EMONET: AN LSTM-BASED RNN FOR EMOTION RECOGNITION FROM PHYSIOLOGICAL SIGNALS

Alec Braynen

ABSTRACT

There has been much research in using neural networks to recognize emotion based on physiological signals [1-12]. Physiological signals are signals triggered by a person's central and autonomous nervous system. These signals occur over time and hence are a temporal (sequential data) dataset. Recurrent Neural Networks (RNNs) are well suited to temporal data as they are a class of neural network that "remembers" information from a temporal dataset *[1]*. In this paper, the author proposes an RNN architecture called EMONET which attempts to recognize emotion solely from temporal Electrodermal Activity (EDA), Blood Pressure (BP), Heart Beats Per Minute (BPM), and Respiratory data. The dataset used for training, testing and validating this network is the BP4D+ dataset. The network achieves an accuracy of ~ 40% on these datasets, with its best accuracy, 45%, on the diastolic blood pressure modality. These accuracies could potentially be improved with additional data (audio/video) to give the network an opportunity to discover better features with more strongly correlated data.

1. INTRODUCTION

Being able to recognize emotion from physiological and other data, could lead to better medical care, mental health care [2], and improve lives in any area where more accurate emotion recognition could lead to quantitative improvements.

There has already been a substantial amount of research on emotion recognition from facial features and other non-physiological data with impressive results [3][4] [5]. However, the research literature in the area of recognizing emotion purely from physiological data such as EDA, BP, BPM and respiratory data is quite sparse by comparison.

Recognizing emotion from purely physiological data is difficult problem, as potential features could be correlated with multiple emotions i.e. an increase in BPM could mean anger, excitement, fear, or simply movement etc.

Despite this, we set out to create a user-independent neural network architecture that could potentially fit to this problem. Our propesed architecture is called EMONET. It was trained on BP4D+ dataset's physiological data only, and it achieves an accuracy of around 40% on this data.

State of the art networks achieve ~80% on Galvanic Skin Response Data (GSR) using Sequential Floating Forward Selection [6], 90% using a KNN machine learning architecture [7]. In [8], the authors use a CNN and achieve ~ 60% accuracy in emotion recognition of arousal. And in [9], the authors achieve an accuracy of 99% and ~90% on the AMIGOS and DREAMER dataset respectively (which use different physiological signals). Most of these works however aim to recognize two to 4 categories of emotions, while B4PD+ and EMONET features 10 categories of emotions (making it a much more difficult problem).

We review and discuss these architectures, along with their design decisions, and propose potential architecture changes in EMONET (adapted to the B4PD+ dataset) that could potentially improve our predictive performance.

2. RELATED WORKS

As stated previously, there is a big body of research on recognizing emotion from non-physiological data such as faces, speech and voice, body movements etc. Non-physiological data describes data that is under conscious control of the participant. On the other hand, research literature using solely physiological data, which is a more obscure data source, is not as bountiful.

It is difficult to find standardized research on this problem to discuss state of the art performance more empiracally. Despite this, we discuss [6] [7] [8] [9] [10] which all use physiological data, but across

different data sets, with different data modalities, and/or combined with non-physiological data.

In [9] the authors achieve an accuracy of 99% on the AMIGOS dataset and ~90% on the DREAMER dataset. They classify the emotions in a 4-class structure composed of high valence, high arousal, low valence and low arousal. The authors here propose a network composed of a Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures that exploit multi-modal fusion to recognize the 4 emotion classes mentioned above. The authors use two neural network designs, one deep 2D CNN for EEG image data and another deep 1D Convolution to LSTM network for ECG and GSR data.

In [6] the authors were able to achieve an accuracy of 80% accuracy in emotion recognition using GSR Data. This work is closely related to the work of this paper as the authors have designed a network to recognize emotion within the constraints of single modality physiological data. The authors captured their own private data within the 4 categories of amusement and used machine learning and statistical techniques SFFS and KNN to predict emotions within these categories.

In [10] the authors use a 3D CNN to recognize the self reported emotions of subjects. The authors achieve an accuracy of ~80% on this task. The dataset used in this work is the B4PD+ dataset and the perceived emotions they aimed to recognize were relaxed, amused, nervous, pained, embarassed and surprised.

In [12] the authors propose combining facial expression data with physiological on the spontaneous emotion corpus. This proposal adds weight to the claim that emotion recognition from physiological data is an extremely difficult and a subjective problem.

[11] reviews the literature surrounding physiological emotion recognition. It explains the type of physiological data and how it is usually captured, psychological emotion theories, datasets the various methods and architectures used as an attempt to better solve this problem etc.

3. METHOD

EMONET is divided into 3 parts. The first part consists of the convolutional layers. The first part begins with a masking layer, and then three 1D convolutional layers. The convolutional layers have filter sizes of 4, 8 and 16 respectively, kernel sizes of 7, 5 and 3 respectively and activation functions of hyperbolic tangent, hyperbolic tangent and ReLu respectively. After the convolutional layers, there is a 1D max pooling layer with a size of 2.

