EMOTION RECOGNITION USING PHYSIOLOGICAL SIGNALS

Xiaowen Li

ABSTRACT

Emotion recognition has gained great attention not only in the medical field but also in education, defense, and commerce. Recognizing emotion can be done by using audio signals, body actions, facial expressions, and many other data that can be collected from the human body. But more researchers are focused on physiological signals the fact that these are spontaneous and not controllable by human beings.

This paper is based the BP4D+ database [1] which is a collection of physiological signals that are naturally temporal. A recurrent neural network(RNN) is designed to train this dataset and a test accuracy around 40% is desired for all data types consisted in this dataset. The experiment is aimed to generate one model from training from each data type and then save them for testing and accuracy analysis. Following metrics are presented: Test accuracy, Recall (both macro and micro), Precision(both macro and micro), F1 score (both macro and micro) and also the confusion matrix.

The experimental analysis shows that with the RNN we designed that will be covered in detailed in the methods section, among all models that one for each data type, the highest test accuracy is 0.4449999928474426 with LA Systolic Blood Pressure data type. And for all data types combined together, the model achieved a 0.4113207459449768 test accuracy.

1. INTRODUCTION

Affect is pervasive in people's daily life. Affect can reflect a person's current situation Physiological and psychological states also have an important impact on people's cognition, communication and decision-making. Emotional changes are usually produced under the stimulation of the external environment and are accompanied by changes in individual representations and psychological reactions, so they can be Measurements, and simulations are carried out using the scientific method. Physiological signal is one of the scientific measurement for emotion recognition and due to it's objectiveness that it's not controllable by individual, it has gain a great attention in an interdisciplinary field of study that spans multiple disciplines including computer science, psychology, and cognitive science.

In this paper, we are going to generate a recurrent neural network with the BP4D+[1]. Specifically, the neural network will be able to handle the temporal data from the database and we are going to evaluate the performance and fine-tune it for the best model based on different data types and evaluation metrics that are introduced in section [4.1].

2. RELATED WORKS

In 1986, one of the founders of artificial intelligence, the winner of the Turing Award – Professor Minsky first proposed the concept of the ability that computers can recognize emotions in one of his book *The Society of Mind* [2].

Moving to recent 10 years, there have been many key articles published in these fields working toward emotion recognition. In the year 2013, Wioleta published a paper[3] at the IEEE international conference that present an actual problem of using physiological signals for emotion recognition. This paper introduced different types of physiological signals and sensors, and also the models of emotions and how emotions can be elicited. This paper gives a survey of all works done before using signals for emotion recognition and again states the possibility for future research.

After that in the year 2017, Ragot and his fellows published a paper[4] that detailed talked about how the physiological signals for emotion recognition can be collected by specific laboratory sensors. Among many wearable devices now in the industries, they analyze them focusing on their emotion recognition capabilities. They specifically focused on the machine learning training result on the Biopac MP150 (from laboratory sensor) and Empatica E4 (from wearable sensor) and concluded that they are similar in terms of accuracy.

Then in 2018, Albraikan from the University of Ottawa [5] attempted to further increase the accuracy of emotion recognition by physiological signals by using peripheral physiological signals. They proposed a hybrid sensor fusion approach based on a stacking model that takes data to be jointly embedded in a user-independent model. They achieved an overall 65.6% accuracy on the E4 dataset.

Some researches focused on one type of signal for emotional recognition. One is ECC; In Xin's 2010 paper [6], they first introduced a novel feature named ECC that is proposed via feature extraction of the Hilbert energy spectrum. This signal outperforms the traditional other signals and they achieved the highest 83.57% emotion recognition rate in 7 classes. For EEG [7], using only 2 channels of the frontal EEG signals, their experiment used a GBDT classifier to produce a mean classification accuracy at 75.18%.

There are also many different data bases with multidata types for emotion analysis. DEAP [8], short stand for *Database for Emotion Analysis Using Physiological Signals* contains 32- channel EEG together with 12 other peripheral physiological signals. It is one of the commonly used data base in this field. Garcia [9] and his fellows made the highest mean accuracy 89.45% on this data in the year 2016 using Gaussian process latent variable models. Another attempt [10], by importing the extracted features in to a linear support vector machine, it achieved the hightest accuracy rate at 85.2%.

The MAHNOB database, which is a database constructed for multimedia labels in new areas, was collected from participants' physiological signals to the multimedia content. The state of art recognition accuracy was 73.8% from a 2017 paper[11] that was done by using a multi-layer LSTM RNN model with a temporall and band attention.

