Affect Detection using SVM and WESAD Dataset

Team 2: K. D. Portillo, T. Stowers
kdportillo, tstowers @uh.edu
University of Houston, Houston, TX, USA

***Abstract*—**

I. Introduction

In recent years, the trend of applying Machine Learing (ML) concepts to improve decision making, predciting market sentiment, recognizing patterns, affective computing, or interpreting text have gained the attention of many. Within the realm of ML concepts are classification problems where the objective is to classify data entries as either categorical, ordinal, or binary [**?**, p. 327-328]. As a result, classification algorithms (CAs) were developed using statistical analysis techniques. When applied to a dataset, CAs can provide great insight into common features, known as class, found within the data. Classifiers in particular use the training data to assess how the data's respective attributes fit within the definitions of a particular class with respect to the ground truth [**?**].

A field of study known as Affective Computing attempts to infer the emotional state of a human being during computer usage. Knowing the emotional state of the user can help machines alter thier content based on this information. Companies can use this on their employee's computers to help improve productivity and health.

This year, a group of researchers introduced a WEarable Stress and Affect Detection (WESAD) dataset that provides a multimodal high-quality dataset with various affective states [**?**]. The experiment tested for three affective states amusement, stress, and neutral. It also could help determine whether a test subject was or was not stressed. In general, stress and emotion are not mutually exclusive since both rely on the body's central nervous system. This has been a problem within the Affective Computing research field. However, this dataset was made with the intention of bridging this problem. For the purpose of this paper, the focus is geared towards detecting if a user is stressed or not.

II. Related Work

The field of Affective Computing has picked up steam in the last couple of decades. This field of study is interdisciplinary which culminates Computer Science, Computer Engineering, Physcology, and Physiology. With the average employee spending 7 hours a day on the computer, 6 at work and 1 at home, it is important to monitor the health conditions of users during thier time behind a screen [**?**].

Affective computing can be defined as the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. While the origins of the field may be traced as far back as to early philosophical inquiries into emotion, the more modern branch of computer science originated with Rosalind Picard's 1995 paper on affective computing. A motivation for the research is the ability to simulate empathy. The machine should interpret the emotional state of humans and adapt its behavior to them, giving an appropriate response for those emotions. Affective computing technologies sense the emotional state of a user(via sensors, microphone, cameras and/or software logic) and respond by performing specific, predefined product/service features, such as changing a quiz or recommending a set of videos to fit the mood of the learner. The more humans rely on computer devices, the more crucial it is to want them to behave politely and be socially aware. Common-sense reasoning requires an understanding of the user's emotional state.

One way to look at affective computing is human-computer interaction in which a device has the ability to detect and appropriately respond to its user's emotions and other stimuli. A computing device with this capacity could gather clues to user emotion from a variety of sources. Facial expressions, posture, gestures, speech, the force or rhythm of key strokes and the temperature changes of the hand on a mouse can all signify changes in the user's emotional state, and these can all be detected and interpreted by the same computer. A built-in camera, for example, can capture images of the user and use classification algorithms to process the datum that yields information about the user's facial expression during computer usage. Speech recognition and gesture recognition are among the other technologies being explored for affective computing applications. Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This can be achieved through ML techniques that process different modalities like speech, natural languages, or facial expressions.

A major area in affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between humans and machines. While human emotions are often associated with surges in hormones, emotions in machines might be associated with abstract states associated with progress (or lack of progress) in autonomous learning systems. In this view, affective emotional states correspond to time-derivatives in the learning curve of an arbitrary learning system. Two major categories describing emotions in machine are: Emotional speech and Facial affect detection. Emotional speech includes: algorithms, databases and speech descriptors. The facial affect detection is based off of physiological monitoring.