The second part of EMONET is just a simple LSTM layer with 10 units and exposed sequential weights.

The final part of EMONET is a 128-neuron dense layer that is fully connected to a 10-feature output layer. [See Figure 2]

We decided not to feature dropout layers in this model as it intuitively seemed that the dataset was too small to accommodate them.

4. EXPERIMENTS AND RESULTS

4.1 EXPERIMENTS

We experimented with different activation functions in the convolutional part of the network. Namely, sigmoid, exponential, and variations on the order of the hyperbolic tangent activation functions and the ReLu activation functions. It was discovered that the sigmoid and exponential activation functions in the convolutional part of the network reduced the accuracy of the output by up to 10%.

Additionally, we experimented with increasing the number of neurons in the dense layer and found that there was no measurable improvement in accuracy.

The model was tested on each of the data modalities of the B4PD+ dataset: EDA, BP (Systolic, Diastolic, mmhg), BPM, Respiratory rate and Respiratory Volts. Additionally, the model was also tested on all of the data at once. The results are shown below in Figure 1 & 3.

4.2 RESULTS

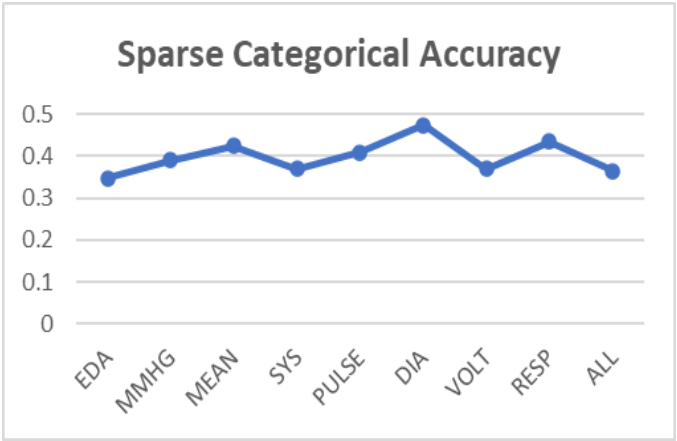

Figure 1: Data Modality and EMONET's Sparse Categorical Accuracy on test data

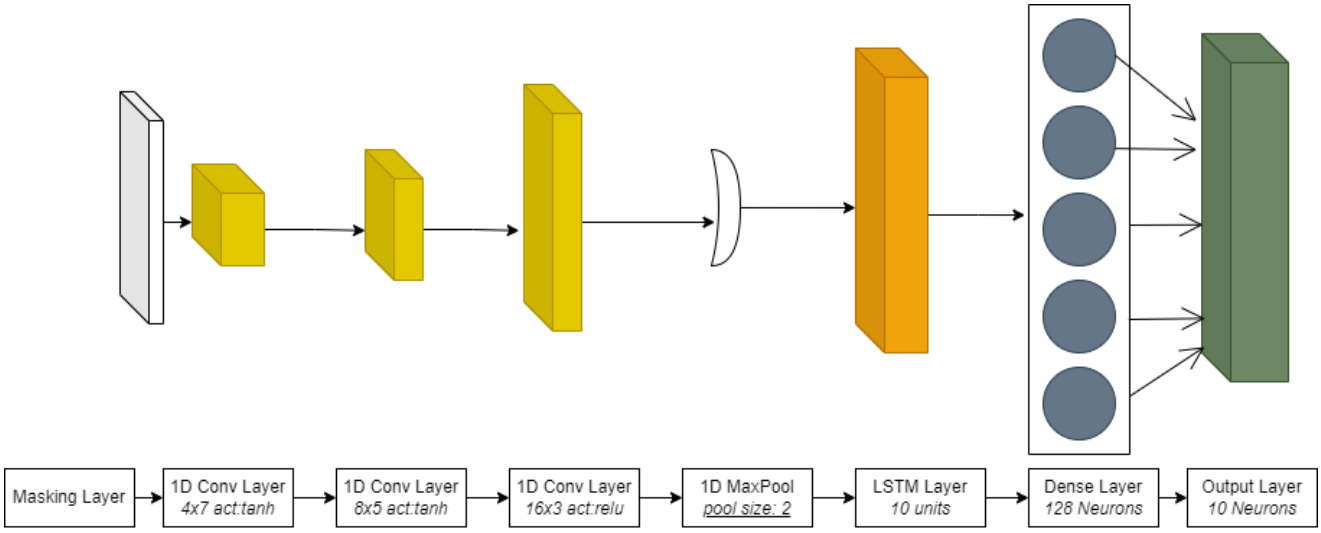

Figure 2: EMONET *Design*

Modality	Sparse Categorical Accuracy	Loss	Macro Precision	Micro Precision	Macro Recall	Micro Recall
EDA	0.3474	1.7189	0.296856848	0.347368421	0.34751462	0.347368421
MMHG	0.39	1.5989	0.382810005	0.39	0.39	0.39
MEAN	0.425	1.6077	0.447393271	0.425	0.425	0.425
SYS	0.37	1.6483	0.31067142	0.37	0.37	0.37
PULSE	0.41	1.5841	0.362665699	0.41	0.41	0.41
DIA	0.475	1.5487	0.475090926	0.475	0.475	0.475
VOLT	0.37	1.6624	0.359610924	0.37	0.37	0.37
RESP	0.435	1.6406	0.339189468	0.435	0.435	0.435
ALL	0.3648	1.6454	0.314290673	0.364779874	0.365153799	0.364779874

Figure 3: Best Results of EMONET on the data modalities of the B4PD+ dataset

As shown in Figures 1 & 3, the model performed best on the Diastolic Blood Pressure Modality (DIA). Diastolic blood pressure is the blood pressure that occurs between heartbeats. Potentially, changes in diastolic blood pressure occur more gradually and stay the same for longer periods of time which allowed our model to better learn the features occurring within the diastolic modality.

It's not unreasonable to presume that if EMONET was adapted to read visual (face) data along with the physiological data, it's accuracy would improve. This intuition comes from the fact that a lot of research in the field typically uses facial emotion recognition along with physiological data to recognize emotions.

5. DISCUSSION

