A regressive approach to emotion prediction applied to EEG signals from passive elicited emotion

Joey de Kruis
STUDENT NUMBER: 2035198

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

B. Nicenboim
M Atzmüller

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
May 2020

Preface

I hereby present the dissertation "A regressive approach to emotion prediction applied to EEG signals from passive elicited emotion", comprehending an EEG study in the emotion prediction domain. This dissertation has been written as an assignment for the Tilburg University master Data Science Society, focused on business, with the purpose to contribute to the emotion prediction domain.

Although the COVID-19 pandemic forced us to adjust to a new working style, this study has been accomplished with a lot of curiosity and pleasure. Coming from little knowledge of EEG and emotion prediction, this work has allowed me to significantly increase my expertise in this topic.

I am grateful for the collaboration with and the supervision of my supervisor B. Nicenboim. I prefer to take a lot of freedom in exploring different solutions and challenging myself to achieve a certain goal. B. Nicenboim has given me that freedom, while always being ready for feedback or help when I did need it, and I am thankful for that.

Isaac Newton once said, "If I have been able to see further, it was only because I stood on the shoulders of giants". By challenging certain ideas and questioning aspects of the research, a study gains the opportunity to improve its quality. Therefore, I would like to thank my giants N. Chauhan and D. van den Corput for challenging my ideas and with that improving my thesis.

I aspire this study to be a contribution and an inspiration to the emotion prediction domain, and I hope that readers will feel the same way.

Joey de Kruis,
Tilburg, May 2020

A regressive approach to emotion prediction applied to EEG signals from passive elicited emotion

Joey de Kruis

This paper addresses the two major EEG emotion prediction challenges regarding a scarcity in emotion prediction research that use passive emotion elicitation stimuli and the implementation of a proper theoretical framework for emotion. By using passively elicited emotion from the reading of sentences, emotion has been regressed based on a two-dimensional framework containing the dimensions 'arousal' and 'valence'. Two setups of EEG features have been evaluated on three regression models, where setup 1 contains the differential asymmetry of the Fractal Dimensions, Statistical features, and Band power features. Setup 2 contains the differential asymmetry of statistical features from the Intrinsic Mode Functions extracted with the Hilbert Huang Transform. A basic linear regression model has been employed in combination with a K-Nearest Neighbor Regressor and a Support Vector Regressor. Results for arousal indicate a better performance of setup 2 with an Mean Absolute Error (MAE) of .605. Valence results show similar predictions for both setups resulting in an MAE of .695. This study shows positive results regarding the possibilities of applying a regressive approach on passive elicited EEG data but emphasizes more research is necessary to obtain practically reliable predictions.

1. Introduction

The effect of emotion in marketing communication with branding has been distinct since the early zeros. Lynch and De Chernatony (2004) stated that companies need to communicate their emotional brand values more effectively. Subsequently, a lot of research has been done into understanding human emotions, and ultimately into emotion prediction (also referred to as emotion recognition and affective computing). Accurate predictions of emotions in marketing communication could eventually lead to more effective marketing campaigns.

A prominent way to predict human emotion is by using electroencephalograms (EEG) signals. In the past 20 years, research has shown that EEG signals are quite valuable features for emotion prediction (Musha et al. 1997; Takahashi et al. 2004). Specifically, EEG is a cheap and relatively easy method for obtaining physiological responses that are suited for EEG emotion prediction (Soroush et al. 2017).

The majority of EEG emotion prediction research experiments are benefiting from methods with active emotion elicitation stimuli (i.e. stimuli that are actively provoking emotion as video clips, music, pictures) to create clear signals (Soroush et al. 2018). Although these are methods that are excellently provoking a reaction inside the brain, there is a real issue regarding ecological validity while using these stimuli (Hu et al. 2019). In the real world, emotion is more passive and comes in lower intensive forms. To meet the ecological validity, more passive emotion elicitation stimuli need to be used as texts and conversations. Considering the scarcity in research for predicting emotion with EEG data using passive emotion elicitation stimuli (i.e. stimuli that are passively provoking emotion like reading and talking), this paper proposes to investigate several predictive algorithms to find out the best way to predict emotion when reading a sentence.

In order to predict emotion, it needs to be properly conceptualized. Because of the complexity of human emotion, scientists disagree on how to conceptualize emotion resulting in different perspectives. The discrete perspective and the dimensional perspective are the most common perspectives used for conceptualizing emotion (Mauss and Robinson 2009). The discrete perspectives categorizes emotional states (e.g. 'sad', 'happy') based on experience, physiology, and behaviour (Panksepp 2007). The dimensional perspective uses continuous dimensions to organize emotional response (Mauss and Robinson 2009). The most common dimensions in the dimensional perspective are arousal, valence, and approach-avoidance (Mauss and Robinson 2009).

Consequently, a major challenge with emotion prediction is the adoption of a consistent and proper theoretical framework for emotion (Hu et al. 2019). Although various researchers use a framework based on the discrete perspective, EEG and neuroimaging studies, indicate there is a link between specific brain regions and the dimensional measures (Mauss and Robinson 2009). More specifically, there is convincing evidence that the frontal EEG asymmetry is considerably sensitive to the approach-avoidance dimension (Davidson 1999). Considering the limited availability of approach-avoidance labels, the proposed study applies the dimensional perspective with the use of the dimensions arousal and valence that are widely available. With the use of gathered EEG data from reading sentences, the interactive emotional effect between the noun and the adjective is utilized to investigate the research question "In what way can we best predict the arousal and valence for the adjective-noun combination when reading a sentence, by using EEG data?" (Nicenboim, Vasishth, and Rösler 2020; Lüdtke and Jacobs 2015).

2. Related Work

Due to its complexity and still limited understanding, emotion prediction is a field with many different preprocessing techniques, features, models, concepts, and emotion elicitation methods. An overview of some recent studies on emotion prediction and their different characteristics is shown in Table 1. In section 2.1, several feature extraction methods are reviewed that are relevant for the current study. Section 2.2 will focus on reviewing different models and a channel selection method.

Although there is a scarcity of research done with passive emotion elicitation methods, one study proposes a new channel selection method applied by using EEG data from recalling past events (Ansari-Asl, Chanel, and Pun 2007). With the use of conversations, Ansari-Asl, Chanel, and Pun (2007) did not actively provoke emotion. On the contrary, the study created an environment where people were free to talk about events that passively elicits emotion (Ansari-Asl, Chanel, and Pun 2007). Ansari-Asl, Chanel, and Pun (2007) proposed synchronization likelihood as a channel selection method. Interesting results are shown with a 5.2% average increase in accuracy when selecting significant channels (Ansari-Asl, Chanel, and Pun 2007). However, the predictions were made based on the discrete model. Consequently, more research is necessary to evaluate a similar experimental setup integrating the dimensional perspective. Additionally, there is limited information on if and how speech and reading are connected inside our brains when it comes to emotion. Although the study done by Ansari-Asl, Chanel, and Pun (2007) shows promising results when it comes to classifying emotion with

Table 1
An overview of recent emotion prediction studies and its different aspects.

