IDENTIFYING ANGER BASED ON SOCIAL SIGNALS

Daniel Istifanus
Ethan Yu
Michael Zhou

ABSTRACT

Identifying different types of anger being displayed by individuals is an important social skill as it often dictates how to socially respond to the angry individual. In terms of human-robot interactions, robots need to be able to effectively identify whether the person is displaying loud anger or subtle anger as the behaviour of the robot should be dependent on the identification of anger. In this study, we aim to create and evaluate various models in how effectively they are able to distinguish between loud and subtle anger. The models will be trained through acted and naturalistic labeled images displaying anger which allows the model to be applied in real-world scenarios. Our work would provide an extension to existing human-robot interaction systems to be able to identify the type of anger that the individual is expressing.

1 INTRODUCTION

During the 1900s, Ekman hypothesized the basic emotion theory. The theory states that humans have 6 central emotions[2]. These emotions have underlying biological and psychological functions that are expressed through behavioural mechanisms. Ekman believes that these emotions were maintained throughout natural selection because their social functions are vital to our survival as a species[2]. These emotions are Happiness, Anger, Sadness, Surprise, Fear, and Disgust. Currently, the literature finds it difficult to agree on the number of basic emotions we display[2]. Different scientists have found similarities between emotions like anger and disgust — raised eyebrows, and wrinkled nose, but evidently, there are clear social and psychological differences in the ways these emotions are expressed. Some researchers believe these differences arose to fulfill social needs rather than contribute to survival[2].

These social needs are even more diverse across cultures. Due to diversity in cultures among humans, the way social signals may be interpreted may vary from one culture to another[4]. In this project, we are focusing on Anger.

Emotions like anger combine with culture to provide various and intricate emotional experiences[4]. Different cultures use and express anger in different ways. These complexities in social signals increase the likelihood of social chaos and conflict in a globalized world[4].

In 2006, Ekman obtained judgment data that illustrated how anger was expressed, with over 11 variants from approximately 50,000 individuals around the world[4]. It is easy to misjudge an emotional expression with many variants[4]. Only a small percentage of individuals were able to point out images with an obvious representation of anger. This research suggests that some aspects of anger expressions are universal but culturally influenced and other appearances of anger are culture specific[4].

In this project, we will be using machine learning algorithms to create a model that accurately differentiates different levels of anger using social signals. The goal of this project is to provide a non-biased method of identifying different forms of anger regardless of cultural differences. This can be used in implementing robots to improve their effectiveness and adaptiveness.

2 APPROACH

As our goal is to classify various levels of anger displayed in images, a dataset containing images of varying levels of anger for the classification model to work with was needed. We theorize that there are action units that are associated with varying levels of anger. Our theory is that passive anger 1 would contain action units 4 (brow lowerer), 5 (upper lid raiser), 7 (lid tightener) and more aggressive anger 2 would also include action unit 9 (nose wrinkler) and 23 (lip tightener) in addition to the action units described for passive anger. Another part of our theory is that the intensity of the action units would be significantly higher for aggressive anger. This was one of the main bases we believe can be used to classify the varying levels of anger and to test this, we objectively annotated the dataset that we created using OpenFace [1] to obtain action unit data for each image in the dataset. We also subjectively annotated the dataset using free annotation where we labeled each image with either a subtle or loud label, modality, valence and arousal.

2.1 Initial Dataset

For our initial dataset, we collected 200 images of naturalistic and acted data where varying levels of anger were displayed. Half of the dataset was acted data and the other half was naturalistic data. The dataset was split between acted data and naturalistic data to reduce any potential bias that may be associated with using purely acted data. Additionally, a full dataset of naturalistic data would be significantly harder to create compared to a mixed dataset of both acted and naturalistic. The task of collecting the data was split evenly amongst the group conducting this study and we were to all collectively annotate the images that we collected. The full dataset was run through the OpenFace software to obtain the action unit data required for the objective annotation of the data. In our preliminary evaluation of the data obtained, we performed principal component analysis (PCA) on both the subjective annotation data and the objective annotation data to reduce the dimensionality of the data. The subjective annotation data was found to be widely dispersed and contained no useful information 3. As the objective data looked to be far more promising, K-Means clustering was also done on the PCA reduced objective data and we found that due to the small sample size of the dataset, there were only two noticeable clusters depicting passive anger and aggressive anger 4.