Recognizing emotion from physiological data continues to garner much attention in scientific communities [1-12]. The ability to predict human emotional experiences from physiological data could provide great benefits in medicine, mental health and probably many other industries. However, as a field we perhaps need to discuss the ethics of reading the emotional state of humans from purely autonomous nervous system data. What would this mean for us as a species if we lose the ability disconnect our emotional state from our outward appearance?

5. CONCLUSION

In this paper we presented a CNN-RNN Neural network called EMONET. It was able to achieve an accuracy of 45% on the Diastolic blood pressure modality in the B4PD+ dataset. We postulate that because diastolic blood pressure changes more gradually and predictably, EMONET was able to

achieve its best results. A worthwhile direction for future research would be to modify the network to learn on face data concurrently with physiological data to improve its accuracy in emotion recognition.

6. REFERENCES

[1] G. H. Ilya Sutskever, "Temporal-Kernel Recurrent Neural Networks,," *Neural Networks*, vol. Volume 23, no. Issue 2, pp. Pages 239-243, 2010.

[2] S. L. L. H. W. G. H. Q. a. G. O. Rui Guo, "Pervasive and unobtrusive emotion sensing for human mental health," in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 2013.

[3] S. K. A. K. P. S. M. Z. Neha Jain, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognition Letters*, vol. Volume 115, pp. Pages 101-106, 2018.

[4] V. M. K. K. R. M. a. C. P. amira Ebrahimi Kahou, "Recurrent Neural Networks for Emotion Recognition in Video," in *In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, New York, NY, USA, 467–474, 2015.

[5] C. P. X. B. P. F. Ç. G. R. M. P. V. A. C. Y. B. R. C. F. M. M. S. J. P.-L. C. Y. D. N. Samira Ebrahimi Kahou, "Combining modality specific deep neural networks for emotion recognition in video," in *In Proceedings of the 15th ACM on International conference on multimodal interaction*, New York, NY, USA, 543–550, 2013.

[6] S. L. L. H. W. G. H. Q. a. G. O. Rui Guo, ""Pervasive and unobtrusive emotion sensing for human mental health," in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 2013.

[7] J. K. a. E. A. . Wagner, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification," in *IEEE International Conference on Multimedia and Expo*, 2005.

[8] G. L. a. J. Tongshuai Song, " Emotion Recognition Based on Physiological Signals Using Convolution Neural Networks," in *In Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, 2020.

[9] M. N. M. U. A. S. G. K. a. A. N. P. 2. Dar, "CNN and LSTM-Based Emotion Charting Using Physiological Signals," 2020.

[10] S. C. a. L. Y. S. Hinduja, "Recognizing Perceived Emotions from Facial Expressions," in *15th IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

[11] L. J. X. M. Y. Z. L. Z. L. D. L. X. X. a. X. Y. Shu, "A Review of Emotion Recognition Using Physiological Signals," *Sensors* , vol. 18, no. 7, 2018.

[12] S.-L. &. H. J. &. K. W. &. P. K. &. Y. B. N. &. K. S. &. S. F. &. P. A. G. Lee, "Fine-grained emotion recognition: fusion of physiological signals and facial expressions on spontaneous emotion corpus," *International Journal of Ad Hoc and Ubiquitous Computing*, 2020.