Perspective	features	Models	Source
Dimensional	HHT, HOC, STFT	RF, SVM	Ackermann et al. (2016)
Combination	FD, BP Statistical	SVR	Lan et al. (2016)
Discrete	STFT, PSD	KNN, DBN SVR, LR	Zheng and Lu (2015)
Discrete	HHT	SVM	Zong and Chetouani (2009)
Dimensional	HHT	SVR	Uzun, Yildirim, and Yildirim (2012)
Discrete	Statistics, BP Wavelet	ANN, KNN SVM	Bhatti et al. (2016)

Note. HHT: Hilbert Huang Transform; HOC: Higher Order Crossings; STFT: Short Time Fourier Transform; FD: Fractal dimensions; PSD: Power Spectral Density; BP: Band powers; RF: Regression forest; SVM: Support Vector Machine; SVR: Support Vector Regressor; LR: Linear regressor; DBN: Deep belief network.

the passive emotion elicitation method, the current study is applying the dimensional perspective rather than the discrete perspective.

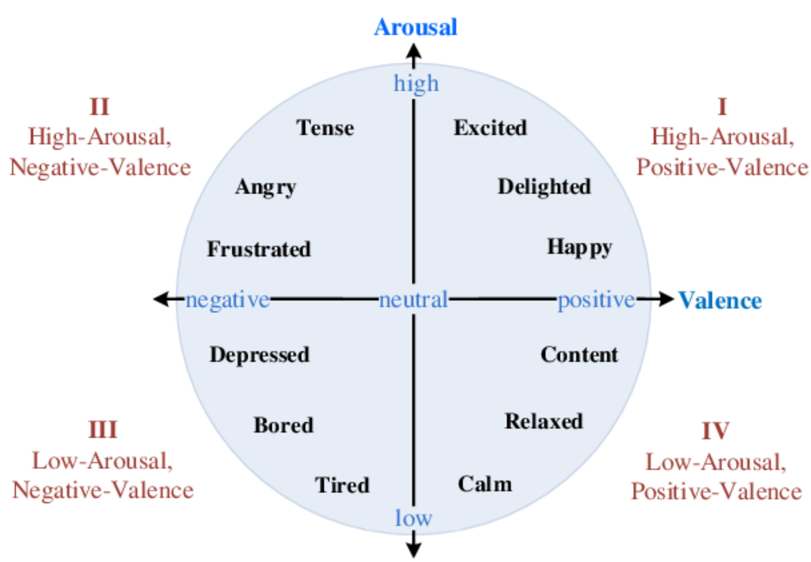
The majority of emotion prediction research applies the discrete perspective for conceptualizing emotion. Some studies do apply the dimensional perspective, but turn it into a classification task by predicting the corner of the matrix (e.g. high arousal and low valence) like done in a study by Ackermann et al. (2016). Although previous knowledge can be taken into account when applying a classification method, it is still unclear what values match with which emotion in the matrix. As shown in Figure 1, different emotions can be found in each edge of the two-dimensional space, which complicates it more when there are more than 2 dimensions (Yu et al. 2016). Ackermann et al. (2016) selected three different emotions to be predicted in a three-dimensional space, resulting in wide assumptions of what that area in the three-dimensional space represents. All things considered, a combination of both perspectives gives uncertainty and can lead to a bias in what actually is being predicted.

2.1 Feature extraction

Lan et al. (2016) attempted to predict valence (controlling for arousal and dominance) based on the International Affective Digital Sounds (IADS) dataset. Lan et al. (2016) picked a regressive approach, so they could reduce the number of training resources that are beneficial for a continuous emotion prediction application. In their experiment, they selected three different sets of features: fractal dimensions (FD), statistical features (STAT), and band power (POWER) features (Lan et al. 2016). Noteworthy is that all feature parameters were extracted by applying a sliding window on the EEG channels, varying from 32 steps (93.75% overlap) to 512 steps (0% overlap) (Lan et al. 2016). Results show that applying a 50% overlap or more brings the best performance (Lan et al. 2016). Additionally, FD features, mean of absolute values of first-order differences, mean of absolute values of second-order differences, and band powers are found to be significantly correlated to valence, achieving a mean absolute error (MAE) of .74 (Lan et al. 2016). In the current study, a similar setup of features is proposed excluding the overlap applied during feature extraction. Due to the significant difference of the study setup, where Lan et al. (2016) uses 60 seconds signals compared to the 2 seconds used in the current study, it can be assumed that applying an overlap in the current study will have a weaker effect and is therefore not implemented.

Another continuous EEG emotion prediction study achieved worse MAE results predicting emotion, based on EEG and other physiological signals (e.g. blood volume pulse, respiration, skin temperature, etc.) (Soleymani et al. 2011). Soleymani et al. (2011) also extracted the band power features using Welsch's method. Additionally, the lateralization (the tendency of neurons to be active in one side of the brain) of several left-right pairs were extracted using the power spectral density (PSD) that were computed for the band powers. The study achieved an MAE for valence of 1.59 purely based on EEG, and an MAE for arousal of 1.53 based on EEG, although the best results came from EEG features combined with music video features. Even though the achieved results were not significantly better as compared to Lan et al. (2016), the use of lateralization is an interesting approach and is used more often in similar studies focused on classification (Jenke, Peer, and Buss 2014). One particular study is interesting regarding its review on many popular features extraction methods used for EEG emotion prediction (Jenke, Peer, and Buss 2014). Jenke, Peer, and Buss (2014) collected EEG data in an experiment using The International Affective Picture System (IAPS) dataset to elicit emotion. All features proposed were computed and several feature selection algorithms were applied to investigate the most frequently used features (Jenke, Peer, and Buss 2014). Results

Figure 1
A conceptual arousal-valence matrix based on the dimensional model.



Note. This figure shows how different emotions can be found inside a conceptual two-dimensional model for arousal and valence.

confirm that rational asymmetry features, which is the relative difference in the signal of different EEG channels (based on the lateralization concept), performs accurately in the study setup of Jenke, Peer, and Buss (2014). Additionally, the study favors the use of 'advanced' methods like the Hilbert Huang Transform (HHT), Higher Order Crossings (HOC), and Higher Order Spectrum (HOS).

Uzun, Yildirim, and Yildirim (2012) applied HHT on the International Affective Digital Sounds (IADS) dataset by using the dimensional perspective (containing arousal/activation, valence, dominance) as a concept for emotion. Empirical mode decomposition (EMD), the first part of the HHT, is used to create 5 Intrinsic Mode Functions (IMFs) per signal (Uzun, Yildirim, and Yildirim 2012). Afterwards, 4 statistics are computed over signal in combination with 7 POWER features, resulting in an average arousal MAE of .65 (SD = .09) and an average valence MAE of 1.11 (SD = .13) (Uzun, Yildirim, and Yildirim 2012). Although valence error was not significantly lower than the baseline created in Lan et al. (2016), arousal error did achieve a relatively good performance compared to what was predicted with Soleymani et al. (2011). Other classification oriented studies have similarly investigated the predictive results coming from the HHT. Zong and Chetouani (2009) likewise applied HHT on their data of the University of Augsburg, comparing it with baseline methods. Zong and Chetouani (2009) concluded that HHT outperforms traditional methods, favoring based on the result of a 'Fission' HHT approach. Due to the accurate performance of the HHT in the study of Uzun, Yildirim, and Yildirim (2012) and in other emotion prediction studies, HHT will be used as a second setup to evaluate the best way to predict emotion using passive emotion elicitation methods.