The process of speech/text affect detection requires the creation of a reliable database, knowledge base, or vector space model, broad enough to fit every need for its application, as well as the selection of a successful classifier which will allow for quick and accurate emotion identification. Currently, the most frequently used classifiers are linear discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms and hidden Markov models (HMMs). Various studies have shown that choosing the appropriate classifier can significantly enhance the overall performance of the system. For this project, SVM is used to classify stress within male patients. The SVM model is a type of linear classifier which decides in which of the two or more possible classes, each input may fall into.

The vast majority of present systems are data-dependent. This creates one of the biggest challenges in detecting emotions based on speech, as it implicates choosing an appropriate database used to train the classifier. Most of the current accessible data was obtained from actors and is thus a representation of archetypal emotions. Those so-called acted databases are usually based on the Basic Emotions theory proposed by Paul Ekman, which assumes the existence of six basic emotions (anger, fear, disgust, surprise, joy, sadness), the others simply being a mix of the former ones. As with every computational practice, in affect detection by facial processing, some obstacles need to be surpassed, in order to fully unlock the hidden potential of the overall algorithm or method employed. The accuracy of modeling and tracking has been an issue, especially in the incipient stages of affective computing. As hardware evolves, new discoveries are made, and new practices are introduced, the lack of accuracy fades; leaving behind noise issues. However, methods for noise removal exist including neighborhood averaging, linear Gaussian smoothing, median filtering, outlier detection, etc.

Physiological monitoring can be used to detect a user's emotional state by monitoring and analyzing their physiological signs. These signs range from their pulse and heart rate to the minute contractions of the facial muscles. This area of research is still in relative infancy as there seems to be more of a drive towards affect recognition through facial inputs. The three main physiological signs that can be analyzed are blood volume pulse, galvanic skin response, and facial electromyography. It can be extremely difficult to ensure that the sensor shining an infra-red light that monitors reflected light is always pointing at the same extremity, especially seeing as subjects often stretch and readjust their position while using a computer. There are other factors which can affect one's blood volume pulse. As it is a measure of blood flow through the extremities, if the subject feels hot, or particularly cold, then their body may allow more, or less, blood to flow to the extremities, all of this regardless of the subject's emotional state. Affective Computing has a bright future due to its limitless capabilities of developing useful data applications that can help researchers find data in effective time tables.

Along with these stress level ratings, in particular, comes error, bias, and variance in technology or simply, the subjects themselves. The goal of modern Affective Computing lies within providing quality, accurate conclusions where other methods are not consistent in. Efforts in facial recognition that classify the affect of a human have been developing in recent years. The changes in muscles across the face lead to attributes that Charles Dawrin infered about universally about humans. These attributes have been consolidated to various class' that have been paired to classification algorithms for further analysis which resulted in 96% accuracy for stationary images [**?**].

With the dependency of computers steadily increasing over time, the need for machines to interpret a user's affective state is important for their overall health. The contrary could result to ruined relationships or even shortness of life. This paper attempts to focus on stress detection during daily computer tasks to mitigate the negative effects of stress over time.

III. Problem Statement

As mentioned in section **??**, stress is one of the main causes for serious health concerns. Because of this, there are stress monitors that supervise blood volume, pulse, emotions, and body temperature. Stress and anxiety can occur at any given time. In particular, during computer usage, stress levels can fluctuate as a result of the content or tasks the user is seeing or doing. The WESAD dataset and the SVM model are used to help analyze the dataset to detect when and why a test subject became stressed during the course of the experiment. Using a 12-Fold Cross Validation approach, the goal is to determine if relative stress detection gives insight into real-time detection in male patients. A detailed analysis of benchmarks for each test is provided to determine which model performed the best.

