Exploring Automatic Translation between Affect Representation Schemes

Video Affective Content Analysis

Ivan Dimitrov[1]

Supervisor(s): Chirag Raman[1]**, Bernd Dudzik**[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Ivan Dimitrov
Final project course: CSE3000 Research Project
Thesis committee: Chirag Raman, Bernd Dudzik, Alan Hanjalic

Abstract

The objective of this report is to establish and present a machine learning model that effectively translates affect representation from emotional attributes such as arousal (passive versus active) and valence (negative versus positive) to dominance (weak versus strong). In the pursuit of this goal, various research questions are addressed. The paper outlines the process of dataset selection, ensuring appropriateness for the problem at hand. Subsequently, a comprehensive investigation into suitable evaluation methods for the developed model is conducted, providing well-reasoned justifications for the chosen approach. An additional research question focuses on assessing different machine learning approaches to determine the optimal performer. The motivation behind this translation lies in the recognition of the interdependence between these affect attributes, supported by both theoretical underpinnings and practical evidence. This contrasts with previous studies that have treated these dimensions as independent descriptors for representing emotions.

1 Introduction

In contemporary times, affective analysis, a field that delves into the study of emotions and their impact, has gained significant relevance. In particular, video-based content has emerged as the foremost means of experiencing entertainment, catching up with major news, and media articles via platforms such as movies and YouTube[1]. With the potential to evoke emotions, it has caught the attention of scientists from various research fields such as psychology, sociology, neuroscience, and computer science, who are attempting to develop models to evaluate and measure emotions[21] during video content exposure.

This task however proves to be quite demanding as emotions are highly multi-dimensional and expressed differently by various individuals. This leads to researchers with contrasting fields of study finding it exceptionally difficult to exchange and reuse each others' work and papers as agreeing upon the way emotions are represented and measured could be onerous.

The study of video affective content analysis has thus gained traction and significant importance due to the increasing use of videos in everyday life[8]. With the growing availability of video content via social media, news outlets, etc. and the increasing importance of understanding emotions on a deeper level and more intelligently in diverse fields, such as marketing[1] and entertainment[6], researchers especially in computer science fields are under increasing pressure to develop reliable and effective methods,models and databases with appropriate multiple affect representations. These methods need to be robust enough to account for the variability in emotional expression and understandings across different individuals, cultures, and contexts.

[1]https://www.youtube.com/

As a result from these demands a lot of related work has already been done in the field of video affective content analysis[24] also in music[9] and image[25]. The aim of this report is more about translating between different affect representations unlike the work listed above which focuses more on collecting affect data from participants using some of the following techniques: SAM(Self-Assessment Manikin)[4] or the AffectButton[5]. In this paper a very methodological way of researching and collecting appropriate datasets for the issue at hand, some with more than one affect representation for the same video by the same person. After the gathering some machine learning models were developed and then evaluated using comparison between mean squared errors (MSE)[12] and R-Squared[10] scores. Then another question which needs answers is how can we generalize the translation models to other unseen datasets and also what is the feasibility of this issue. And at the end, a best performer out of the models would be picked and analyzed. Due to some difficulty finding satisfactory large dataset with multiple affect representations we decided to tackle a more niche problem whether there is correlation between pleasure,arousal and dominance.

This particular question of dependence in this 3D affect representation has already been tackled as you will be presented in the related work section (See section 2) this paper aims to go further into that question while also observing the intercorrelation between each of the emotion representations. Our goal is to develop the machine learning models, optimize them, evaluate them and remark what the results achieved mean for the problem at hand and also discuss what does this mean for further papers and experiments which use PAD[2] as their affect representation.

Overall, the study of video affective content analysis is an area of growing importance that requires collaboration and coordination across multiple disciplines to develop effective models for evaluating and measuring emotions during video content exposure. Therefore any development done in that direction is of help and will benefit the future work done in that field of study.

2 Related work

Previous studies have shown dependencies between arousal and dominance[20; 19; 13].

