Long Short-Term Memory Network Predicting Negative Affect using Bilateral Amygdala Activation during Naturalistic Viewing

Kathryn Adams
Department of Psychology
Georgetown University
Washington, D.C., U.S.A
ka741@georgetown.edu

I. *ABSTRACT*

Naturalistic viewing data, often captured by collecting data while participants watch movies, increases the ecologically validity of emotion perception research. Taking further advantage of the timeseries nature of naturalistic viewing data we utilized a LSTM network. Participants underwent fMRI scanning while watching a full-length movie. We hypothesized that amygdala response during movie watching could predict participants' sensitivity to negative affect above chance. Our model was able to label participants as having low, medium, or high sensitivity to negative affect. This provides preliminary evidence that a LSTM network is a critical tool in timeseries analysis of emotion perception data.

II. INTRODUCTION

Naturalistic viewing, or movie watching, during functional magnetic resonance imaging (fMRI) presents a unique opportunity to understand dynamic changes in emotion perception. [1] Prior research was often limited to capturing responses to stagnant stimuli such as images of emotional expressions. [2] Naturalistic viewing utilizes movies in order to capture more ecologically valid conditions where emotion perception is captured across the entire timeseries rather than at isolated timepoints. [1] Utilizing fMRI during naturalistic viewing is especially relevant in clinical populations to understand dysfunctional neural responding during emotion perception.

Capturing dynamic responding through naturalistic viewing is still limited by current analysis techniques. The most popular of which, Multivariate Pattern Analysis (MVPA), extracts individual timepoints to detect patterns of responding within the brain. [1] Thus, brain responses are analyzed in an individual fashion rather than continuously across the entire timeseries.

In the current study, we therefore utilize a Long Short-Term Memory (LSTM) network to investigate emotion perception. [3] The LSTM network leverages the timeseries aspect of naturalistic viewing data allowing us to predict participants negative affect ratings more accurately. Participants underwent neuroimaging while watching a full-length movie. We hypothesized that the LSTM network would be able to predict above chance, participants' level of negative affect, a measure linked to high physiological arousal during negative emotional states (e.g., anger, fear, sadness). [4]

III. RELATED WORK

LSTM: The Long Short-Term Memory (LSTM) network utilizes information from prior inputs to learn long term relationships between inputs. [3] A LSTM network learns these relationships by adding and removing information from cells utilizing gates. An LSTM has three gates, an input gate, output gate, and forget gate. The input gate controls what information will come into the cell using a tanh function. The output gate controls what information is used as an input to the following cell using a sigmoid function. Finally, the forget gate utilizes a sigmoid function to control what information is forgotten as new information comes into the cell.

IV. DATASET

Participants: Eighty-six healthy adult participants (Age: *M* = 26.87 , *S.D.* = 10.03, Gender: %Male = 51.12, Ethnicity: %White = 50, %Asian = 27.91, %Black or African American = 11.63, %Other = 10.47) were recruited for a naturalistic viewing study. [1] Exclusion criteria included left-handedness, a history of neurological/psychiatric illness, hearing or vision impairments, medication use, contradictions to MRI scanning and being a non-native English speaker.

Naturalistic Viewing: Participants were pseudo randomly assigned to one of 10 full length movies they had not seen before (500 Days of Summer, 12 Years a Slave, Citizen Four, The Usual Suspects, Pulp Fiction, The Shawshank Redemption, The Prestige, Back to the Future, Split, or Little Miss Sunshine). Movies were stopped in 40-50-minute intervals or when a break was requested by participants. Participants completed 2-6 BOLD-fMRI runs depending on stop requests and movie length.

Image Acquisition: Functional and structural images were collected using a 1.5 T Siemens MAGNETOM Avanto with a 32-channel head coil. A multiband sequence (TR = 1 s, TE = 54.8 ms, flip angle of 75°, 40 interleaved slices, resolution = 3.2 mm isotropic), with 4x multiband factor and no in-plane acceleration was used to collect Function T2*-weighted images. All functional volumes were registered to the subject's anatomical grid collected at the start of each session via the 10-minute high-resolution T1-weighted MP-RAGE anatomical scan (TR = 2.73 s, TE = 3.57 ms, 176 sagittal slices, resolution = 1.0 mm).

V. METHODS

Image Analysis: Functional images were analyzed using the AFNI software suite. [5] Feature extraction was performed using the Montreal Neurological Institute (MNI) atlas left and right amygdala masks (Figure 1). [6] Estimated left and right amygdala mean activation values were calculated for each 1 second timepoint (Figure 2). Final values were then standardized for left and right amygdala separately.