The BioVid Emo database, which consists of SC, ECG and EMG type of physiological signals. Had a highest classification accuracy of 79.51% done by this research paper [12].

Also the database we are using in this paper, BP4D+ is a physiological signal database that consists of 8 data types. In a 2019 paper [13], Fabiano and Canavan, our professor at the University of South Florida proposed a new method for emotion recognition using fused physiological signals that takes the variance of those signals to boost the effectiveness of the part of the signals that contribute to the accuracy and down weight the rest that does not.

3. METHOD

We finalized our recurrent neural network with the following architecture: First starting with an embedding layer to get the input from the database we read in from files. Details about the database can be found in the Experiment section. Followed by a 64 filter kernel size 3 1D convolutional layer with ReLu activation function. After that, a size 2 max-pooling is added. Then we apply the bidirectional LSTM layer1 which is our key concept that makes this an RNN. We applied this LSTM twice with the first number of units to be 64 and the second to be 128. After LSTM is done, we have one 64 neurons densely connected layer and finally a softmax activation output layer with 10 possibly classes as described in section 4.1.

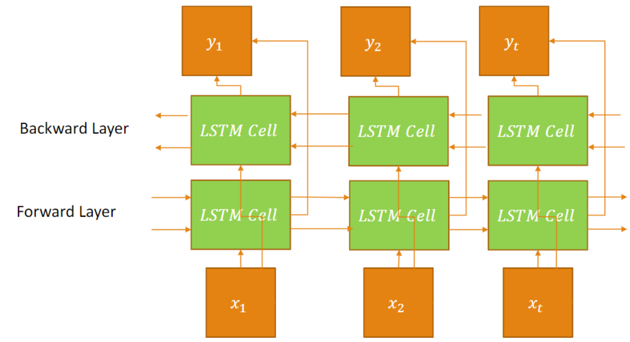

Fig. 1: Diagram explaining bidirectional LSTM[14]

Specifically, For bidirectional LSTM we take the tanh activation function and sigmoid function for recurrent activation. We are taking the dimension of the output space in an increasing order layer by layer, from 64 double to 128.

The choice of using Bidirectional LSTM is pretty straightforward. The database we are using is temporal, so an RNN will be needed for a good fit. Both Unidirectional and bidirectional LSTM falls into this category. Bidirectional LSTM takes both right and left context from one prediction so unless some specific dataset only works well for one direction timeline, Bidirectional LSTM tends to give a better accuracy but also requires more training. Applying one convolutional layer before LSTM is a common strategy for people doing research in the NLP area, the convolutional layer plays the role of a feature extractor before we run into the complicated double layer of Bidirectional LSTM.

4. EXPERIMENTS AND RESULTS

4.1. Datasets

The data - BP4D+ which is collected by researchers at Binghamton University is a dataset that consists of physiological signals that are naturally temporal. It is an extension of the BP4D dataset which is a 3D+time database. It contains multimodal datasets including synchronized 3D, 2D, thermal, physiological data sequences, and meta-data (facial features and FACS codes). And are collected from 58 males and 82 females. The data falls into 10 categories - (1) Happy; (2) Sad; (3) Surprise; (4) Pain; (5) Disgust; (6) Afraid; (7) Startled; (8) Skeptical; (9) Embarrassment; (10) Fear.

The dataset has 8 data types: (EDA) electrodermal activity, (mmHg) blood pressure, (mean) LA mean blood pressure, (sys) LA systolic blood pressure, (pulse) pulse rate, (DIA) blood pressure diastolic, (volt) respiration rate in voltage and (resp) respiration rate in BPM.

Data came in a zipped file and split into 3 folders for training, testing, and validation purpose. The data text file has the title that tells the subject, class, and then data type. Data inside the text files contain one float number for each line with a maximum of 1000 lines but is not guaranteed.

4.2. Implementation and Required Packages

We run our experiments on TensorFlow 2 with python version 3.9. Keras API is used in experiments.