Russell et al. (1977) argues for showing that pleasure, arousal and dominance are all sufficient and necessary to describe emotions unlike some beliefs that dominance because of being newer dimension is not as important.It conducts two studies and they both show in different ways that not only are the 3 dimensions important equally but also spots some intercorrelations in them which accounts for variance in scores.For example an intercorrelation of .40 for pleasure and dominance,.15 for arousal and dominance. The paper also defends the use of dominance because despite the fact that there is correlation and one could argue that it can be done without the dominance there are examples of some emotions needing the dominance dimension to be assessed accurately such as distinguishing angry from anxious, alert from surprised and relaxed from protected.

Another study conducted by Russell et al. (1978) also consists of several experiments which strengthen his research conclusion in the forementioned article with some similar intercorrelations and reliability of the dimensions examined.

These observations suggest that there is a realistic need and possibility of a solution which uses valence and arousal to predict dominance values. This is even more appealing considering the fact that some other affect representation schemes such as discrete categories of emotions like happiness, anger and sadness or PANAS(Positive and Negative Affect Schedule)[23] have major drawbacks.For example the using discrete emotions has negative effect when needed to scale and might not always be consistent across different people and stimuli but using the 3D representation of PAD[2] accounts for the same amount of variability despite the widely different stimuli used [13].Extending the number of classes leads to sparseness and unbalance in the distribution of emotional classes.For those reason using few emotional attributes is highly appealing.

Another paper tackles an issue very similar to this one[16]. In this article we S.Parthhasarathy and C.Busso show that there are interrelations between the 3 dimensions and develop a multi-task learning (MTL) framework implemented with deep neural networks (DNN) with shared hidden layers. Their model using MTL jointly predicts arousal, valence and dominance by setting a target emotional attribute considered as primary task and treating the others as secondary tasks.

The contribution this paper aims to provide is further explore the possibility and reliability of translating valence and arousal emotional dimensions to dominance. Relying on the previously discussed correlation between the emotional categories a various number of machine learning models are implemented from simple linear regression to tree regressors and their performance is compared to each other and to a dummy regressor which should provide concrete and further evidence to whether such dependences exist.

3 Methodology and Problem Description

In this section the reader will be presented with the methodology of our work and then a formal problem description.

3.1 Methodology

This subsection outlines the exploration procedure employed to explore automatic translation between affect representation schemes in videotape affective content analysis and address the exploration questions. The methodology encompasses data collection, evaluation procedure, dataset analysis, machine literacy approaches, performance evaluation, and limitations. All of these steps in development were treated as sub-research questions.

For data collection, a methodical approach was taken. A comprehensive literature review was conducted, consulting academic databases and repositories. A plethora of datasets were found a lot of them only using one affect representation in the following article[3] such as LIRIS-ACCEDE, HUMAINE, EMDB etc. some also used physiological emotion depiction but that is of no use for our goal. Another set of datasets was found such as FilmStim and CP-QAE-I but the first one lacks in size and for the second no EULA's or PDF's were found. And since the named datasets were needed to be applicable to videotape affective content analysis and no appropriate ones were identified the research question at hand was altered to examining the dependence of dominance on arousal and valence in the 3D PAD representation.

After settling on using the Mementos dataset[7] dataset we moved onto the second step of pre-processing the data and then developing various ML models. To ensure that there was no bias while training the model the data was split in train and test sets while guaranteeing that there were no entries by the same person or the same movie in both sets. This assures that the machine learning models would meet only unique entries while predicting on the test set. Then the problem of evaluating the performance of these models was tackled by testing a few approaches and choosing the the best measures: MSE and R-Squared. MSE was chosen to keep track of how much better the current model is in predicting closer to the correct value of dominance. The more important score and the one we focused more on while deciding how good a model is is R-Squared since it is a statistical measure used to assess the goodness of fit of a regression model. It provides an indication of how well the dependent variable is explained by the independent variables in the model. Together with these scores a number of graphs such as QQ Plot[15] have been generated to better visualize the results which we obtained.

To insure reproducibility and validity, a detailed experimental setup will be handed. This included information on tackle, software, hyperparameters, and any preprocessing was applied to the datasets. The performance of the machine learning approaches was estimated using the named criteria . Statistical and qualitative analysis were performed to compare the models' performance and gain perceptivity into their strengths and limitations. Ethical considerations were taken into account throughout the exploration process, including proper citation, acknowledgments, and warrants for the datasets used.

