6-2017

Machine Learning Approaches to Sentiment Analytics

Wangchuchu Zhao
Missouri University of Science and Technology, wzkt2@mst.edu

Keng L. Siau
Missouri University of Science and Technology, siauk@mst.edu

Machine Learning Approaches to Sentiment Analytics

Wangchuchu Zhao
Missouri University of Science and Technology
wzkt2@mst.edu

Keng Siau
Missouri University of Science and Technology
siauk@mst.edu

ABSTRACT

One key aspect of sentiment analytics is emotion classification. This research studies the use of machine learning approaches to classify human emotion. Two different machine learning approaches were compared in an experimental study. In one approach, emotions from both genders were used to train the machine. In another approach, genders were separated and two separate machines were used to learn the emotions of the two genders. We also manipulated the training sample sizes and study the effect of training sample sizes on the two machine learning approaches. Our preliminary results show that the approach where the genders were separated produces a higher accuracy in classifying emotions. We also observe that training sample sizes have different impact on the two approaches.

Keywords

Sentiment Analytics, Emotion classification, Machine Learning

INTRODUCTION

Sentiment analytics (Pang & Lee, 2008; Adeborna & Siau, 2014; Yuan & Siau, 2017) is becoming an increasingly appealing exercise because of large amount of Internet data (Lee & Siau, 2001; Cambria, Schuller). One important dimension of sentiment analytics is emotion classification. Emotion classification has been increasingly used in various domains such as computer science (Peng, Zhu, Zheng, & Lu, 2014), multimedia (Soleymani, Pantic, & Pun, 2012), marketing (Rao, 2016), and commercial area (Feldman, 2013). Correctly classifying emotions can have huge impact on businesses. For example, marketing messages and advertisements can be tailored to customers if their emotional state can be identified in real-time.

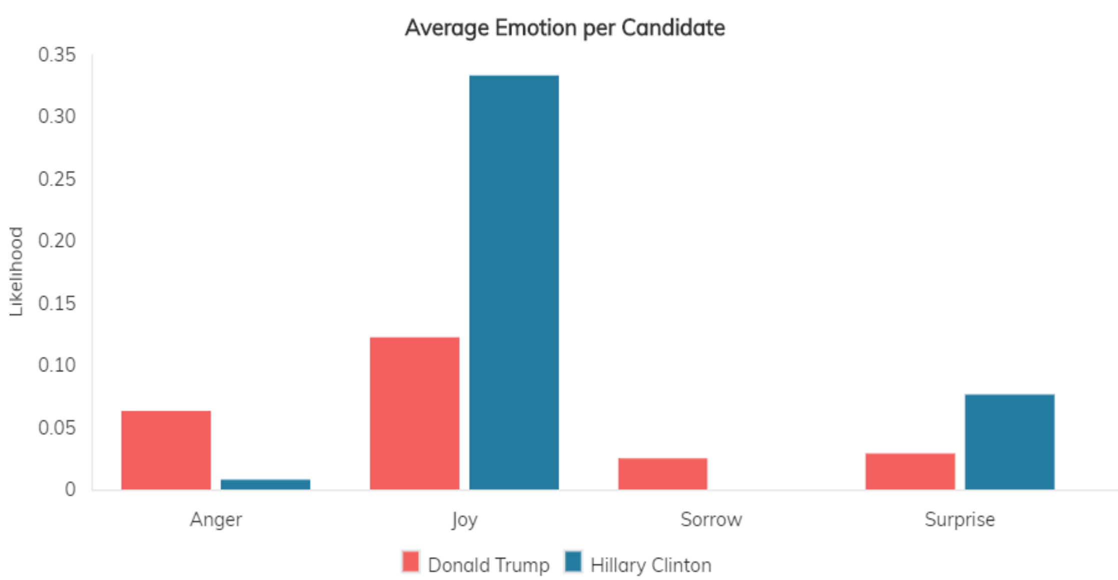

Figure 1. Emotion identified using facial expression of Clinton vs Trump (http://wonderif.pt/emotions_us_elections)

In Figure 1, emotion analysis of the facial expressions of the two presidential candidates has been performed on 205 photos (102 for Hillary, 103 for Trump) during public speeches collected from Google Images. At one glance, we notice the differences between the two candidates' emotions detected by Google Cloud Vision (https://cloud.google.com/vision/), and the results are similar in Figure 2. As the software does not distinguish between the male and female group during the analysis, the gender difference may affect the precision of the analysis.

