Comparative study of Emotion recognition systems

sagar , nishanth
King's College London, UK
sagar.joglekar, nishanth.sastry}@kcl.ac.uk

1. ABSTRACT

Emotions are a very fundamental part of how humans communicate. A major context of what we say is communicated using our body language and our facial expressions. Our paper tries to benchmark several expression recognition systems in the wild, using comparative studies on datasets and sample data. We also experiment the problem of emotion recognition using asymmetric convolutional network designs and show that mutually complimentary networks can be trained to improve the overall recognition performance of the problem. We further try to understand analytically what can make an ideal emotion recognition system and propose a combined approach. ‘

2. INTRODUCTION

Online social networks (OSNs) have seen a massive surge in usage over the past decade. The surge is also going hand in hand with the explosion of smart phone industry. More and more social interactions are now driven by media content like selfies and group selfies because of the ubiquitous nature of cameras. A sharp change in cultural aspects of online social interactions are evident and have also been studied in detail in papers like [8]. Naturally, properties of real life interactions are now going to emerge in virtual social interactions. A study by Daniel Perez et.al [6] observes that non-verbal communication has a strong effect on how a person is perceived at times of real world situations like interviews. These real world behaviours can also be extended to the virtual social world. With the rise of social media networks like Vine and Instagram, human to human non-verbal interactions have another dimension to manifest. The [8] paper explores several of these properties of Instagram users, where they explore correlation of facial orientation, poses and smiles with parameters like country of origin of the selfie user, post frequency, likes received , number of faces in the pictures, gender and smile scores. Such studies give us interesting insights about the sharp rising phenomenon of selfies. However, there is another dimension that could be of potential use to understand the psychological and social reasons behind the posters. That dimension is sentiment. Sentiment is a emergent quantity, which can be quantified in terms of emotion, ontology and context.

WWW 2015, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3469-3/15/05.
http://dx.doi.org/10.1145/2736277.2741671.

3. A SOCIAL APPROACH TO SENTIMENTS:

Sentiments are fundamental part of our day to day social interactions. A face to face social interaction is generally augmented with facial expression, body language and linguistic sentiment to convey the exact meta information. These properties are very human in nature and are mimicked in the social networks as well. Studies like [4] have explored the world of linguistic sentiment in social networks, by comparing several popular textual sentiment analysis methods used for analysing tweets. Our paper tries to benchmark methods that explore the sentiment of a human face. This is an effective way to approach a problem as human faces are dominant factors of social media of the day. With phenomenons like selfies and video selfies on the rise, our hypothesis is that facial expressions play a major role in expressing the overall sentiment of a social media.

When it comes to perceptual sentiments, there are two broad categories that could be explored. The first category looks at the perceptual sentiment evoked by a social media content. The second category talks about the actual latent perceptual sentiment that comes with the context of the content itself. We will discuss about the research problems about both these categories.

3.1 Evoked perceptual sentiment

Several works have done in depth studies using methods like crowdsourcing to understand the different shades of a particular evoked emotion. Works like UrbanGems [1] and StreetScore [7] use crowdsourcing methods to understand degrees of human sentiment evoked because of pictures of real urban neighbourhoods. Sentiments like the feeling of safety and aesthetics are especially hard to quantify and crowdsourcing helps the authors to do some interesting modelling. On the other hand there are papers like [3] by L. Jeni et.al. describe utility of actual facial expression detection for understanding content consumer reaction. Such approaches help us understand the very effect of a particular content on the consumer.

3.2 Latent perceptual sentiment

This approach is what this paper stresses on. By latent perception, we mean the hidden parameters, which are part of the very content. Social networks like reddit have specific sub-reddits that work on appealing to these types media sentiments that evoke emotions like empathy, disgust, contempt and love . One such popular sub-reddit is labelled R/aww which contains images and GIFs that showcase cute animals and animal behaviours. Another one called R/cringe appeals to the sentiments of awkwardness and discomfort by exhibiting videos and Gifs about people in awkward situations. These specific social channels are popular because the content shared over these channels have a certain type of latent sen-

timental response, which the consumers of these channels resonate with.

Our paper focuses on this part of the story, and tries to survey and benchmark certain state of the art methodologies out there. We also propose certain hybrid approaches, which show that we can attain much better performance if a heuristic approach to combine certain methods is taken.

4. SENTIMENT ANALYSIS METHODS

To the best of our knowledge we have evaluated certain popular approaches in solving the problem of extracting latent sentiment in a media content. The sentiment analysis methods broadly fall into two bins. One is the Content based Image retrieval (CBIR) [5] set of approaches, which actually analyse the image structure and contents to extract features and inferences about the image. The second bin is emotional semantic image retrieval (ESIR) [9] which aim at trying to extract the semantic gist of a particular image. Human brain is great at extracting such semantics. For example it is very natural for a person to describe a particular image as "picturesque" or "scenic" or to describe someone's clothing as "tacky" , "classy" or "elegant". These semantic classes, no matter how subjective, are also sufficiently descriptive for another human being to process. In the subsections to come, we will discuss some of the popular perceptual sentiment analysis methods.

4.1 SENTIBANK

Sentibank [2] is a method that proposes a Visual Sentiment Ontological approach towards image perceptual sentiment retrieval. The method tries to match an image with an Adjective noun pairs which give a visual ontology about the structure of an image. The pairs are extracted using trained detectors which train on a dataset acquired from Flickr images. The adjective noun pair concept labels are verified using Mechanical Turk. The result is

5. REFERENCES

[1] ADAM BARWELL, DANIELE QUERCIA, J. C. http://www.cam.ac.uk/research/news/how-to crowdsource-your-happy-space, 2012.

[2] BORTH, D., JI, R., CHEN, T., BREUEL, T., AND CHANG, S.-F. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM International Conference on Multimedia* (New York, NY, USA, 2013), MM '13, ACM, pp. 223–232.

[3] JENI, L. A., LŐRINCZ, A., NAGY, T., PALOTAI, Z., SEBŐK, J., SZABÓ, Z., AND TAKÁCS, D. 3d shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing 30*, 10 (2012), 785–795.

[4] JOO, J., LI, W., STEEN, F. F., AND ZHU, S. C. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 216–223.

[5] LIU, Y., ZHANG, D., LU, G., AND MA, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition 40*, 1 (2007), 262 – 282.

[6] MARCOS-RAMIRO, A., PIZARRO, D., MARRON-ROMERA, M., AND GATICA-PEREZ, D. Let your body speak: Communicative cue extraction on natural interaction using rgbd data. *IEEE Transactions on Multimedia 17*, 10 (Oct 2015), 1721–1732.

[7] NAIK, N., PHILIPOOM, J., RASKAR, R., AND HIDALGO, C. Streetscore – predicting the perceived safety of one million streetscapes. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on* (June 2014), pp. 793–799.

[8] SOUZA, F., DE LAS CASAS, D., FLORES, V., YOUN, S., CHA, M., QUERCIA, D., AND ALMEIDA, V. Dawn of the selfie era: The whos, wheres, and hows of selfies on Instagram. In *Proceedings of the 2015 ACM on Conference on Online Social Networks - COSN '15* (2015), pp. 221–231.

[9] WANG, W., AND HE, Q. A survey on emotional semantic image retrieval. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* (Oct 2008), pp. 117–120.