After observing the performance some other machine learning algorithms were implemented such as decision trees and a simple neural network. The exploration methodology acknowledges limitations similar as implicit impulses in dataset selection and constraints in the evaluation procedure. The impact of these limitations on the validity and generalizability of the findings was precisely considered. In conclusion, the exploration methodology employed a methodical approach for data collection, a valid evaluation procedure, dataset analysis, disquisition of machine literacy approaches, detailed experimental setup, comprehensive performance evaluation, ethical considerations, and acknowledgment of limitations. The methodology aligns with established practices and aims to effectively address the exploration questions.

3.2 Problem description

Automatic emotion recognition systems can be categorized into two main tasks within the research domain. The first task involves the identification and classification of discrete categories of emotions, encompassing emotions like happiness, anger, and sadness, among others. The second task involves the prediction of values associated with emotional attributes,

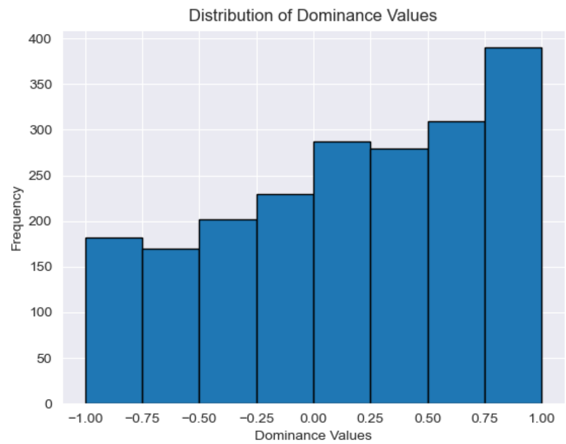

Figure 1: Distribution of Dominance Values in the original data in Mementos.

including arousal, valence, and dominance. The challenge of classifying emotions lies in the intricate nature of human interaction, as acknowledged in the study on human interaction conducted[14]. This complexity makes it difficult to categorize emotions into a limited number of distinct classes. Moreover, expanding the number of classes to accommodate a wider range of emotions leads to issues of sparseness and imbalance in the distribution of emotional classes. Consequently, the utilization of a reduced set of emotional attributes appears highly advantageous. In order to predict the values of emotional attributes, many systems rely on machine learning algorithms that are trained using signals from one or more modalities. These modalities typically encompass acoustic, facial, and physiological signals, such as electrocardiograph (ECG) and electroencephalogram (EEG), as evidenced in studies[18; 22].

Given the disadvantages with systems trained using categorical affect representation this has sparked our interest in further researching the dependence in the 3D PAD space and also the importance of dominance in evaluating emotions.

3.3 Dataset selection

In the search for dataset as previously stated we found some with multiple affect representations but they were not satisfactory in size so we decided to aim for ones that use 3D PAD representation. A multimodal corpus called Mementos fit our requirements so we used it to computationally describe how affect and memory are processed in response to video input. It includes 1995 unique replies from 297 individual viewers who responded to 42 different music video portions. It was gathered online through crowdsourcing. The dataset contains the three dimensions pleasure, arousal and dominance stored as continuous values in the range of [-1,1] (Figure 1) representing viewers' rating for the emotional experience that the presented video stimulus elicited in them.

In addition to 2012 minutes of webcam recordings of their upper-body behavior and self-reports of their emotional experience, it also includes in-depth explanations of the incidence and content of 989 individual memories that were brought on by the video content. The dataset also includes viewer-specific self-report measures of individual differences in participants' backgrounds and circumstances (Demographics, Personality, and Mood), which makes it easier to explore key contextual elements in research that builds on it and this helps building the model for this study.

Another dataset found with the desired affect representation is the DEAP[11] dataset. In it the ratings from an online self-assessment where 120 one-minute extracts of music videos were each rated by 14-16 volunteers based on arousal, valence and dominance are recorded and the participant ratings, physiological recordings and face video of an experiment where 32 volunteers watched a subset of 40 of the above music videos. EEG and physiological signals were recorded and each participant also rated the videos as above. For 22 participants frontal face video was also recorded.