2.2 Revised Dataset

Due to the lack of useful information obtained from the subjective annotation data in the initial dataset, we decided to forgo the subjective annotation of the revised dataset and focus solely on

Figure 1: A Photo of Passive Anger in a Naturalistic Representation & Property of National Basketball Association. [Public domain], via YouTube. (ttps://www.youtube.com/watch?v=8xsotflQGP4).

the objective annotation using OpenFace software to obtain action unit data. Choosing to go with an objective annotation over subjective annotation also helps to eliminate bias from the dataset as subjective annotation varies between individuals.

In the K-Means clustering of the PCA reduced objective data 4 we obtained a silhouette score for n=2(0.417). Due to such a low silhouette score, we determined that our initial dataset was too small to be able to accurately depict that there existed two types of anger. This led to the revision of the dataset where we added 4,976 anger images from various databases to our pre-existing 200 image dataset for a total of 5,176 images displaying varying levels of anger. The datasets which we used for adding these novel anger images are from the following datasets: Labeled Faces in the Wild Database [3], Face Expression Recognition Database from Kaggle [6] and AffectNet [5]. After running all the images through OpenFace, 4,000 images were annotated. With the size of the revised database, we are now able to redo the K-Means clustering 5 which resulted in a silhouette score of n=2(0.896). With a much higher silhouette score, we are more confident with our annotations used during the supervised machine learning portion.

2.3 Classification of the Images

For the classification of the images, we compared supervised models in terms of their efficacy in classifying the anger images with the correct label. The supervised models that we used for classification are K Nearest Neighbors (kNN), Gaussian Mixture Model (GMM) and Support Vector Machine (SVM). For the kNN model,

Figure 2: A Photo of Aggressive Anger in an Acted Representation [Public domain], via CBT Counselling Kent (https://www.cbtandcounsellingkent.co.uk/how-emotionally-intelligent-people-express-anger-frustration/).

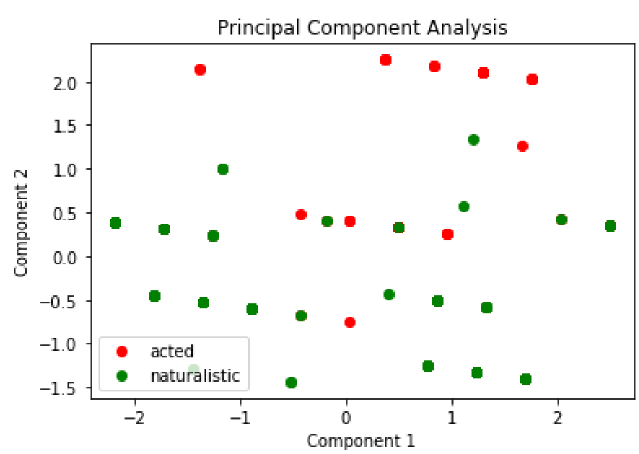


Figure 3: Principal Component Analysis of free annotation data

we previously confirmed through using unsupervised learning on

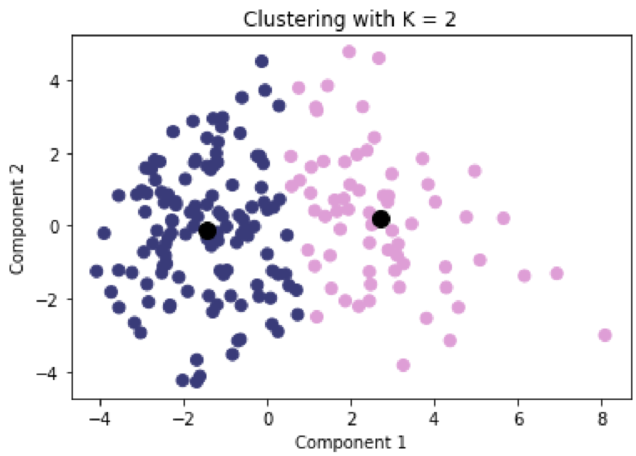


Figure 4: K-Means Clustering of OpenFace annotation data from the initial dataset

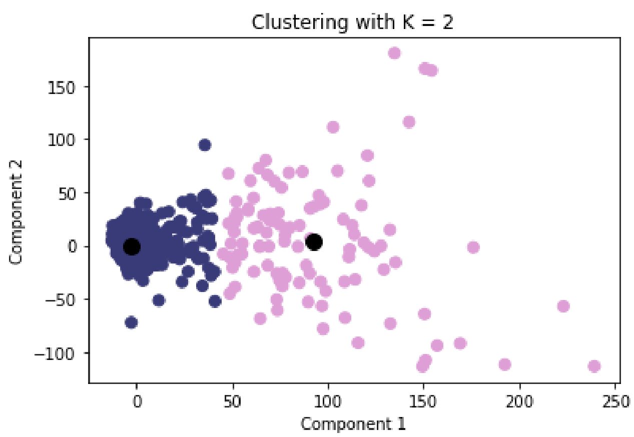


Figure 5: K-Means Clustering of OpenFace annotation data from the revised dataset