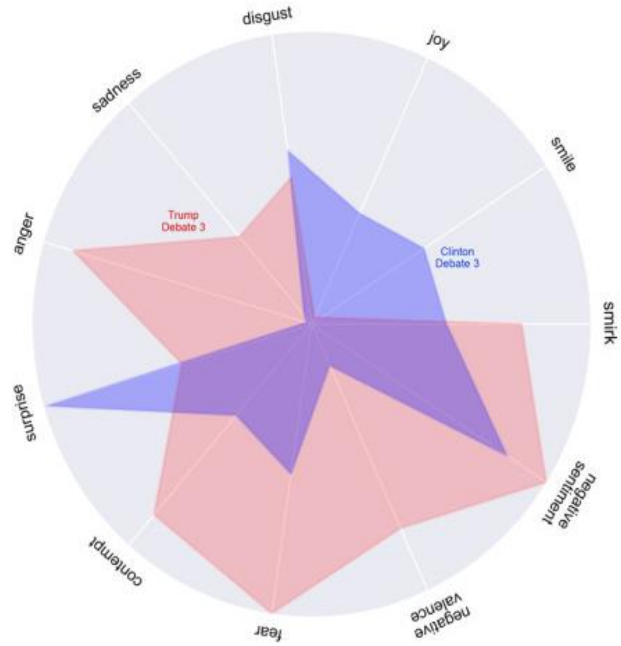

Figure 2. Emotion of Clinton vs Trump (https://fast-company-res.cloudinary.com/image/upload/w_623,c_fill,g_face/v1477064796/fastconews/y14ud3idgif0sv1zgbne.jpg)

In this research, we hypothesize that separating the two genders in the emotion classification process will produce a more accurate classification of emotions. We also hypothesize that separating the two genders in the emotion classification process will require a smaller training set in machine learning than the approach where the two genders were not separated. A smaller training set will be beneficial in real-life applications.

LITERATURE REVIEW AND THEORETICAL FOUNDATION

There are three main steps in visual image analysis -- detecting the objects, tracking the objects, and evaluating tracking results to describe and infer semantic events and latent phenomenon (Cazzato, Leo, & Distante, 2014). Moravec worked on stereo matching and he came up with the Moravec detector in 1981 (Moravec, 1981). In 1988, Harris and Stephens refined the processing and made image matching more repeatable for small image variations and corners (Harris, 1988). In 1995, correlation window was used to select matches and remove outliers for large image range (Zhang, Deriche, Faugeras, & Luong, 1995). In 1997, Schmid and Mohr used rotationally invariant descriptor for image recognition (Schmid & Mohr, 1997). With the advancement in image recognition, automatic recognition and classification of images using image recognition technology become possible and viable. Emotion recognition and classification are the most typical forms of facial expression analyses on images because a facial expression can reflect a human being's emotion and sentiment (Ekman, 1993). Further, a picture is worth ten thousand words (Siau 2005).

In this research, we argue that separating the two genders during machine learning and emotion analysis will result in better emotion classification. One reason is the emotional differences between genders, which are well recognized and documented. In the book, On the Origin of Species, Darwin proposed that psychological traits, like physical traits, evolve through the process of gender selection and result in the differences between genders (Darwin, 1859). Previous studies concluded that women are more emotional compared to men (Barrett, Robin, Pietromonaco, & Eyssell, 1998) and are more willing to express their emotion than men (Allen & Haccoun, 1976). Further, women express stronger emotions (Ashmore & Sewell, 1998). Also, gender difference is found in the emotion in social network using data mining (Thelwall, Wikinson, & Uppal, 2010). Using fMRI, an inevitable gender gap in emotion recognition is found (McRae, Ochsner, Mauss, Gabrieli, & Gross, 2008).