4.3. Experiment Methods

Our experiment is aimed to build an appropriate neural network structure that at least meets the average test accuracy on BP4D+ data [1] which is about 40%. We also give it the ability to save the best model based on data type and then do an accuracy analysis from the saved model. The following metrics are taken for the evaluation:

- Accuracy
- Recall (macro and micro)
- Precision (macro and micro)
- F1 score (macro and micro)
- Confusion Matrix

4.4. Experiment Observations

We run the training with the architecture specified in Method section [3]. And we finalized with the following values for hyper-parameters:

The number of batches to be 16 and maximum number of epochs to be 20. Also a softmax activation function is used for the output layer. We are getting the following metrics for each data type:

sys:	
Test accuracy:	0.4449999928474426
Macro precision:	0.36815389538090687
Micro precision:	0.445
Macro recall:	0.4388888888888889
Micro recall:	0.445
Macro F1:	0.4004214825090421
Micro F1:	0.44499995000000564

EDA:	
Test accuracy:	0.38427671790122986
Macro precision:	0.377792532218104
Micro precision:	0.38427672955974845
Macro recall:	0.3823043547488258
Micro recall:	0.38427672955974845
Macro F1:	0.38003500273012175
Micro F1:	0.38427667955975503

Table 1: Best models of sys(Left) EDA(Right)

mmHg:	
Test accuracy:	0.41194969415664673
Macro precision:	0.41697414546199035
Micro precision:	0.4119496855345912
Macro recall:	0.4158494878327097
Micro recall:	0.4119496855345912
Macro F1:	0.41641100727011177
Micro F1:	0.41194963553459735

Mean:	
Test accuracy:	0.4000000059604645
Macro precision:	0.4182618153682621
Micro precision:	0.4
Macro recall:	0.39722222222222225
Micro recall:	0.4
Macro F1:	0.40747055672274696
Micro F1:	0.3999999500000063

Table 2: Best models of mmHg(Left) Mean(Right)

Volt:	
Test accuracy:	0.38999998569488525
Macro precision:	0.280019423823104
Micro precision:	0.39
Macro recall:	0.36111111111111105
Micro recall:	0.39
Macro F1:	0.31543688527636715
Micro F1:	0.3899999500000065

Resp:	
Test accuracy:	0.4099999964237213
Macro precision:	0.3337197684910103
Micro precision:	0.41
Macro recall:	0.3916666666666667
Micro recall:	0.41
Macro F1:	0.3603786477708302
Micro F1:	0.40999995000000605

Table 3: Best models of Volt(Left) Resp(Right)

Pulse:	
Test accuracy:	0.375
Macro precision:	0.3229249178717264
Micro precision:	0.375
Macro recall:	0.3805555555555556
Micro recall:	0.375
Macro F1:	0.3493795739794089
Micro F1:	0.3749999500000067

DIA:	
Test accuracy:	0.3566037714481354
Macro precision:	0.33467727087525945
Micro precision:	0.35660377358490564
Macro recall:	0.36397120938707994
Micro recall:	0.35660377358490564
Macro F1:	0.3487100510219182
Micro F1:	0.3566037235849127

Table 4: best models of Pulse(Left) DIA(Right)

ALL:	
Test accuracy:	0.4113207459449768
Macro precision:	0.37553764091703484
Micro precision:	0.41132075471698115
Macro recall:	0.4034455191996391
Micro recall:	0.41132075471698115
Macro F1:	0.3889916158046533
Micro F1:	0.4113207047169873

Table 5: Metrics for best models of all data types combined together

Choice of number of epochs When a complete dataset passes through the neural network and returns the result, this one process is counted as one epoch. A general observation is that neural networks will get better accuracy when the number of epochs is increasing. But it is not always the larger the better. First of all, more epochs mean longer run time. But even if we have enough time to run the training process, the accuracy will still reach a point where it converges and no longer improves. This is the case for this experiment. As shown in figure [3], the accuracy and also the loss are converging to a certain limit when the number of epochs goes beyond 10. We also see that the accuracy of the validation data keeps going up and down from the converging limit. Therefore we picked the maximum number of epochs to be 20 and use the checkpoint to find the best accuracy over all accuracy produced at each time one epoch finished.