the revised dataset that the ideal number for K is 2 as we ran K Means Clustering on the revised dataset and received a high silhouette score of 0.896. Prior to classification by the models, we preprocessed the data that we received after running our dataset through OpenFace. We selected the action units from the OpenFace data that we mentioned previously which were AU_4, AU_5, AU_7, AU_9 and AU_23. We also included all of the additional data that was included by OpenFace such as X, Y, Z and gaze. This data was then scaled through a standard scaler and then principal component analysis (PCA) was done to reduce the dimensionality of the data to 2 dimensions. After all the preprocessing was completed, the data was then fed into the supervised models by splitting the dataset into training and test sets with 5-fold cross-validation which splits the data into 80% training and 20% test set. The results of the models are then plotted into confusion matrices to easily compare the results between models. The categories that we will be using to evaluate the models to each other are accuracy, precision, recall and F1 score. All of the preprocessing and supervised models that we used come from the sklearn library [7] in Python.

3 RESULTS

Model	Accuracy	Precision	Recall	F1 Score
kNN	0.50	0.51	0.50	0.51
GMM	0.44	0.46	0.44	0.43
SVM	0.54	0.48	0.54	0.49

Table 1: Results of the supervised training models.

The results that we received from each of the supervised training models (kNN, GMM and SVM) were inserted into a table 1 for comparison between the models. The results of the models were also plotted into confusion matrices to allow for easy visualization of the results.

We selected five images from our revised dataset and ran the best classification model, which in this case is SVM, to classify these images. As seen in the table 2 below, the SVM labels compared to our annotated labels are correct in the first three cases but incorrect in the last two cases.

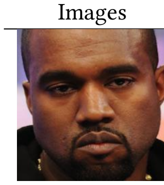
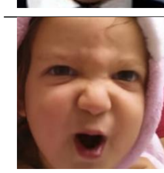
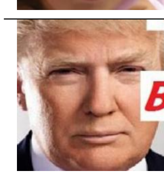
Images	SVM Classification Label	Free Annotation Label
	Subtle	Subtle
	Subtle	Subtle
	Loud	Loud
	Subtle	Loud
	Loud	Subtle

Table 2: Classification Comparision between SVM and Free Annotation.

4 DISCUSSION

Based on the results that we obtained from the different models, we can see that the lowest-performing model was GMM as it had the lowest scores in all 4 of the evaluation categories. kNN outperformed SVM in two aspects of F1 score and Precision when we averaged the scores of the test runs. However, SVM outperformed kNN in the other two evaluation categories which were accuracy and recall. Overall, based on the results of the tests that were run, the SVM model was the slightly more effective model at classifying anger images as loud anger or subtle anger, therefore, was used to classify the unlabeled images.

4.1 Dataset for Supervised Learning

In this study, the size of our dataset for supervised learning was only 200 labeled images containing either loud or subtle anger. As a result of this, the training data and the test data for the supervised learning models were limited and with a larger dataset, we could potentially see increases in accuracy, precision, recall and F1 score across all the supervised models. The increase would be a result of the model being trained with more images and therefore having more data to work with when predicting the correct label to match the test image. To aid in increasing the size of the dataset, future work could include having more annotators to annotate anger images to determine whether they display loud or subtle anger as we already have compiled a larger dataset of anger images which we used with K-Means Clustering that we described previously.

Additionally, the dataset that we used was each annotated by a single member of the group conducting the study. As for labeling the images, whether they displayed loud or subtle anger is a subjective measure resulting in cases where the labeling may be ambiguous. It may be beneficial to have multiple annotators for each image to reduce the ambiguity when labeling the images with the correct label. This would lead to the supervised models having better training data to work with to make more informed and accurate predictions on the labels to classify the test images.

4.2 Image Classification

The SVM model was able to accurately classify 3 out of the 5 unlabeled images. It was unable to classify image 4 correctly which is an image of a baby displaying loud anger. The inconsistency in the classification for the baby image could be due to the model being trained on primarily images featuring adults which have more prominent facial features compared to a baby. The inconsistency in the last image could be a result of the individual in the image having a naturally angry resting face which the model may have perceived to be displaying loud anger as a result of the data obtained from OpenFace. Alternatively, our SVM model was tested to only have a 54% accuracy so inconsistencies in labeling of the images was expected based on the current efficacy of the model.