3.4 Machine learning models

Algorithms and packages used in the development:

- Scikit-learn(1.0.2.)[17]:
  - Linear Regression - fits a linear model with coefficients w = (w1, . . . , wp) to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation.
  - Neural Network - A sequential model is created using the Sequential class from the Keras library. The model consists of three densely connected layers. The first two layers have 64 units and use the ReLU activation function, and a dropout layer is added after each of them to prevent overfitting. The final output layer has a single unit with a linear activation function.
  - GradientBoostingRegressor - This estimator builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.
  - Mean (control) - In a mean model, the predicted value for every input sample is the mean value of the target variable in the training data. This means that regardless of the input features, the model predicts the same constant value for all samples.
  - Median (control) - Similarly, in a median model, the predicted value for every input sample is the median value of the target variable in the training data. Again, this means that the model predicts the same constant value for all samples.
  - Grid Search - Implements a "fit" and a "score" method. It also implements "score_samples", "predict", "predict_proba", "decision_function", "transform" and "inverse_transform" if they are implemented in the estimator used. The parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid.

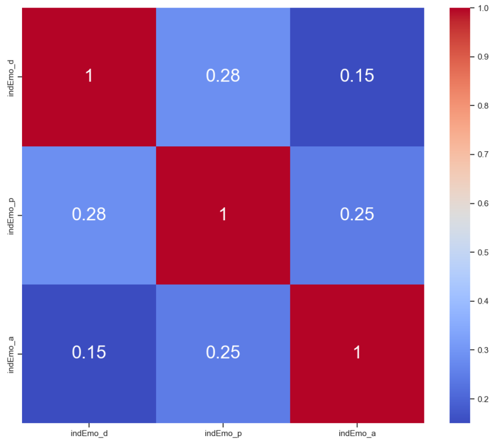

Figure 2: Intercorrelation between Pleasure,Arousal,Dominance for the Mementos dataset.

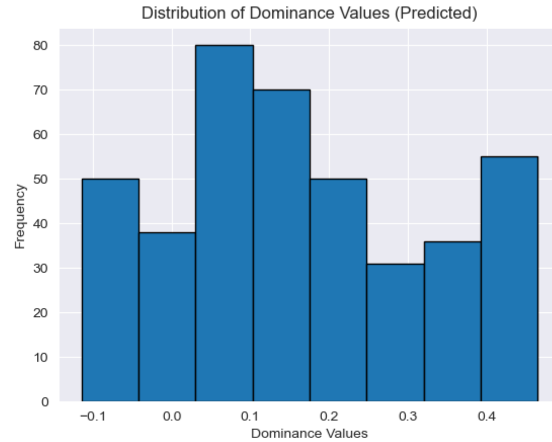

Figure 3: Distribution of dominance values predicted using Linear Regression

4 Experimental Setup and Results

Before starting with the modelling process a decision was made to keep the data in the original continuous state and not translate it to discrete since it is not the aim of this research to achieve great accuracy which will be highly unlikely, but to show that there is some interdependence between pleasure,arousal and dominance. So that way we keep the pre-processing to a minimum since we only need to set the valence and arousal labels as training set and the dominance as the target.

It was already observed in the related work section that intercorrelation between the dimensions is not unlikely and the same was found in the mementos dataset(Figure 2) we can see that there is intercorrelation between pleasure and dominance with a rating of 0.28 and arousal and dominance with a rating of 0.15. Which already display some better correlation than what was observed by Russel al. 1977. So we can expect better R-Squared results than what he was able to achieve of 5%.

Setup One

The following models and results are done using a constant random_state = 0 to compare all results in one particular instance of the dataset test/train split.

Before any models started to be developed a simple mean/median approach was taken in order to obtain a mean squared error(MSE) and R-Squared scores to which we can compare the latter models and examine either the benefits or the losses compared to this approach. Results can be seen here in the following table (Table 1). As expected the R-Square scores obtained are very close to 0 which means we have a good baseline for MSE and looking if our model can also obtain better accuracy in addition to improving the proportion of the variance in dominance explained by pleasure and arousal. This was also tackled so the feasibility question could be answered. It would show whether our models are better than chance/proportional guessing and if the there is a possibility of finding a solution to the problem of dependence between the dimensions.