2.2 Models and channels

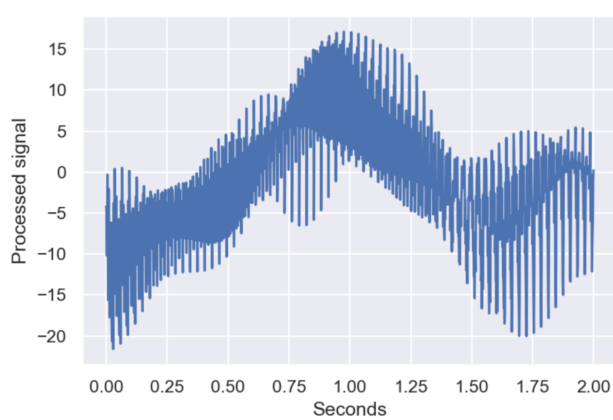
Notable results have been made across different combinations of study setups using the classification model K-Nearest Neighbor (KNN), and the Support Vector Machine(SVM) (Soroush et al. 2018; Hassanien et al. 2018). A study evaluating different algorithms using the IAPS database indicates that KNN predicts best out of a set of five classifiers (i.e. KNN, SVM, Bayesian Network, Regression Tree, Artificial Neural Networks) (Sohaib et al. 2013). A comparative study using self-reported emotional states confirms accurate classification performance of KNN, although it displays the competitive performance of other classifiers where SVM and Regression Tree (RT) were outperforming KNN (Rani et al. 2006). Based on that similar regression studies applied an SVR as done in the study Lan et al. (2016) and Uzun, Yildirim, and Yildirim (2012), the proposed study will explore results obtained from the SVR. Additionally, the regression equivalent of the KNN model (K-Nearest Neighbor Regressor, KNNR) will be used to compare the effectiveness among different models, due to KNN's significant results in classification studies. In order to investigate the predictive differences between the models, a basic linear regression model will also be applied to test whether a minimum model would already be effective.

At last, the EEG channels that are picked for an emotion prediction study are of importance. Oftentimes EEG studies are dealing with many signal samples and different channels, which creates a heavy dataset requiring a lot of processing time or power and storage. In emotion prediction there is also a strong impact from the way emotion has been elicited. Different emotion elicitation stimuli process different information, which makes the selection of EEG channels a more complicated challenge (Ansari-Asl, Chanel, and Pun 2007). Even though there is still not enough knowledge about which channels work best for which elicitation method, some studies are attempting to find channel selection methods that can possibly identify relevant channels. Ansari-Asl, Chanel, and Pun (2007) proposed synchronization likelihood (SL) as a channel selection method using MATLAB, showing that a dataset of 60 channels could be reduced to only 5 channels with just slightly reducing accuracy performance. Ansari-Asl, Chanel, and Pun (2007) concluded that frontal channels F3, F4, and AFz performed best, maximizing the SL values. A more recent review on channel selection methods for emotion prediction concluded similarly that the frontal pairs of channels give better results compared to other channel combinations. Due to the absence of support for SL in the used processing languages, this method will not be used in the current work. However, the current research will make use of the frontal pairs of channels that seem to perform accurately in emotion prediction study.

3. Experimental Setup

In order to explore the best way to predict arousal and valence with a passive emotion elicitation method, two advanced feature extraction setups are proposed. To evaluate the performance and the impact of several models, 2 top-performing models in the EEG emotion prediction domain are utilized in combination with a basic linear regression model. In section 3.1 more information is given on the EEG data and the emotion labels. Section 3.2 describes the preprocessing methods and the two feature extraction setups that are proposed. Section 3.3 elaborates on the decision in which models and channels are being picked. Lastly, in section 3.4 the evaluation metrics are proposed that fit the current study. All code, models, and scripts can be found in https://github.com/Joeydekruis/emotion-prediction.

Figure 2
Processed EEG signal example



Note. This figure is an example of a randomly picked EEG channel of a randomly picked EEG signal that has been pre-processed.

3.1 Data

This study adopts the EEG data obtained by Nicenboim, Vasishth, and Rösler (2020) that investigated the effect of gender during the pre-activation of the noun. EEG data of 120 respondents were collected while they were reading German sentences. The dataset can be found at https://osf.io/ut7xq/files/ under the "data" folder and the "eeguana_preproc_clean" folder. In contrary to the research of Nicenboim, Vasishth, and Rösler (2020), the current study is focused on the emotion obtained from passive elicitation methods. To realize this, the variables constraining (i.e. whether the context is forced to make the continuation predictable) and completion (i.e. whether the sentence meets the predicted continuation) are being controlled and the dataset is filtered on 'non-constraining' and 'not-completed'. Additionally, the dataset is filtered on the regions of the adjective and the noun, due to their interactive emotional effect. An example of a randomly picked EEG signal for a randomly picked channel is showed in Figure 2.

The consecutive data necessary for this study are the emotional labels. As described in the introduction, arousal and valence are used according to the dimensional model. A recent study constructed a dataset containing 350.000 German words rated on the attributes of abstractness-concreteness, arousal, valence, and imageability (Köper and Im Walde 2016). The ratings were based on a 10 point scale (0 - 10) and range from 1.780 - 6.273 for arousal and 3.081 - 7.402 for valence. The study used the unsupervised Turney and Littman algorithm to classify the words based on several prominent studies (Vo et al. 2009; Lahl et al. 2009; Kanske and Kotz 2010) as training data (Köper and Im Walde 2016). With the purpose of regressing arousal and valence for single sentences, the emotional values are matched with every word in every sentence and are averaged over both regions (i.e. adjective and the noun). If there was no exact match for a word, a smaller word up to 3 character differences at the end of the word is matched. This

Figure 3
The probability distribution of arousal and valence.

Note. A plot showing the probability distribution for arousal and valence, coming from the kernal density estimation was necessary due to the different forms of German words (e.g. past and present). A coverage of 100% of all words was achieved. The probability distribution of both arousal and valence is showed in Figure 3. As showed in Figure 3, both distributions differ from each other. A Shapiro-Wilk test indicates a non-normal distribution for both arousal (F = .99, p = .000) and valence (F = .99, p = .000) where arousal is skewed to the left with .13 and valence is skewed to the right with -.22.

3.2 Preprocessing and feature extraction methods

(Nicenboim, Vasishth, and Rösler 2020) filtered EEG data applying a zero-phase band-pass finite impulse response (FIR) filter combined with a pass band-edge frequencies of .1 and 30 Hz. More in-depth is the width of transition band .10 and 7.50 Hz for low and high edges. Additionally, Independent Component Analysis (ICA) has been applied by Nicenboim, Vasishth, and Rösler (2020) using the FastICA algorithm to correct eye movements. R package Tibble is used to combine all respondents and filter channels and samples based on the variables picked (Müller and Wickham 2017). Python is used to create a two-dimensional array consisting of all channels times the number of samples present for the word (i.e. approximately 650). Furthermore, for each sentence, the EEG data for the adjective and the noun are combined and the correct averaged label is added to the sentence. All sentences containing just the noun, or just the adjective, have been removed. Similarly, all sentences without a label have also been removed. The benefit of averaging the labels over the sentence is that it will not lose the context of the 10-point scale and with that, the study can still be compared to other studies. Two relevant feature extraction setups are proposed in section 3.2.1 and in section 3.2.2.