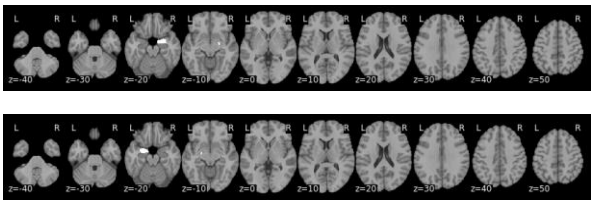

Figure 1. MNI left (bottom) and right (top) amygdala mask in horizontal brain slices. Amygdala location is shown in white.

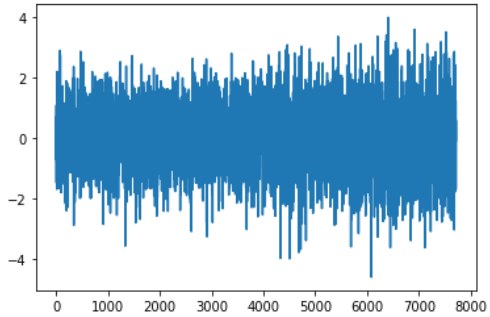

Figure 2. Timeseries of mean left amygdala response for a single participant. Standardized functional amygdala activation is on the y axis and time in seconds is on the x axis.

Data preprocessing: Participant timeseries data initially varied in length from 5470-8882 seconds (Figure 3). To have input data of the same size and reduce the length of the input needed to make a prediction each participant's timeseries was split into 60 second (1 minute) segments. This resulted in an input size of *7826 * 60 * 2* equating to 7826 total timeseries, each timeseries 60 seconds in length, with 2 values, one for left, and one for right amygdala response. Amygdala response values were standardized and therefore ranged from -4 to 4.

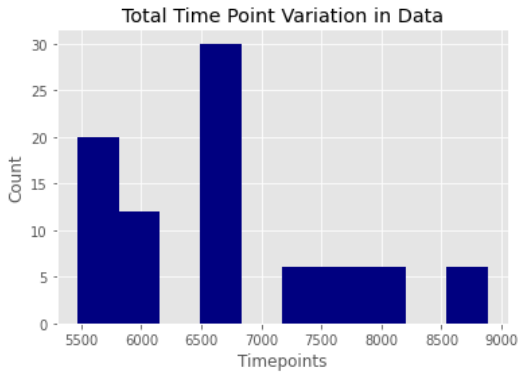

Figure 3. Variation in timeseries length across participants. Length depended on the movie to which participants were assigned.

Next, to predict negative affect, the continuous negative affect variable needed to be transformed into a grouping variable. Participants were separated into one of three groups (High, Medium, Low) based on Negative Affect scores. A Lowess plot was used to identify shifts in the relationship between average amygdala response across the timeseries and negative affect scores (Figure 4). Visual inspection revealed three separate linear relationships thus, low was categorized as a score less than 45 (n = 18), medium as 43- 57 (n = 37), and high as greater than 57 (n = 31). Group membership (low: 0, medium: 1, high: 2) was used as the predicted outcome in our LSTM network.

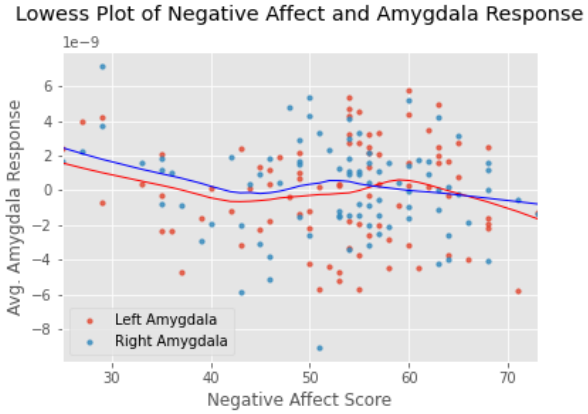

Figure 4. Lowess plot depicting the relationship between left and right amygdala response (averaged across the timeseries).

Model construction: A LSTM was the optimal network being that it can leverage the sequence of responding within a timeseries rather than individual inputs.

- LSTM input: The LSTM network took an input of size *7826 * 60 * 2*. To prevent overfitting and train the model on the most predictive features a dropout rate of .2 was used on the inputs to the model. [7]
- LSTM output: The LSTM had an output space of 128 units. An additional recurrent dropout rate of .15 was applied to the outputs of the LSTM network. [7]
- Dense layer: A dense layer was used to capture the output from the LSTM network and via a SoftMax function calculate a probably rating.

Output & Visualization: Predictions were made for each timeseries resulting in an output of size *7826 * 3*. These ratings quantified the probability that a timeseries belonged to each of the three negative affect groups (low, medium, high). The complete model is depicted in figure 5.