The first model that was implemented and tested was linear regression there the following results were achieved (Table 1). By investigating both of these results it can be seen that a simple model like linear regression already performs slightly better than mean/median approach but still not great in terms of accuracy. But where large improvement can be observed is the R-Squared score of 0.106 which means that even a straightforward linear regression model can be used to show some dependence between the emotional dimensions. Another valuable remark can be made when looking in the distribution of dominance values predicted by the linear regression (Figure 3) is that the values are only in the range [-0.1,0.4] which shows us that the model is not complex enough to capture the full range of variation in the data because of the possibility that the relationship between the labels is non-linear.

That is why the next model developed was Decision Tree Regressor. Here a Grid Search was used so optimal hyperparameters would be achieved.The following results were gathered from this model (Table 1) together with the best parameters found:

- Best Parameters: 'max_depth': 3, 'max_features': 'sqrt', 'min_samples_leaf': 2, min_samples_split': 2

It can be seen that no benefits in terms of score were achieved by the Decision Tree.

The next model implemented is Gradient Boosting Regressor(GBR). Again Grid Search was used to achieve the best result possible (Table 1) together with tuned hyperparameters:

- Best Parameters: 'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 1000, 'random_state': 0, 'subsample': 1.0

Here some sizeable gains can be seen in both measures. Firstly, the MSE is down from the other two models by approximately 0.069 which is 22.5%. But the value which helps

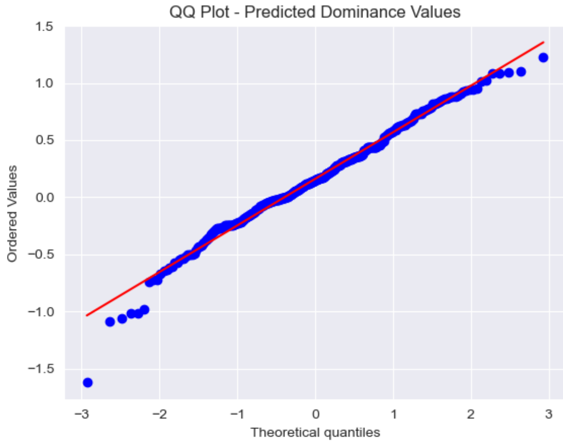

Figure 4: Gradient Boosting Regressor QQ plot.

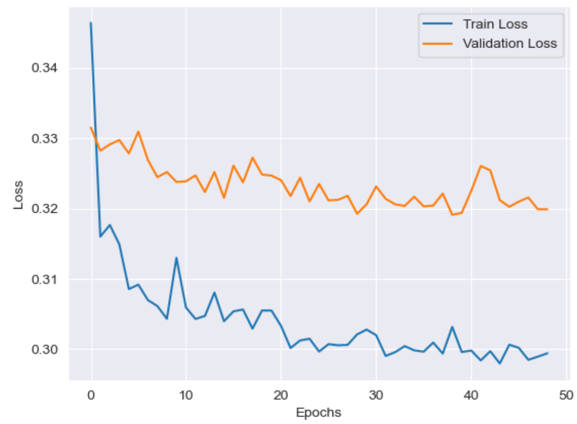

Figure 5: Training and Validation Loss Over Epochs.

the aim of this research is the R-Squared value which indicates that approximately 34.1% of the variance in the target (in our case dominance) can be explained by the independent (shown in [20]) variables (in our case pleasure and arousal). We have also generated a QQ Plot for the results of this model (Figure 4) where we can clearly observe that points fall roughly in a straight line, which indicates that the dataset follows the theoretical distribution closely and also presents more evidence of how the GBR captures the dependence of our target label.

