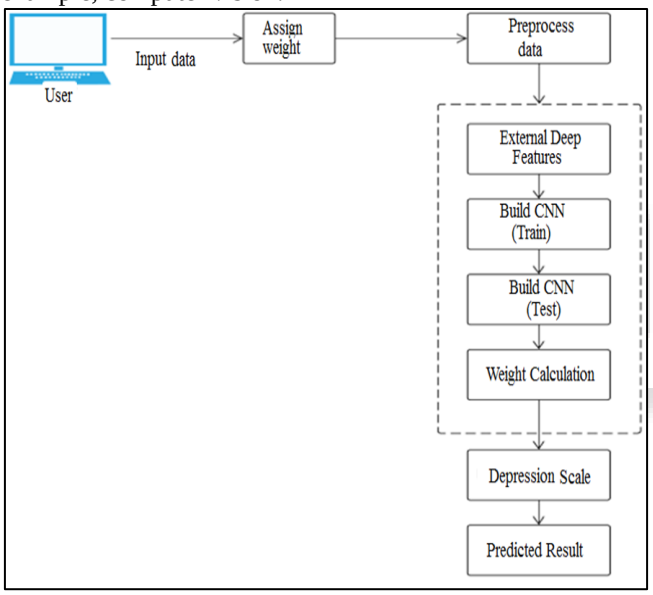

Fig. 1: system basic overview

The above Figure 1 shows the overall architecture of the proposed visual sentiment prediction framework. The CNN is first trained on a large scale dataset (ImageNet) for image classification. The parameters of pre-trained layers are transferred to the problem of sentiment prediction for generating image representations using domain specific fine tuning. Individuals share a great deal of substance via web-based networking media as images - be it individual, or ordinary scenes, or their sentiments portrayed as kid's shows or images. Breaking down substance like this from internet based life sites and additionally photograph sharing sites like Flickr, Twitter, Facebook, and so forth. It can give bits of knowledge into the general estimation of individuals about state Presidential races. Likewise, it is valuable to comprehend the sentiment an image delineates to consequently anticipate enthusiastic labels on them - like bliss, dread, and so forth. As a piece of this task, It intends to foresee the passionate class an image falls into from 5 classifications - Love, Happiness, Violence, Fear, and Sadness. This is finished by tweaking 3 distinctive convolutional neural systems for the undertakings of sentiment identification and notion process.

VGG-16, a pretrained convolution model, [15] is known for its state of the art performance in various applications [15] like image classification, object detection etc.; It is trained on ImageNet dataset. VGG16 consists of a total of 13 layers with 5 convolution layers interleaved with pooling layers. Features are extracted from the FC2 layer of VGG16, which gives 2048 dimensional feature vector for each image. Static feature representation is used where all the images are represented with equal size.

IV. Dataset Used

A. Deep Learning on Flickr Dataset

For this research we can used Flicker dataset like dynamically select 70-75% images from the half million Flickr dataset which is consider for training dataset to build the rain module according to select features. The continuing 30-25% images considered as testing dataset. So we can train the convolutional Neural Network using number of iteration and each iteration should be hold n number of images respectively.

B. Twitter Testing Dataset (Real time dataset)

We can also build some real time image dataset from various social media web application as well as image tweets. Image tweets denote to persons tweets that contain some input images. We constructed a large set as testing images for entire research to validate system accuracy with real time dataset.

V. Observations

Our study mirrors that Deep Learning gives promising outcomes in both the arrangement of classification just as in performing assessment investigation, even on the raw data gathered straightforwardly from Flickr. The following stage could be to run these investigations on greater and cleaner datasets and check whether they would improve the outcomes. We accept they would. It would likewise be fascinating to perceive what the remarkable locales in the images are, maybe by utilizing [6], and feed these pre-processed images to pre-trained neural systems. The thought behind this is people would choose the feeling distinguished by a picture by taking a gander at some striking locales in the picture. By advising the system of these already may help in the preparation procedure. Our exploration show that profound learning provides promising outcomes with a presentation practically identical to certain strategies utilizing carefully assembled highlights on feeling characterization task, and furthermore a couple of techniques utilizing profound learning for conclusion investigation. Assumption order in images has applications in programmed labelling of images with passionate classes, consequently classifying video groupings into sorts with emotional categories, automatically categorizing video sequences into genres like thriller, comedy, romance, etc.

VI. CONCLUSION

Emotion classification in images has applications in automatic tagging of images with emotional categories, automatically categorizing video sequences into genres like thriller, comedy, romance, etc. According to literature study deep learning does provide promising results with a performance comparable to some methods using handcrafted features on emotion classification task, and also a few methods using deep learning for sentiment analysis. For the goal of autonomous accuracy is much important during the execution more categories of image in each class for train the module which could be active higher accuracy of system.

REFERENCES

[1] Stuti Jindal and Sanjay Singh "Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning", International Conference on Information Processing (ICIP) 2015.
[2] Jyoti Islam, Yanqing Zhang "Visual Sentiment Analysis for Social Images Using Transfer Learning Approach" IEEE International Conferences on Big Data and Cloud Computing 2016.
[3] Malhar Anjaria, Ram Mahana Reddy Guddeti "Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning" 978-1-4799-3635-9/14 2014 IEEE.
[4] Y. Wang, Y. Hu, S. Kambhampati, and B. Li, "Inferring sentiment from web images with joint inference on visual and social cues: A regulated matrix factorization approach." in ICWSM, 2015, p. 21.
[5] Yuhai Yu1, Hongfei Lin, Jiana Meng, † and Zhehuan Zhao "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks."
[6] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," 2015.
[7] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, and J. Tang, "How do your friends on social media disclose your emotions?" in Proc. AAAI Conf. Artificial Intelligence (AAAI), 2014, pp. 306–312.
[8] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, M. A. Ranzato, and T. Mikolov, "DeViSE: A deep visual-semantic embedding model," in Proc. Advances in Neural Information Processing Systems (NIPS), 2013, pp. 2121–2129.
[9] Yilin Wang and Baoxin Li "Sentiment Analysis for Social Media Images."