3.2.1 Setup 1. The first feature extraction method is based on the emotion prediction study attempting to predict valence (Lan et al. 2016). In the current study, fractal dimensions (FD) , statistical features (STAT) and band power features (POWER) are used. In the description of these features, it is to be mentioned that $\xi(t) \in \mathbb{R}^T$ denotes the time series vector of an EEG channel. Fractal dimensions is a time-domain feature used for characterizing non-linear time series. Fractal dimensions can be applied in several ways. According to (Jenke, Peer, and Buss 2014), the Higuchi algorithm is recommended for FD, because it comes close to the theoretical FD values. The Higuchi algorithm is implemented in the current work, where the finite time series $\xi(t)$, $t = 1,...,T$ is rewritten as (Liu and Sourina 2013):

$$\left\{\xi(m), \xi(m+k), ..., \xi\left(m + \left[\frac{T-m}{k}\right] \cdot k\right)\right\}, \tag{1}$$

where $[\cdot]$ denotes the Gauss' notation and where m and k are both integers and respectively represent the initial time and the time interval. A time interval that is equal to k, gives k sets of new time series. K sets are

$$L_m(k) = \frac{T-1}{\left[\frac{T-m}{k}\right]K^2} \sum_{i=1}^{\left[\frac{T-m}{k}\right]} \left|\xi(m+ik) - \xi(m+(i-1)k\right|. \tag{2}$$

Table 2
An overview of the statistics and its functions that are applied.

	Statistic	Function
1	Mean	$\mu_\xi = \frac{1}{T}\sum_{t=1}^{T}\xi(t)$
2	Standard deviation	$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\xi(t)-(\mu_\xi)^2}$
3	1st difference	$1T-1\sum_{t=1}^{T-1}\left\|\xi(t+1)-\xi(t)\right\|$
4	1st difference normalized	$\overline{\delta_\xi} = \frac{\delta_\xi}{\sigma_\xi}$
5	2nd difference	$\gamma_\xi = \frac{1}{T-1}\sum_{t=1}^{T-2}\left\|\xi(t+2)-\xi(t)\right\|$
6	2nd difference normalized	$\overline{\gamma_\xi} = \frac{\gamma_\xi}{\delta_\xi}$

After calculating K sets, the average value of K sets, denoted as $\langle L(k)\rangle$, comprehends of the following relationship with the fractal dimension FD_ξ(Jenke, Peer, and Buss 2014; Liu and Sourina 2013):

$$\langle L(k)\rangle \propto k^{-FD}\xi. \tag{3}$$

The fractal dimension dim_H is obtained from the logarithmic difference between t and and the associated k (Liu and Sourina 2013). The Higuchi algorithm is applied using the pyeeg library (Bao 2018).

Additionally, the other time-domain features are the 6 statistics that are applied using the Numpy and Scipy package in Python (Oliphant 2006; Virtanen et al. 2020). An overview of the 6 statistics used is to be found in table 2. These 6 statistics are picked based on the recommendation of Jenke, Peer, and Buss (2014). Several other studies similarly used these statistics in their emotion prediction study and on top of that, good results were showed in the feature selection experiment showing significantly better results compared to most other techniques (Jenke, Peer, and Buss 2014).

In contrary to the other 2 features, the band power features are frequency domain features. The band power features are extracted with the power spectra density (PSD) using Welch's method and are applied using Scipy (Virtanen et al. 2020). Welch's method is chosen over other methods, due to the results seen in other emotion prediction studies (Soleymani et al. 2015; Jenke, Peer, and Buss 2014). Band powers delta (.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz) are selected and computed in accordance with the similar emotion prediction study Lan et al. (2016).

3.2.2 Setup 2. In the second setup, features are extracted using a time-frequency domain feature which is the Hilbert Huang Transform (HHT). Python library pyhht is being used for the computation of the HHT (Deshpande 2015). This feature extraction method is fairly new and consists of two major steps: Empirical Mode Decomposition (EMD) and Hilbert Spectral Analysis (HSA) (Uzun, Yildirim, and Yildirim 2012). The first step EMD decomposes the EEG signal into a certain amount of Intrinsic Mode Functions (IMFs), which can be compared to other transformations as the fourier transform and the wavelet transform (Huang et al. 1998). An IMF represents the oscillation mode that is embedded in the data and is defined as a function where the number of extrema and the number of zero-crossings differ at most by one (Huang et al. 1998; Uzun, Yildirim, and Yildirim 2012). Additionally, the mean values of the envelopes that are formed by the local minima and local maxima are equal to zero (Uzun, Yildirim, and Yildirim 2012). After the EMD process, the signal is represented by a n amount of IMFs:

$$\xi(t) = \sum_{i=1} nIMF_i + r_n, \tag{4}$$

where r_n equals the residue obtained from EMD for every IMF (Huang et al. 1998). After the decomposition, the second step, the Hilbert transform, is applied on every IMF to obtain the analytical signal ($\hat{x}(t)$ which can be defined by its amplitude (θ) and its instantaneous frequency (Uzun, Yildirim, and Yildirim 2012). The instantaneous frequency, $w(t)$ can be computed by taking the derivative of an analytical signal using the phase function (Uzun, Yildirim, and Yildirim 2012). The following equation describes the Hilbert transform:

$$\widetilde{\xi}(t) = \xi(t) + \hat{\xi}(t) = G(t)e^{j\theta(t)}, \tag{5}$$

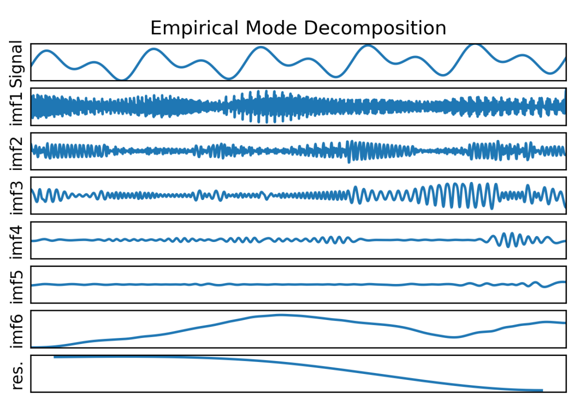
where

$$\hat{\xi} = Hx(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{g(\tau)}{t - \tau} d\tau, \tag{6}$$

and

$$w(t) = \frac{d\theta(t)}{dt}. \tag{7}$$

Figure 4
IMFs plot example



Note. A plot showing an example of all IMFs extracted from a signal, including the signal and the residu.

Combined results of equations 5, 6, and 7 give a 3D representation of the time-frequency-energy distribution, also known as the Hilbert Huang Spectrum (HHS) (Uzun, Yildirim, and Yildirim 2012). The current study does not apply the second step of the HHT. In line with Uzun, Yildirim, and Yildirim (2012), an arbitrary number of IMFs are created of which a set of statistics has been applied. Unlike what has been done in Uzun, Yildirim, and Yildirim (2012), 7 IMFs has been created on which zero paddings has been applied when the algorithm was not able to find a number of 7 IMFs. Additionally, statistics from Table 2 are being used to extract the features from the IMFs

3.2.3 Differential Asymmetry. Supporting evidence has shown that the emotions can be found in EEG by looking at the hemispherical asymmetry of channels (Jenke, Peer, and Buss 2014). Due to this phenomenon, the lateralization has been computed by focussing on the differences between the left hemispherical and the right hemispherical. A feature extraction review shows several methods for computing the asymmetry (Jenke, Peer, and Buss 2014). The current study applies differential asymmetry, which is the most common method for calculating differences in signal. Differential asymmetry is applied by Lan et al. (2016), which is the study on which setup 1 is formed. To stay in line with the feature setup researched by Lan et al. (2016), this work implements differential asymmetry in setup 1, and to control for lateralization it is applied in setup 2. Differential asymmetry can be computed by:

$$\delta x = x_l - x_r, \tag{8}$$

where x_l includes all EEG channels on the left hemisphere and where x_r includes all EEG channels on the right hemisphere.