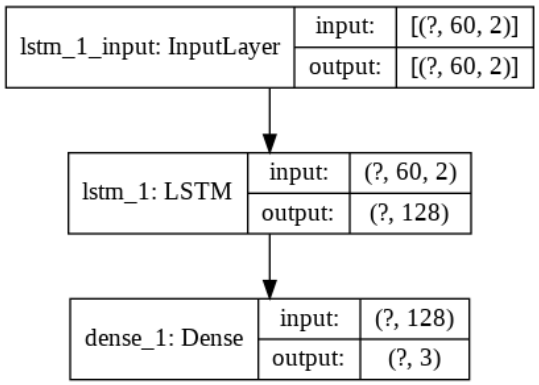

Figure 5. The complete LSTM network model.

VI. RESULTS

To test the hypothesis that the LSTM network was more accurate than chance a bootstrapped resampling of testing and training data was used to calculate model accuracy with 25 replications. Chance accuracy was set at ~.33 because of the presence of three possible group labels. The results of the bootstrapped test are depicted in figure 6.

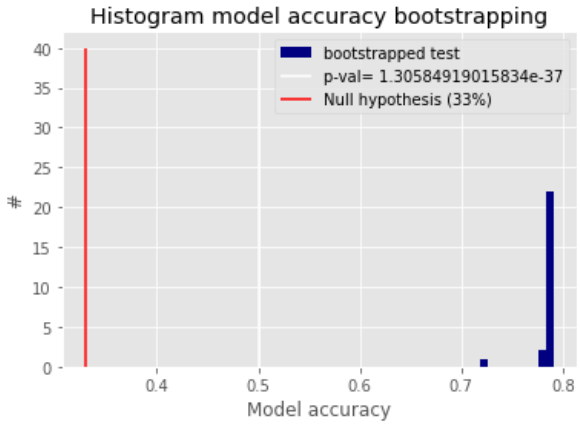

Figure 6. Results of a bootstrapped test for model accuracy.

VII. DISCUSSION

The analysis of naturalistic viewing data via a LSTM network leverages the dynamic nature of emotion perception by capturing not only individual emotional states but also the sequence and transition from one emotion to the next. Using this approach our findings provide evidence that a LSTM network can predict participant's sensitivity to negative affect. Specifically, our results suggest that 60 second increments of bilateral amygdala activation during movie watching contains meaningful information about a participant's sensitivity to fear, anger, and sadness.

These results support prior research on the amygdala as a key brain structure involved in emotion perception. [2] This literature, however, leaves a few critical questions unanswered due to its limited ecological validity. No prior study on naturalistic viewing and emotion perception has utilized a LSTM network.

VIII. CONCLUSION

In the current study, we therefore utilize a LSTM network to capture how the brain transitions between emotional states. We were able to predict participants' sensitivity to negative affect based on 60 seconds of bilateral amygdala activation during a movie. This finding highlights the potential for LSTM networks to be utilized in functional timeseries analyses especially when investigating emotion perception.

Limitations of the current study include the lack of labels for each 60 second segment. It is possible that participants were not viewing anything that would provoke a strong emotional response. Additionally, only using the amygdala limits the conclusions that can be drawn about larger brain networks that function together to process emotions.

Future research should include other brain regions as inputs to the model. Future research should also investigate the impact of different actors in a movie displaying the same emotion. Thus, suggesting that the brain is sensitive to not just emotional states but also the actor portraying that state.

We conclude that the amygdala is a key region in emotion perception and varies in its sensitivity to negative affect (fear, anger, sadness) across individuals. We also highlight the importance of the LSTM network and its ability to leverage the dynamic nature of fMRI naturalistic viewing data.

IX. REFERENCES

[1] Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. I. (2020). A naturalistic neuroimaging database for understanding the brain using ecological stimuli. Scientific Data, 7(1), 347.

[2] Kristen A. Lindquist, Ajay B. Satpute, Tor D. Wager, Jochen Weber, Lisa Feldman Barrett, The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature, Cerebral Cortex, Volume 26, Issue 5, May 2016, Pages 1910–1922.

[3] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[4] Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurology. 2013 Mar 12;80(11 Suppl 3):S2-6. doi: 10.1212/WNL.0b013e3182872e5f. PMID: 23479538; PMCID: PMC3662335.

[5] Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173 (1996).

[6] Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A probabilistic atlas of the human brain: Theory and rationale for its development. The International Consortium for Brain Mapping (ICBM) Neuroimage. 1995;2:89–101.

[7] Gal, Y., & Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. Advances in neural information processing systems, 29, 1019-1027.