In this research, we compared two different machine learning approaches to emotion classification.

RESEARCH METHODOLOGY AND OPERATIONALIZATION

Based on the evolution theory (Darwin & Peckham, 1959), females are more emotionally expressive than males. Thus, separating machine learning into male set and female set should produce more accurate emotion classification than clustering male and female expressions into a single set. Further, we hypothesize that classifying female emotions will be more accurate than classifying male emotions as females are more expressive. The research model is shown in Figure 3. The approach with "No Gender Separated Dataset" represents many existing techniques to classifying emotions. The approach, "Separated by Gender Dataset", is the approach proposed in this research. In the first approach, the machine learning algorithm is trained using both male and female facial expressions and emotions. In the second approach, we suggest that the subject's gender should be identified first and then the appropriate gender emotion algorithm is applied based on the gender. Size of the training set will moderate the emotion classification outcome.

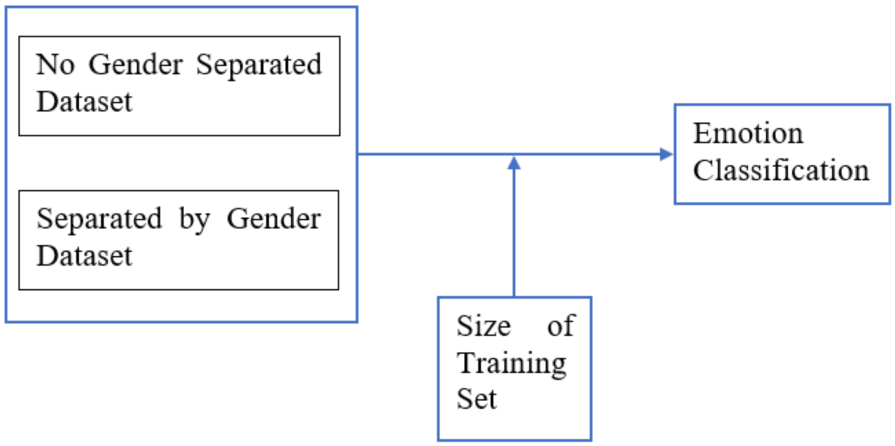

Figure 3. Research Model

H1: Human frontal facial expressions (consisting of female and male facial expressions) can positively determine the subjects' emotions.

H2: Frontal facial expressions of female group can positively determine the subjects' emotions.

H2a: The results from the female group are more accurate than the combined group.

H3: Frontal facial expressions of the male group can positively determine the subjects' emotions.

H3a: The results from the male group are more accurate than the combined group.

H3b: The results from the female group are more accurate than the male group.

H4: Size of the training set will moderate the emotion classification results.

The research is operationalized as follows: The subjects' emotions are analyzed using their facial expressions. We use the databases from Cohn-Kanade, which provide images of frontal face emotions. For the machine learning algorithm, we use a pre-coded function packaged in OpenCV, named Fisherfaces (van Gent, 2016). Fisherfaces uses facial images to calculate and form an eigenface corresponding to the least-square solution (Martinez, 2011). The approach of using Eigenfaces for recognition was developed by Sirovich and Kirby (Sirovich & Kirby, 1987), and Eigenfaces is the name given to a set of eigenvectors. Emotion classification can be determined by comparing how faces are represented to the base set.

PRILIMINARY RESULTS

Our preliminary results show that there are differences between the two genders. Our preliminary results also show that emotion classification is more accurate when the male and female sets are separated than when the sets are combined (i.e., both males and females in one set). The preliminary results also show that the size of training sets has an effect on emotion classification. When the machine learning is separated by gender, a smaller training set is required to achieve equivalent emotion classification results than when the combined gender set is used for machine learning.