3.3 Models and channels

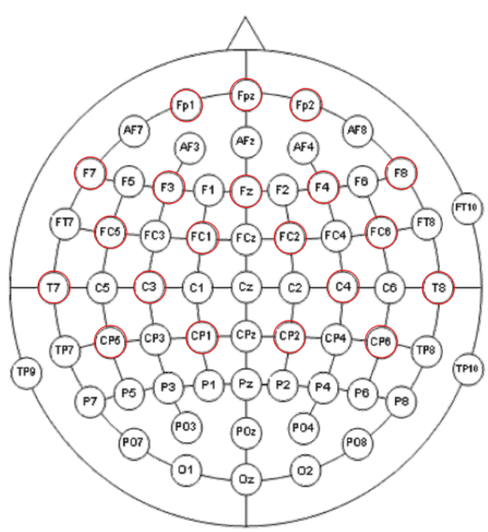
As discussed in relevant work, two classifiers (KNN, SVM) have shown significant performance in emotion prediction research. Additionally, the regression version of the SVM, the SVR, has proven to be suited for predictions according to the dimensional model. The SVR is used in this work as one of the models for regressing emotion, where the regularization parameter (C) and the epsilon parameter are used as hyperparameters. These parameters are optimized using steps of .2 - .5. Additionally, the KNNR is used, derived from the KNN algorithms. The KNNR algorithm is optimized using the number of neighbors variable taking steps of 1. In order to evaluate how these more advanced models perform compared to a basic model, the LR model is applied as a third model. No optimization is done with the LR model. The dataset is split into a 70% training set and a 30% test set with a random state of 42. During processing, the Numpy library was used to get the data in the right shape (Oliphant 2006). All other models are applied using the scikit-learn library in Python (Pedregosa et al. 2011).

All channels selected are shown in Figure 6. A wide selection of channels is taken in line with the study done by Lan et al. (2016). It is chosen to drop several channels that are located the furthest away from the frontal channels. As concluded by Ansari-Asl, Chanel, and Pun (2007), the frontal channels are most relevant for EEG emotion prediction, whereas the relevance of southern channels is significantly less. The channels picked are controlled and are constant for both experiment setups. As described in section 3.2, the differential asymmetry is calculated by taking the difference of the left hemisphere and the right hemisphere. The following pairs (left - right) are selected: FP1 - FP2, F3 - F2, F7 - F8, FC5 - FC6, FC1 - FC2, C3 - C4, CP5 - CP6, CP1 - CP2. Two channels, Fpz and Fz, have been excluded from the differential asymmetry calculation, because of their central position. Although these channels are centered, they are highly relevant due to their frontal position.

3.4 Evaluation

Due to the appliance of the dimensional model, other evaluation metrics are needed for regression rather than classification metrics. In line with other emotion regressing studies, MAE was selected for evaluating the performance of the model. The MAE metric provides a strong indication of the absolute error between the prediction and real value. Considering the 10 point scale rating of the labels, MAE brings a better interpretation compared to a metric as Root Mean Squared Error (RMSE). Several other studies applying the dimensional approach are using MAE, allowing a comparison between studies. Additionally, another evaluation metric is used to evaluate the models from multiple perspectives. Unfortunately, most evaluation metrics are not suited for this kind of task. EEG data is very noisy and the predictions are therefore not spot-on accurate. In common evaluation metrics as Mean Squared Error and R^2, small error stacks up quickly while outliers skew the performance to their side. Haag et al. (2004) proposed a new evaluation metric called bandwidth accuracy, which considers the sample as correct (1) when it is within a certain bandwidth of β. It considers the

Figure 5
EEG channel selection



Note. An illustration showing all EEG channels. Channels with a red border, are selected in the current study.

Table 3
An overview of MAE and Bandwidth results for predicting **arousal** using the two different setups and three algorithms.

	Setup 1		Setup 2	
	MAE	Bandwidth	MAE	Bandwidth
Linear Regressor (LR)	0.626	0.429	0.718	0.407
K-Nearest Neighbor Regressor (KNNR)	0.627	0.442	**0.605**	**0.457**
Support Vector Regressor (SVR)	**0.617**	**0.446**	0.611	0.456

Note. KNNR parameters: *N = 20 and 17; Weights = Uniform and Distance*.
SVR parameters: *C = 0.6 and 0.6; E = 0.05 and 0.1*.

prediction as incorrect (0) when the prediction is not within the range of the bandwidth. The bandwidth of formulated as:

$$A_{bandwidth} = \frac{1}{N} \sum_{i} I(\hat{y}_i, y_i), \tag{9}$$

where

$$I(\hat{y}_i, y_i) = \begin{cases} 0 & \text{if } \|\hat{y}_i - y_i\| < \beta(y_{max} - y_{min}) \\ 1, & \text{else} \end{cases}, \tag{10}$$

where $\| \cdot \|$ denotes the Euclidean distance. The advantage of this evaluation metric is the interpretability. Compared to accuracy for classification, it is a straightforward metric to understand how much test labels are accurately predicted. Additionally, the bandwidth metric normalizes the scale of the bandwidth by applying the max and min. In this way, it is comparable between studies with different data. MAE is in contrary to bandwidth, not easily interpret if the distribution of the labels is uncertain or not clearly mentioned (Jenke, Peer, and Buss 2013). Although there is no domain-width compliance of using the bandwidth, this metric will be used upon recommendation of the evaluation metric review Jenke, Peer, and Buss (2013). A threshold bandwidth of 10% is chosen for this work, in line with how it is carried out in Jenke, Peer, and Buss (2013). As indicated in the related work, an MAE of .65 for arousal and an MAE of .74 for valence is the baseline. The MAE is computed using the Scikit-learn library in Python, while the bandwidth accuracy is computed using a custom function (Pedregosa et al. 2011). All graphs are plotted using the Matplotlib library and the Seaborn library (Hunter 2007; Waskom et al. 2017).

4. Results

Three models are applied to the training dataset using two different experimental setups. Prediction results in terms of MAE and bandwidth are displayed in Table 3 for arousal and in Table 4 for valence. When it comes to the dimension 'arousal' and

Table 4
An overview of MAE and Bandwidth results for predicting **valence** using the two different setups and three algorithms.