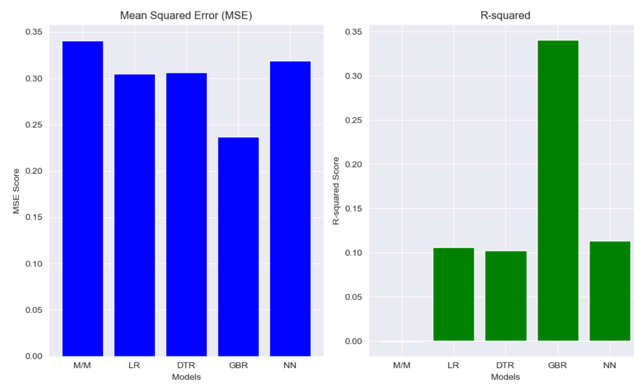

Figure 6: Results for MSE and R-Squared achieved by the models.

Table 1: MSE and R-Squared results for each model using random_state = 0.

Model	MSE	R-Squared
Mean/Median	0.36	-0.00
Linear Regression	0.305	0.106
Gradient Boosting Regressor	0.237	0.341
Decision Tree Regressor	0.306	0.102
Neural Network	0.32	0.114

We also developed a Neural Network and we observed overfitting on the training data which indicates two things: either the model is too complex or we have not enough data to work with. First the NN was ran with Grid Search to find optimal parameters and after noticing that the network was achieving good MSE in later epochs but the model's overall MSE when tested on the test set was not good it was drawn the conclusion that overfitting is happening so an early stopping criteria was added to the model (Figure 5) which stops training after not gain is achieved in a period of 10 epochs and the model constantly stopped learning after 40-50 epochs which shows that the NN is sufficiently trained on that time but only can achieve R-Squared of 0.114 barely any better that the other two models.

As expected none of the models can achieve even marginal accuracy but that is not the aim of this research but it is to simply show that there is some dependence between pleasure,arousal and dominance which is partly achieved by the Gradient Boosting Regressor. Also for better generalization all results were plotted in Figure 6.

Setup Two

The following models and results are done using a random repeated sampling for random_state from 0 to 100 so we can achieve robustness, reliability and sensitivity analysis.

For this data setup we observe more or less the same behaviour amongst the models. Since we have a larger set of results here we can also look at best/worst values achieved and do some statistical evaluation of how good or bad are the models. For example in both Figure 7 and Figure 8 we can observe the significantly better performing GBR model when compared to Linear Regression and Decision Tree Regressor but also these two models' boost in performance when brought up to the control models of the Mean/Median approach we discussed in the beginning of the section.

Another interesting statement that can be reached while monitoring the best/worst (Tables 2 and 3) results from these models are e.g. the best R-Squared score for Gradient Boosting Regressor which is as high as 0.488 which is a witness to how much of the variation in the dominance dimension can

Figure 7: Results for MSE and R-Squared achieved by the models.

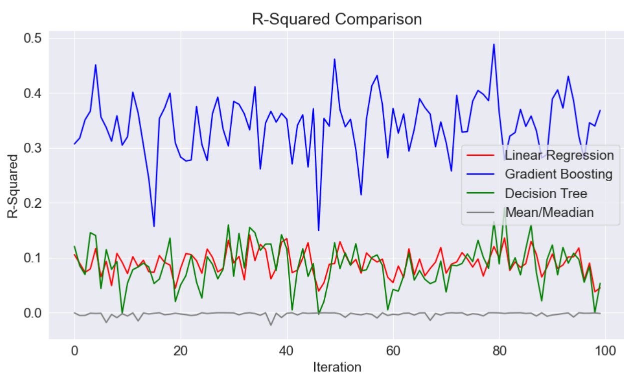

Figure 8: Results for MSE and R-Squared achieved by the models.

be explained by valence and arousal variables which further strengthens our stand that there is high dependence between the emotion dimensions. Also another remarkable score is the best MSE score for the same model which gets 0.173 showing that not only it accounts the variance but also gets considerably close to the true values of the dominance label.

Table 2: Best and Worst MSE results achieved by the models during random repeated sampling.