5 CONCLUSION

This project displayed how machine learning algorithms can be used to differentiate different forms of anger. The goal of this experiment is to come up with algorithms to improve robots' effectiveness and adaptiveness. After implementing the kNN, GMM and SVM models on our data sample we ended up with the highest accuracy of 54% using SVM, with which we concluded that this project may not accurately differentiate the two different forms of anger. On the other hand, with our K-Means clustering, we can confidently conclude that there do exist two types of anger that are expressed throughout different cultures.

However, to achieve a much more accurate result we hypothesize we will need a much larger data set and with that, we will be able to come up with a more accurate algorithm to differentiate more levels of anger beyond the two we have found but our initial results are promising.

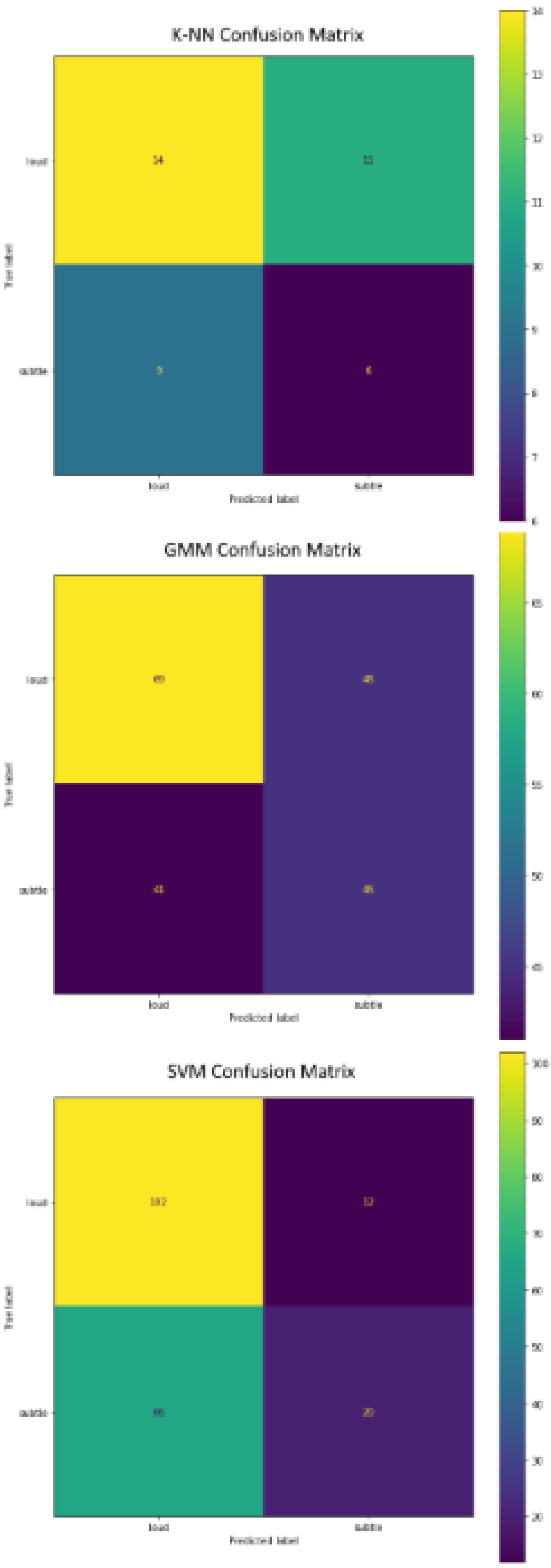


Figure 6: Confusion Matrices of

REFERENCES

[1] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. *OpenFace: A general-purpose face recognition library with mobile applications.* Technical Report. CMU-CS-16-118, CMU School of Computer Science.

[2] Simeng Gu, Fushun Wang, Nitesh P. Patel, James A. Bourgeois, and Jason H. Huang. 2019. A Model for Basic Emotions Using Observations of Behavior in Drosophila. *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.00781

[3] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. 2007. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* Technical Report 07-49. University of Massachusetts, Amherst.

[4] David Matsumoto, Seung Hee Yoo, and Joanne Chung. 2010. *The Expression of Anger Across Cultures.* 125–137. https://doi.org/10.1007/978-0-387-89676-2_8

[5] Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. 2019. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing* 10, 1 (jan 2019), 18–31. https://doi.org/10.1109/taffc.2017.2740923

[6] Jonathan Oheix. 2019. Facial expression recognition dataset. (2019).

[7] Scikit. 2022. *scikit-learn.* Retrieved Apr 15, 2022 from https://scikit-learn.org/stable/

A APPENDIX