	Setup 1		Setup 2	
	MAE	Bandwidth	MAE	Bandwidth
Linear Regressor (LR)	0.722	0.363	0.819	0.322
K-Nearest Neighbor Regressor (KNNR)	**0.695**	**0.379**	**0.696**	**0.358**
Support Vector Regressor (SVR)	0.721	0.381	0.717	0.349

Note. KNNR parameters: *N = 25 and 17; Power = Euclidean distance and Manhatten distance; Weights = Uniform and Distance.*
SVR parameters: *C = 3 and 0.8; E = 0.1 and 0.1.*

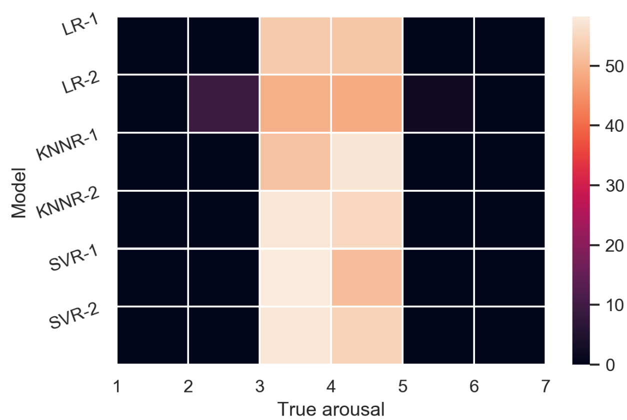
experiment setup 1, relatively similar results are achieved across all three models. The basic Linear Regression model performs just fractionally less compared to the more advanced models. The best performance in this section is observed for the SVR model achieving an MAE of .617 and a bandwidth accuracy of 44.6%. The parameters are set at the optimal point of C = .6 and E = .05. In experiment setup 2 a notable difference can be observed comparing the basic Linear Regression model with the advanced models. Performance of advanced models KNNR and SVR are relatively close where KNNR (N = 17, weights = "distance") performs best with an MAE of .605 and a bandwidth accuracy of 45.7%.

Looking at the prediction results of valence, one model stands out compared to the other models. In experiment section 1, the basic Linear Regression model performs with an MAE of .722 similarly as to the SVR (C = 3, E = .1) model with an MAE of .721. The KNNR model (N = 25) achieves the best MAE score with an MAE of .695. However, the SVR (bandwidth accuracy of 38.1%) scores slightly better compared to KNNR (bandwidth accuracy of 37.9%), when looking at the bandwidth accuracy. In experiment setup 2, the valence performance of the basic Linear Regression model is worse compared to the other advanced models. Similarly as to the arousal scores for experiment setup 2, KNNR (N = 17) performs significantly better with an MAE of .696 and a bandwidth accuracy of 35.8% as compared to the SVR (C = .8, E = .1) with an MAE of .717 and a bandwidth accuracy of 34.9%.

Comparing the different experiment setups for arousal, there is a difference between both setups' best performance. An independent t-test indicates the performance of experiment setup 2 is significantly better compared to experiment setup 1, $t(815) = 6.136$, $p = .000$. Looking more specifically at the results of the predictions in Figure 6, every sector represents the percentage of correctly regressed labels according to the bandwidth per bin. Figure 6 shows distinct results for the 3 bottom models. More specifically, KNNR-2 and SVR-1 perform relatively strong for labels in the third bin (arousal of 3-4) width a respectively 57.3% and a 58.2% bandwidth accuracy. However, KNNR-1 is particularly performing relatively strong in the fourth bin with a bandwidth accuracy of 57.1%. LR-2 has a bandwidth accuracy of 9.3% in the second bin, and a bandwidth accuracy of 2.4% in the fifth bin making it the only model performing better than 0% in the second and fourth bin.

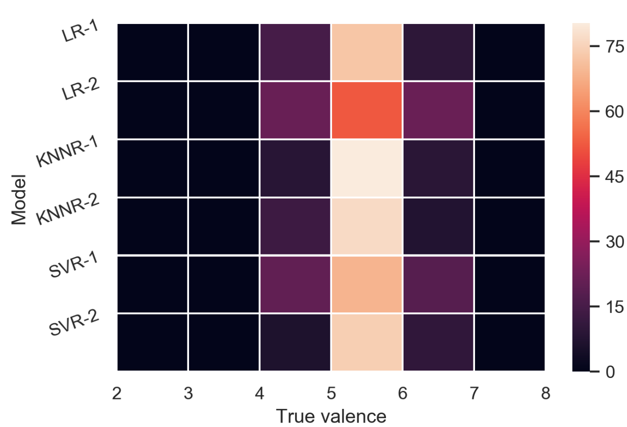
In contrary to results found for arousal, the best performing model in experiment setup 1 and experiment setup 2 for valence achieve similar performance. An independent t-test indicates it is to be concluded that no significant distinctiveness can be found between the two setups for the KNNR model, $t(815) = 1.575$, $p = .116$. Figure 7 shows the percentage correctly regressed labels according to the bandwidth for every model per bin. Best performing models KNNR-1 and KNNR-2 achieve a bandwidth accuracy of respectively 80.4% and 76.8% in the fourth bin (valence of 5-6). SVR-2 achieves a similar

Figure 6
Prediction heatmap for Arousal



Note. A heatmap overviewing which arousal predictions were accurately predicted for experiment setup 1 (model names that end with 1) and experimental setup 2 (model names that end with 2).

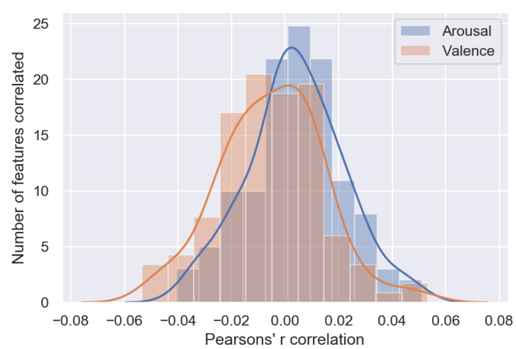
Figure 7
Prediction heatmap for Valence



Note. A heatmap overviewing which valence predictions were accurately predicted for experiment setup 1 (model names that end with 1) and experimental setup 2 (model names that end with 2) bandwidth in bin 4 with 74.2%, although other bins do not perform well enough to achieve similar overall scores as KNNR-1 and KNNR-2. Notable results are predicted for bin 3 and 5 with LR-2, achieving a bandwidth accuracy of 21.4% and 21.7%. However, the bandwidth accuracy of the fourth bin is 52.2%, lowering the overall score of the model.

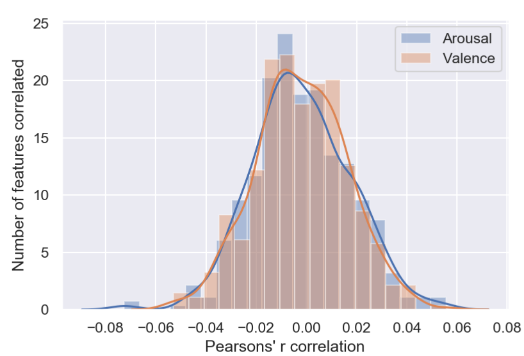
A two-dimensional matrix is showed in Figure 10 comprehending predictions from the best performing arousal model KNNR-2 and the best performing valence model KNNR-1, compared with true values for arousal and valence. It is to be observed that the predictions are centered at an arousal of around 4 and a valence of around 5.5, which is close to the mean of the true labels that are predicted (arousal mean = 4.04, valence mean = 5.44). To further analyze the impact of which features create a two-dimensional matrix

Figure 8
Correlation plot for arousal



Note. A plot indicating the correlations of arousal and valence with all features extracted in setup 1.