Model	Worst	Best
Mean/Median	0.372	0.301
Linear Regression	0.348	0.266
Gradient Boosting Regressor	0.301	0.173
Decision Tree Regressor	0.357	0.266

5 Responsible Research

In translating the pleasure-arousal dimension to the dominance dimension, we have taken a responsible and rigorous approach to ensure accurate representation and meaningful interpretation of the data. The process involved several key steps:

- Understanding the original framework: We thoroughly studied the work of Bakker et al. (2014) to gain a comprehensive understanding of the pleasure-arousal dimension and its relationship to dominance.[2]
- Conceptual alignment: We carefully examined the conceptual and theoretical foundations of both dimensions to identify any overlaps or distinctions. This step helped us understand the underlying constructs and their nuances.
- Statistical analysis: We conducted appropriate statistical analyses to explore the relationship between the pleasure-arousal and dominance dimensions. This involved examining correlation coefficients, regression models, or other relevant statistical techniques to identify any patterns or trends.
- Interpretation and validation: We critically analyzed the findings and interpretations, ensuring that the translation from pleasure-arousal to dominance was supported by empirical evidence and consistent with established theories and frameworks.
- Dataset ethics: Our paper also heavily relies on ethical aspects held by the researchers who performed the data gathering for the datasets DEAP and Mementos. Ultimately, the ethical aspects of creating a model using data from different experiments revolve around respecting participant autonomy, ensuring privacy and confidentiality, addressing biases, and promoting transparency and accountability throughout the research process. By upholding these ethical principles, researchers can contribute to the development of AI models that have a positive impact while safeguarding the rights and well-being of individuals involved.

Table 3: Best and Worst R-Squared results achieved by the models during random repeated sampling.

Model	Worst	Best
Mean/Median	-0.023	0.00
Linear Regression	0.038	0.142
Gradient Boosting Regressor	0.150	0.488
Decision Tree Regressor	-0.001	0.180

6 Discussion

Since we were unable to obtain a multiple affect representation dataset that met our specific requirements, we cannot defend the usage of dominance as done by Russell et al. (1977), where it was shown that certain emotions are challenging to capture without the inclusion of dominance as a dimension. However, despite this limitation, our study makes a general contribution by demonstrating a correlation between the pleasure-arousal-dominance (PAD) dimensions and highlighting that pleasure and arousal can account for some of the variance in the dominance field.

Although we acknowledge that a comprehensive understanding of affective states necessitates the inclusion of multiple dimensions, our findings suggest that there exists a significant correlation between the PAD dimensions. This suggests that pleasure and arousal, as captured by the available datasets, can provide valuable insights into the dominance dimension. While it is important to recognize the limitations of this approach, our study contributes to the broader understanding of affective psychology by showcasing the interconnectedness between these dimensions.

7 Conclusions and Future Work

In this study, we addressed several sub-questions that aimed to help in investigating the relationship between pleasure-arousal (PA) and dominance (D) dimensions, as well as explore the availability of datasets with suitable representations for this analysis. Based on our findings, we draw the following conclusions.

First, the process of gathering datasets revealed that many existing datasets predominantly provide two-dimensional or three-dimensional representations. This highlights the need for further research and dataset development that incorporate the multidimensional nature of affective states, including PA and D. By expanding the availability of datasets with richer representations, we can enhance our understanding of the complex interplay between affective dimensions.

Specifically addressing the research question of the dependence between PA and D, our analysis revealed a significant relationship with a results mainly between 20% and 40% with our observations reaching as high as 48.8%. This finding is particularly noteworthy when compared to other relevant studies in the field, as it contributes to the growing body of evidence supporting the interdependence of affective dimensions. However, it is important to note that further research is required to explore the specific mechanisms underlying this dependence and to validate the findings across diverse populations and contexts.

Moving forward, there are several areas for future work that warrant attention. First, improvements in the translation process from PA to D could be explored. While our study provided a valuable initial investigation, refining the translation methodology may lead to more accurate and nuanced representations of affective dimensions. Consideration of alternative translation approaches and the incorporation of additional dimensions could contribute to a more comprehensive understanding of affective states.

Furthermore, the identified dependence between PA and D holds potential implications across various domains. Future research could focus on exploring the practical applications of this relationship, such as in marketing, user experience design, or psychological interventions. Understanding how the interplay between PA and D influences individuals' emotional experiences and behaviors can provide valuable insights for developing tailored strategies and interventions.