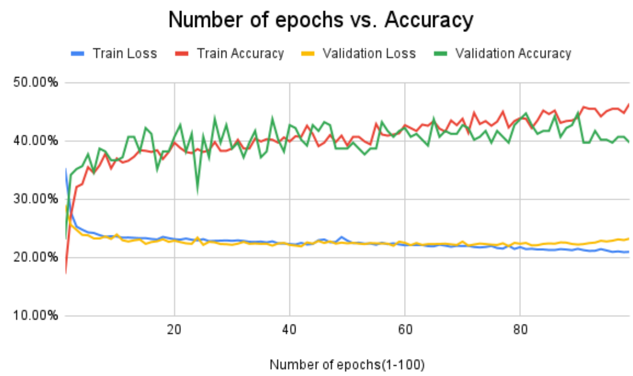

Fig. 2: Accuracy/Loss vs number of epochs

Number of Batches vs. Number of epochs From the previous paragraph, we learned that increasing the number of epochs will be limited due to the increase in run time. A way to balance this is to split data into batches. More batches get fewer data elements in one epoch while running. But also may lead to a worse overall accuracy for the training result. We examined our model with 8, 16, and 32 batches and found that 16 is a good trade-off between efficiency and accuracy.

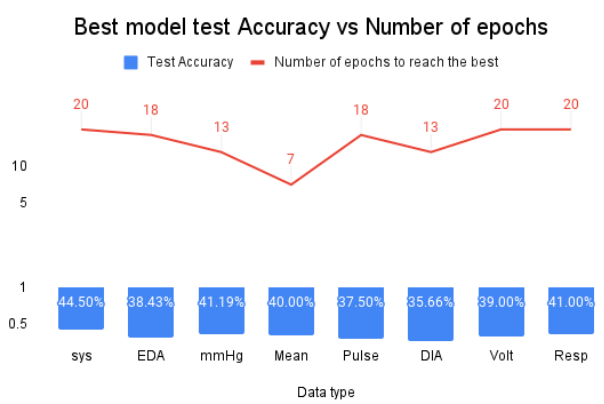

Fig. 3: Best model test Accuracy vs Number of epochs to reach the best for every data type

As a result, we evaluate the best model based on the test accuracy. The best model we found was with (sys) LA systolic blood pressure data type. It achieves the highest test accuracy 0.4449999928474426 with 20 epochs.

5. CONCLUSION

In this paper we introduced the following neural network to train the BP4D+[1] database: 1D convolutional layer followed with two bidirectional LSTM layer. With 16 batches and 20 epochs, we achieved our best model with (sys) LA systolic blood pressure data type with a test accuracy at 0.4449999928474426. We also got all data types training results with an average test accuracy of % which meets the baseline accuracy for this database.

6. REFERENCES

[1] Binghamton University, "3d facial expression database," .

[2] Marvin Lee Minsky, *The Society of Mind*, Simon amp; Schuster Paperbacks, 2007.

[3] Szwoch Wioleta, "Using physiological signals for emotion recognition," in *2013 6th International Conference on Human System Interactions (HSI)*, 2013, pp. 556–561.

[4] Martin Ragot, Nicolas Martin, Sonia Em, Nico Pallamin, and Jean-Marc Diverrez, "Emotion recognition using physiological signals: laboratory vs. wearable sensors," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 15–22.

[5] Amani Albraikan, Diana P. Tobón, and Abdulmotaleb El Saddik, "Toward user-independent emotion recognition using physiological signals," *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8402–8412, 2019.

[6] Li Xin and Li Xiang, "Novel hilbert energy spectrum based features for speech emotion recognition," in *2010 WASE International Conference on Information Engineering*, 2010, vol. 1, pp. 189–193.

[7] Shiyi Wu, Xiangmin Xu, Lin Shu, and Bin Hu, "Estimation of valence of emotion using two frontal eeg channels," in *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2017, pp. 1127–1130.

[8] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[9] Hernán F García, Mauricio A Álvarez, and Álvaro Á Orozco, "Gaussian process dynamical models for multimodal affect recognition," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 850–853.

[10] Wei Liu, Wei-Long Zheng, and Bao-Liang Lu, "Emotion recognition using multimodal deep learning," in *International conference on neural information processing*. Springer, 2016, pp. 521–529.

[11] Jiamin Liu, Yuanqi Su, and Yuehu Liu, "Multi-modal emotion recognition with temporal-band attention based on lstm-rnn," in *Pacific rim conference on multimedia*. Springer, 2017, pp. 194–204.

[12] Zi Cheng, Lin Shu, Jinyan Xie, and CL Philip Chen, "A novel ecg-based real-time detection method of negative emotions in wearable applications," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. IEEE, 2017, pp. 296–301.

[13] Diego Fabiano and Shaun Canavan, "Emotion recognition using fused physiological signals," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 42–48.

[14] S. Canavan, "Lecture10_timebasedmodelspart2.pdf," University of South Florida, 2022, pp. 1–10.