Figure 9
Correlation plot for Valence



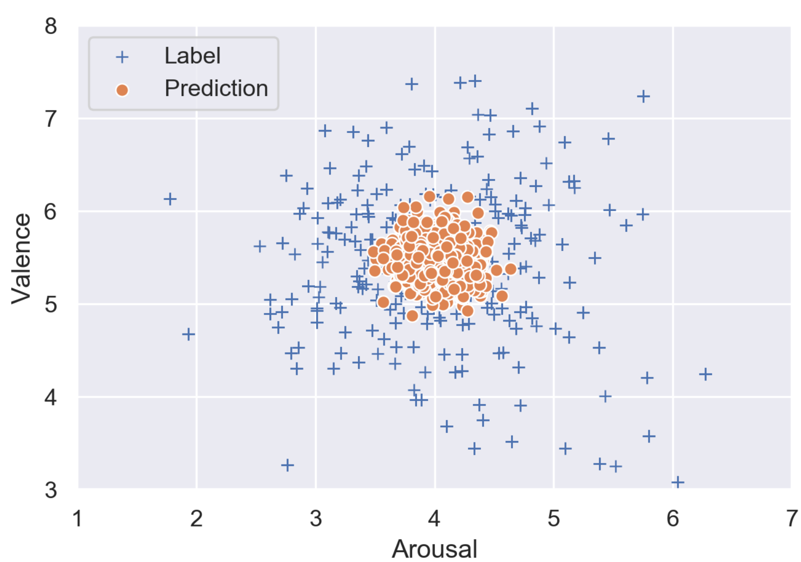
Note. A plot indicating the correlations of arousal and valence with all features extracted in setup 2.

as shown in Figure 10, correlations between emotion (arousal and valence) and all features are extracted using Pearson's r correlation. In Figure 8 and Figure 9 are showing a plot visualizing a distribution of Pearson's r correlations between arousal and valence, and the features that are used. In Figure 8, it is to be noted that in the distribution of setup 1, a significant part of the features (43.8% for arousal, 41.3% for valence) are slightly or not correlated with emotion (-.01 - .01). The majority of features have a small correlation, where 56.2% for arousal and 58.7% for valence have a correlation between -.01 and -.1 or between .01 and .1. A similar distribution is observed for setup 2 wherein percentages there are more small correlations. More specifically, with 60.4% and 59.7% of the features for arousal and valence, a small correlation (between -.01 and -.1 or between .01 and .1) is observed, which is significantly more features considering that the second setup contains almost 4 times as many features as the first setup.

5. Discussion

Despite the considerable amount of progress that has been made in the emotion prediction field, this study focuses on emotion prediction while tackling two major challenges in the domain of EEG emotion prediction: Applying a passive emotion elicitation method with extracting EEG signals; Adopting a proper theoretical framework for emotion. As is well known in the EEG domain, EEG signals quite noisy and volatile. It is therefore logical that strong emotion elicitation methods have been applied to draw out strong emotional responses for analysis. Soleymani et al. (2011) introduced the possibility of classifying emotion while applying a passive emotion elicitation method.

Figure 10
A comparison of predictions and true values



Note. A two-dimensional matrix showing predictions that are made, in orange, with the KNNR-2 model for arousal and with the KNN-1 model for valence. True values are showed as blue "+" symbol.

For this reason, emotion prediction for a similar emotion elicitation method based on the regression approach is introduced.

Additionally, many different studies have applied the discrete perspective on emotion, treating different emotions as unique combinations of physiology, experience, and behavior. However, scientific evidence has shown that a dimensional perspective would favor emotion prediction. As a consequence, several studies have applied a regression approach creating an MAE baseline of .65 for arousal and an MAE of .75 for valence. As shown in Table 3, the best performing model achieves an MAE of .605 for arousal, improving the baseline score with .045. A significant difference in result is to be found between the two setups, favoring setup 2 where the Hilbert Huang Transform has been applied. Results in Table 4 show that the best performing model for valence achieves an MAE score of .695, which is an MAE that is .045 lower than the baseline score for valence. However, it is to be noted that the MAE can be biased due to the range of the labels. Labels in have been computed by averaging emotion scores for the adjective and for the noun, to predict the emotion of the sentence. Although the scores are based on a 10-point scale, the range of arousal and valence might be smaller compared to other studies, depending on their study setup. Therefore, the distribution of both arousal and valence are displayed in Figure 3 for future studies to compare. As a consequence, a new metric is implemented, bandwidth accuracy, to be used for emotion prediction regression as is introduced in Haag et al. (2004). The threshold bandwidth that is applied is 10%, based on how it is implemented in Jenke, Peer, and Buss (2013). However, there is currently no consensus regarding what this threshold bandwidth should be in the EEG emotion prediction domain. More future research should be done to investigate what level of bandwidth threshold is suited for emotion prediction.

Furthermore, the current research investigated two major experiment setups that were in line with two leading studies for emotion prediction that were built on the regression/dimensional approach. In the first setup, statistical features, band powers, and fractal dimensions were extracted from the EEG signals. Eventually, right hemisphere channels were deducted from the left hemisphere channels, to calculate the lateralization. Two channels, Fpz and Fz, have been excluded from this process. The second setup applied a Hilbert Huang Transform method by calculating IMFS and taking the statistical features of those IMFS. Similar to the first setup, the lateralization has been calculated. A comparison between the models clearly shows a significant difference for arousal, although for valence there is no significant difference. For arousal, setup 2 would favor with a small improvement in metrics over setup 1.

Although the model performance was significantly better than baseline methods, the performance does not allow it yet to be practically implemented in a business application. As shown in Figure 10, predictions are centered around the mean, meaning that sentences with a valence or arousal level around the mean would be more likely to be predicted correctly. This kind of pattern is not suitable for a business application that needs to give reliable predictions. An emotion prediction application for, for example, a marketing communication campaign would require to have accurate predictions so that it can benefit decision making. Therefore, more research is necessary to investigate several other aspects of emotion prediction. One of these aspects would be the implementation of either self-reported emotion labels or collectively agreed labels coming from reliable datasets as Vo et al. (2009). Additionally, more research is necessary to understand which features combined with which channels are allowing better performance.

6. Conclusion

The current study introduces an EEG emotion prediction research using EEG signals that are relatively passive elicited by reading German sentences. Additionally, the current study proposes a regression approach based on the dimensional models, comprehending the dimensions ’arousal’ and ’valence’. Labels for arousal and valence per sentence are retrieved by averaging the adjective and noun labels, retrieved from a corpus created by Köper and Im Walde (2016). EEG signals have been processed with a zero-phase band-pass FIR filter combined with a pass band-edge. Additionally, ICA has been applied to correct for eye movements. Two different experiment setups are proposed containing several different feature extraction methods. The first experiment setup consists of a combination of FD features, STAT features, and POWER features. The second setup consists of statistical features extracted from the Hilbert Huang Transform. All features are computed using the differential asymmetry formula to obtain information about the lateralization of the signals. Furthermore, performance on three models (LR, KNNR, SVR) has been evaluated by looking at the metrics MAE and the relatively new metric bandwidth accuracy.

The best performance for arousal is found in the KNNR model with experimental setup 2. The KNNR model for setup 2 is found to be significantly better compared to the best model in setup 1. The KNNR model for setup 2 achieved an MEA of .605 and a bandwidth accuracy of 45.7%, improving the MAE ’arousal’ baseline of .645. For valence, similar results are found for setup 1 compared with setup 2, where the KNNR model performs best in both setups with the best MAE of .695 and a bandwidth accuracy of 37.9%, improving the MAE ’valence’ baseline of .75. Correlations between emotion and features show us that the majority of features have a small correlation, although no strong correlations are found. Quantitywise, more small correlations are found in setup 2.