REFERENCES

1. Adeborna, E., Siau, K. (2014). An Approach to Sentiment Analysis – The Case of Airline Quality Rating. *Paper presented at the Pacific Asia Conference on Information Systems (PACIS 2014)*, Chengdu, China.
2. Allen, J. G., & Haccoun, D. M. (1976). Sex Differences in Emotionality: A Multidimensional Approach. Human Relations, 29(8), 711-722. doi: doi:10.1177/001872677602900801
3. Ashmore, R. D., & Sewell, A. D. (1998). Sex/Gender and the Individual. In D. F. Barone, M. Hersen, & V. B. Van Hasselt (Eds.), Advanced Personality (pp. 377-408). Boston, MA: Springer US.
4. Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are Women the “More Emotional” Sex? Evidence from Emotional Experiences in Social Context. Cognition and Emotion, 12(4), 555-578. doi:10.1080/026999398379565
5. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 15-21.
6. Cazzato, D., Leo, M., & Distante, C. (2014). An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation. *Sensors (Basel), 14*(5), 8363-8379. doi:10.3390/s140508363
7. Darwin, C. (1859). On the origin of species; On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life [microform]. London: J. Murray.
8. Darwin, C., & Peckham, M. (1959). The origin of species: a variorum text. Philadelphia Pa.: University of Pennsylvania Press.
9. Ekman, P. (1993). Facial expression and emotion. Am Psychol, 48(4), 384-392.
10. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 82-89.
11. Harris, C. S., M. (1988). A Combined Corner and Edge Detector. Paper presented at the Proceedings of the 4th Alvey Vision Conference.
12. Lee, S., Siau, K. (2001). A Review of Data Mining Techniques. *Industrial Management & Data Systems*, 101(1), 41-46.
13. Martinez, A. (2011). Fisherfaces, Scholarpedia, 6, 4282.
14. McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. *Group processes & intergroup relations*, 143-162.
15. Moravec, H. P. (1981). Rover visual obstacle avoidance. Paper presented at the 7th international joint conference on Artificial intelligence - Volume 2, Vancouver, BC, Canada.
16. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 1-135.
17. Peng, Y., Zhu, J. Y., Zheng, W. L., & Lu, B. L. (2014). EEG-based emotion recognition with manifold regularized extreme learning machine. Conf Proc IEEE Eng Med Biol Soc, 2014, 974-977. doi:10.1109/EMBC.2014.6943755
18. Rao, Y. (2016). Contextual Sentiment Topic Model for Adaptive Social Emotion Classification. IEEE Intelligent Systems, 31(1), 41-47. doi:10.1109/MIS.2015.91
19. Schmid, C., & Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(5), 530-535. doi:10.1109/34.589215
20. Siau, K. (2005). Human-Computer Interaction–The Effect of Application Domain Knowledge on Icon Interpretation. *Journal of Computer Information Systems*, 45(3), 53-62.
21. Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4(3), 519-524. doi:10.1364/JOSAA.4.000519
22. Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal Emotion Recognition in Response to Videos. IEEE Transactions on Affective Computing, 3(2), 211-223. doi:10.1109/T-AFFC.2011.37
23. Thelwall, M., Wikinson, D., & Uppal, S. (2010). Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the Association for Information Science and Technology*, 190-199.

24. van Gent, P. (2016, April 1). Emotion Recognition with Python, OpenCV and a Face Dataset. Retrieved from A tech blog about fun things with Python and embedded electronics.: http://www.paulvangent.com/2016/04/01/emotion-recognition-with-python-opencv-and-a-face-dataset/
25. Yuan, B., Siau, K. (2017). Lexicons in Sentiment Analytics. Paper presented at the *Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017)*, Springfield, Illinois.
26. Zhang, Z., Deriche, R., Faugeras, O., & Luong, Q.-T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence, 78*(1), 87-119. doi:http://dx.doi.org/10.1016/0004-3702(95)00022-4