A. WESAD Dataset

The WESAD dataset was composed after a stress test that was performed on to twelve males and three females. To collect their data, they used both a chest- and a wrist-worn devices: a RespiBAN Professional2 and an Empatica E43. The RespiBAN itself is equipped with sensors to measure accelerometer and respiratory data, and can function as a hub for up to four additional modalities. Other datum were also recorded like from photoplethysograph, electrodermal activity, and temperature were recorded. All signals were sampled at 700 Hz.The RespiBan was placed around the subject's chest; The respiratory data was recorded via a respiratory inductive plethysmography sensor. In order to allow the subject to move as freely as possible, the electrodermal activity signal was recorded on the rectus abdominis, and the temperature sensor was placed on the sternum. All subjects wore the Empatica E4 on their non- dominant hand. The Empatica E4 recorded blood volume pulse at 64Hz, temperature at 4Hz, and electrodermal activity at 4Hz.

The goal of the WESAD study was to elicit three different affective states (neutral, stress, amusement) in the participants. In addition, the subjects were asked to follow a guided meditation in order to de-excite them after the stress and amusement conditions. After the subjects had been equipped with the sensors, a 20 minute baseline test was recorded. During the baseline, the subjects were asked to sit or stand at a table. After the baseline condition results were recorded, there was an amusement condition test. During the amusement condition, the subjects watched a set of eleven funny video clips. In total, the amusement condition had a length of 392 seconds. The final test conducted was the Trier Social Stress Test (TSST) which consists of a public speaking and a mental arithmetic task. In their version of the TSST, the study participants first had to deliver a five minute speech on their personal traits in front of a three-person panel, focusing on strengths and weaknesses. After the TSST the study participants were given a ten-minute rest period. At the end of the protocol, the sensors were again synchronised via a double tap gesture. In total, the study had a duration of about two hours.

IV. Methods & Results

A. Methods

From the WESAD Dataset, the male patients were selected to train and test the SVM model. The dataset indexes each male or female patient as S_n where $n \in [2, 17]$ and $n \neq 12$ which yields a total of 15 patients. From S_n, the female patients were patients $n = 8, 11, 12$. The resulting set of male only patients, M, is thus

$$M = \{S_n, n \neq 8, 11, 12, 17\}$$

and has a cardinality of 12. The M set was then remapped to fit a more intuitive indexing for SVM model and 12-Fold Cross Validation purposes. Table **??** denotes how each male patient's study index was remapped.

Study Index	Remapped Index
S_2	1
S_3	2
S_4	3
S_5	4
S_6	5
S_7	6
S_9	7
S_{10}	8
S_{13}	9
S_{14}	10
S_{15}	11
S_{16}	12

TABLE I
REMAPPING WESAD INDEXES FOR IMPLEMENTATION

As noted in section **??**, a 12-Fold Cross Validation technique was chosen to partition the WESAD dataset into training and testing data. For the purpose of this experiment, each male is gaureenteed to be the training set and subsequently be the testing set. For example, M_1 would be the fitted as the first trainer in the SVM model and males M_{2-12} would be tested using the fitted M_1 model. This process repeats itself a total of 12 times, once for each male. The average accuracies from each training and testing pair within each fold are recorded. To ensure that the training and teting data was of the same length each time, a function, `preprocessing`, was used to ensure both datum satisfied this condition. At the end of each fold, meaning using the example above would be M_1 as trainer and M_{12} as testing, the mean average is computed and recorded along with the remapped index. Finally the 12 final average and fold pairs are ranked in decsending order by average accuracy. The following pseudocode denotes how the methods above were expressed.

Algorithm 1 12-Fold Cross Validation using SVM

Require: 12 male patients datum (EDA and HR)
for $M_i \in M$ **do**
  training $= [M_i^{EDA}, M_i^{HR}]$
  for $M_j \in M$ and $i \neq j$ **do** ▷ Ignore case when training set is the testing set
    testing $= [M_j^{EDA}, M_j^{HR}]$
    if $len(training) \neq len(testing)$ **then**
      training, testing $= preprocessing(training, testing)$
    end if
    results $= [classify_{i,j}(training, testing)]$
  end for
  averages $= (i, mean(results))$ ▷ Append final averages after each fold
end for
$rank(averages, "descending")$

B. Results

The table below shows the ranked results from the 12-Fold Cross Validation experiment.