It can be derived from these results that a passive emotion elicitation method can be used for improving the ecological validity. Additionally, a regression approach for emotion prediction has been recommended by the literature and shows improving performance in the scarce studies that are available. The current work shows improvements of that approach and introduces a new evaluation metric, bandwidth accuracy, to better evaluate the regressive performance of emotion predictions. Although performance is improving, there is quite some room for improvement. In order to practically implement emotion predictions in business applications such as a marketing communication evaluation tool, predictions need to be more accurate. More research is necessary to better understand emotion, the impact of features, and the impact of EEG channels. Conclusively, an advanced feature extraction method such as the HHT is recommended in opposition to a combination of FD, STAT, and POWER features. In addition, a KNNR model is a popular model in emotion classification, and the regressing version, KNNR, is also to be recommended for future EEG emotion prediction studies selecting the regressive approach.

References

Ackermann, Pascal, Christian Kohlschein, Jó Ágila Bitsch, Klaus Wehrle, and Sabina Jeschke. 2016. Eeg-based automatic emotion recognition: Feature extraction, selection and classification methods. In *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)*, pages 1–6, IEEE.

Ansari-Asl, Karim, Guillaume Chanel, and Thierry Pun. 2007. A channel selection method for eeg classification in emotion assessment based on synchronization likelihood. In *2007 15th European Signal Processing Conference*, pages 1241–1245, IEEE.

Bao, Forrest. 2018. pyeeg.

Bhatti, Adnan Mehmood, Muhammad Majid, Syed Muhammad Anwar, and Bilal Khan. 2016. Human emotion recognition and analysis in response to audio music using brain signals. *Computers in Human Behavior*, 65:267–275.

Davidson, Richard J. 1999. Neuropsychological perspectives on affective styles and their cognitive consequences.

Deshpande, Jaidev. 2015. pyhht.

Haag, Andreas, Silke Goronzy, Peter Schaich, and Jason Williams. 2004. Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems*, pages 36–48, Springer.

Hassanien, Aboul Ella, Moataz Kilany, Essam H Houssein, and Hameed AlQaheri. 2018. Intelligent human emotion recognition based on elephant herding optimization tuned support vector regression. *Biomedical Signal Processing and Control*, 45:182–191.

Hu, Xin, Jingjing Chen, Fei Wang, and Dan Zhang. 2019. Ten challenges for eeg-based affective computing. *Brain Science Advances*, 5(1):1–20.

Huang, Norden E, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995.

Hunter, J. D. 2007. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Jenke, Robert, Angelika Peer, and Martin Buss. 2013. A comparison of evaluation measures for emotion recognition in dimensional space. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 822–826, IEEE.

Jenke, Robert, Angelika Peer, and Martin Buss. 2014. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 5(3):327–339.

Kanske, Philipp and Sonja A Kotz. 2010. Leipzig affective norms for german: A reliability study. *Behavior research methods*, 42(4):987–991.

Köper, Maximilian and Sabine Schulte Im Walde. 2016. Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 german lemmas. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2595–2598.

Lahl, Olaf, Anja S Göritz, Reinhard Pietrowsky, and Jessica Rosenberg. 2009. Using the world-wide web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 german nouns. *Behavior Research Methods*, 41(1):13–19.

Lan, Zirui, Gernot R Müller-Putz, Lipo Wang, Yisi Liu, Olga Sourina, and Reinhold Scherer. 2016. Using support vector regression to estimate valence level from eeg. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 002558–002563, IEEE.

Liu, Yisi and Olga Sourina. 2013. Real-time fractal-based valence level recognition from eeg. In *Transactions on computational science XVIII*. Springer, pages 101–120.

Lüdtke, Jana and Arthur M Jacobs. 2015. The emotion potential of simple sentences: additive or interactive effects of nouns and adjectives? *Frontiers in psychology*, 6:1137.

Lynch, Joanne and Leslie De Chernatony. 2004. The power of emotion: Brand communication in business-to-business markets. *Journal of Brand management*, 11(5):403–419.

Mauss, Iris B and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and emotion*, 23(2):209–237.

Müller, Kirill and Hadley Wickham. 2017. tibble: Simple data frames. r package version 1.3. 3.

Musha, Toshimitsu, Yuniko Terasaki, Hasnine A Haque, and George A Ivamitsky. 1997. Feature extraction from eegs associated with emotions. *Artificial Life and Robotics*, 1(1):15–19.

Nicenboim, Bruno, Shravan Vasishth, and Frank Rösler. 2020. Are words pre-activated probabilistically during sentence comprehension? evidence from new data and a bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, page 107427.

Oliphant, Travis E. 2006. *A guide to NumPy*, volume 1. Trelgol Publishing USA.

Panksepp, Jaak. 2007. Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on psychological science*, 2(3):281–296.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Rani, Pramila, Changchun Liu, Nilanjan Sarkar, and Eric Vanman. 2006. An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9(1):58–69.

Sohaib, Ahmad Tauseef, Shahnawaz Qureshi, Johan Hagelbäck, Olle Hilborn, and Petar Jerčić. 2013. Evaluating classifiers for emotion recognition using eeg. In *International conference on augmented cognition*, pages 492–501, Springer.

Soleymani, Mohammad, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. 2015. Analysis of eeg signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, 7(1):17–28.

Soleymani, Mohammad, Sander Koelstra, Ioannis Patras, and Thierry Pun. 2011. Continuous emotion detection in response to music videos. In *Face and Gesture 2011*, pages 803–808, IEEE.

Soroush, Morteza Zangeneh, Keivan Maghooli, Seyed Kamaledin Setarehdan, and Ali Motie Nasrabadi. 2017. A review on eeg signals based emotion recognition. *International Clinical Neuroscience Journal*, 4(4):118.

Soroush, Morteza Zangeneh, Keivan Maghooli, Seyed Kamaledin Setarehdan, and Ali Motie Nasrabadi. 2018. Emotion classification through nonlinear eeg analysis using machine learning methods. *International Clinical Neuroscience Journal*, 5(4):135–149.

Takahashi, Kazuhiko et al. 2004. Remarks on emotion recognition from bio-potential signals. In *2nd International conference on Autonomous Robots and Agents*, pages 186–191, Citeseer.

Uzun, S Sinem, Serdar Yildirim, and Esen Yildirim. 2012. Emotion primitives estimation from eeg signals using hilbert huang transform. In *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, pages 224–227, IEEE.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*.

Vo, Melissa LH, Markus Conrad, Lars Kuchinke, Karolina Urton, Markus J Hofmann, and Arthur M Jacobs. 2009. The berlin affective word list reloaded (bawl-r). *Behavior research methods*, 41(2):534–538.

Waskom, Michael, Olga Botvinnik, Drew O'Kane, Paul Hobson, Saulius Lukauskas, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, Jordi Warmenhoven, Julian de Ruiter, Cameron Pye, Stephan Hoyer, Jake Vanderplas, Santi Villalba, Gero Kunter, Eric Quintero, Pete Bachant, Marcel Martin, Kyle Meyer, Alistair Miles, Yoav Ram, Tal Yarkoni, Mike Lee Williams, Constantine Evans, Clark Fitzgerald, Brian, Chris Fonnesbeck, Antony Lee, and Adel Qalieh. 2017. mwaskom/seaborn: v0.8.1 (september 2017).

Yu, Liang-Chih, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Lai, and Xuejie Zhang. 2016. Building chinese affective resources in valence-arousal dimensions.

Zheng, Wei-Long and Bao-Liang Lu. 2015. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175.

Zong, Cong and Mohamed Chetouani. 2009. Hilbert-huang transform based physiological signals analysis for emotion recognition. In *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 334–339, IEEE.