In conclusion, this study sheds light on the dependence between pleasure-arousal and dominance dimensions, revealing a significant relationship with implications for understanding affective states. While further research is needed to refine the translation process and investigate the underlying mechanisms, our findings contribute to the growing body of knowledge in affective psychology. We encourage future research to build upon these findings and explore the diverse applications and potential benefits of understanding the dependence between PA and D.

References

[1] Techfunnel Author. What is emotional marketing? https://www.techfunnel.com/martech/what-is-emotional-marketing/, Feb 2022.

[2] Ingrid Bakker, Theo van der Voordt, Peter Vink, and et al. Pleasure, arousal, dominance: Mehrabian and russell revisited. *Current Psychology*, 33(4):405–421, 2014.

[3] Yoann Baveye, Emmanuel Dellandréa, Christel Chamaret, and Liming Chen. Liris-accede: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, 6(1):43–55, 2015.

[4] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[5] Joost Broekens and Willem-Paul Brinkman. Affectbutton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, 71(6):641–667, 2013.

[6] CHERESEARCHERS. Emotions and media. https://historiesofemotion.com/2016/03/04/emotions-and-media/, Mar 2016.

[7] Bernd Dudzik, Hayley Hung, Mark Neerincx, and Joost Broekens. Collecting mementos: A multimodal dataset for context-sensitive modeling of affect and memory processing in responses to videos. *IEEE Transactions on Affective Computing*, 14(2):1249–1266, 2023.

[8] Przemysław Falkowski-Gilski and Tadeus Uhl. Current trends in consumption of multimedia content using online streaming platforms: A user-centric survey. *Computer Science Review*, 37:100268, 2020.

[9] Jui-Lin Hsu, Yu-Lin Zhen, and Ting-Chou Lin. Affective content analysis of music emotion through eeg. *Multimedia Systems*, 24:195–210, 2018.

[10] Investopedia. R-squared. https://www.investopedia.com/terms/r/r-squared.asp. [Online; accessed June 13, 2023].

[11] S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.

[12] Onur Köksoy. Multiresponse robust design: Mean square error (mse) criterion. *Applied Mathematics and Computation*, 175(2):1716–1729, 2006.

[13] A. Mehrabian and J. Russell. *An Approach to Environmental Psychology*. The MIT Press, Cambridge, MA, US, June 1974.

[14] Emily Mower, Angeliki Metallinou, Chi-Chun Lee, Abe Kazemzadeh, Carlos Busso, Sungbok Lee, and

Shrikanth Narayanan. Interpreting ambiguous emotional expressions. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–8, 2009.

[15] National Institute of Standards and Technology. Quantile-quantile (q-q) plots. NIST/SEMATECH e-Handbook of Statistical Methods. [Online]; accessed June 25, 2023.

[16] Srinivas Parthasarathy and Carlos Busso. Jointly predicting arousal, valence and dominance with multi-task learning. In *Interspeech*, volume 2017, pages 1103–1107, 2017.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[18] Fabien Ringeval, Florian Eyben, Eleni Kroupi, Anil Yuce, Jean-Philippe Thiran, Touradj Ebrahimi, Denis Lalanne, and Björn Schuller. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. *Pattern Recognition Letters*, 66:22–30, 2015. Pattern Recognition in Human Computer Interaction.

[19] J. Russell. Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, 36(10):1152–1168, October 1978.

[20] J. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, September 1977.

[21] K. R. Scherer. What are emotions? And how can they be measured. *Social Science Information*, 44(4):695–729, 2005.

[22] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *International Workshop on Audio/Visual Emotion Challenge*, pages 3–10, Amsterdam, The Netherlands, October 2016. Association for Computing Machinery.

[23] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality & Social Psychology*, 54(6):1063–1070, 1988.

[24] M. Xu, L.-T. Chia, and J. Jin. Affective content analysis in comedy and horror videos by audio emotional event detection. pages 4 pp.–, 2005.

[25] Sicheng Zhao, Xingxu Yao, Jufeng Yang, Guoli Jia, Guiguang Ding, Tat-Seng Chua, Björn W. Schuller, and Kurt Keutzer. Affective image content analysis: Two decades review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6729–6751, 2022.