Male Patient	Accuracy (%)
M_2	1
M_3	2
M_4	3
M_5	4
M_6	5
M_7	6
M_9	7
M_{10}	8
M_{13}	9
M_{14}	10
M_{15}	11
M_{16}	12

TABLE II
FINAL AVERAGES AFTER 12 FOLDS OF TESTING

The average computation time is 2 minutes and 19.480 seconds. This experiment was perfromed on a MacBook Pro, Late 2013, 6th Gen i7 Intel Core Processor, 16 GB RAM machine. Another experiment was perfromed on a HP x360 Spectre with an 8th Gen i7 Intel Core Processor, 16 GB RAM machine. The average computation time for this platform was is 2 minutes and 19.480 seconds. The 8th Generation Processor performed NUM% better than the 6th Generation one.

V. Analysis

Before diving into the results noted in table **??**, here are some statistics about the males patients sampled from the WESAD dataset.

- Average age: 27.5 yrs, Variance of 7.18 years
- Average height: 179.67 cm, Variance of 32.06 cm
- Average weight: 77 kg, Variance of 77 kg
- All were right hand dominant
- 16.67% Drank coffee the day of their test
- No one drank coffee an hour before taking the exam
- 1 patient played sports the day of the exam

- 1 patient is a smoker
- No one smoked within an hour prior to the exam
- Only 1 patient felt ill the day of the exam

M_2 proved to be the best trainer overall when testing against the other male patients with an average accuracy of NUM%. Below is a list of attributes about M_2.

- Right hand dominant
- Age: 27
- Weight: 66 kg
- Height: 129 cm
- Did not drink coffee the day of the study
- Did not drink coffee 1 hour before the study
- Did not participate in sport activities the day of the study
- Is not a smoker
- Did not smoke an hour before the study
- Did not feel ill the day of the study
- Additional Notes
    1) During the baseline condition, the subject was sitting in a sunny workplace
    2) Subject provided a valence label of 7 after the stress condition, claiming that he was looking forward to the next condition and was therefore cheerful

The runtime for this experiment was exceedingly high. This is a result of single threading the entire experiment. A possible optimization would be to initiate 12 threads, one for each fold. The threads could run concurrently without the fear of stepping over one another. Since each fold simply steps through the dataset and selects their respective training and testing samples, a multithreaded solution would significantly increase runtime execution.

Another potential optimization technique would be to import the entire WESAD dataset at once. For each fold, the program is reading from two files 4 times. Reading from disk is a costly function that results in longer execution times. By reading only once, the number of total times the program opens and reads from files and directories. Alternatively, formatting the data so that it only requires a single read operation instead of traversing multiple directories for the data.

VI. Conclusion

References

[1] Jiawei Han, Micheline Kamber, Jian Pei *Data Mining Concepts and Techniques 3rd Ed.*. Morgan Kaufmann, 2012.
[2] Sidath Asiri Machine Learning Classifiers, 2018
[3] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger and Kristof Van Laerhoven *Introducing WESAD, a multimodal dataset for Wearable Stress and Affect Detection*, ICMI 2018, Boulder, USA, 2018
[4] Ron Winbeta Microsoft US Workers, 2013
[5] Denise M. Martz The Relationship Between Feminine Gender Role Stress, Body, And Eating Disorders, 1995
[6] Gross, C., Seebab, K., *The Standard Stress Scale (SSS): Measuring Stress in the Life Course*, (NEPS Working Paper No. 45). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study, 2014.
[7] Alice Boyes, Ph.D Five Types of Good Stress, 2012
[8] Saeed Turabzadeh, Hongying Meng 1, Rafiq M. Swash, Matus Pleva, Jozef Juhar *Facial Expression Emotion Detection for Real-Time Embedded Systems*, MDPI, Basel, Switzerland. 2018.