Pain Identification using SVM and Random Forest

Heamnath Balasubramanian, Lalithabhinaya Mallu, Mathan Kumar Venkateswaran, Padmavathi Siddi, Sravya Kodali

Affective Computing – CAP 5627

ABSTRACT

Identifying emotion based on the physiological signals has become a very important topic for research and applied in many areas like health care, armed forces and many other areas. This paper present studies on systems build for identifying pain from other emotions using machine-learning algorithms. Here we build systems which analyses the data from two datasets having 8 physiological data. We have classified the data as pain or no pain with the help of Random Forest and Support Vector Machine classifiers. The results are then compared with Correlation prediction to get information about the systems and analyze the machine learning algorithm. The role of this XAI is discussed. These can be used as a guide for further studies in the increasingly important area of Military[3][4].

Keywords*: XAI; Correlation; SVM; Random Forest; physiological signals;*

1. INTRODUCTION

Injured soldiers experience pain must be addressed before they return to their duty or civilian life. So, US armed forces have taken many initiatives to address these issues and ways to manage soldiers' pain and treat this pain as soon as possible. Emotion recognition method [5] is broadly classified into two categories based on the human physical signals and other is using the internal signals. The first one, using the physical signals is based on the facial, speech posture, gesture etc., is easy to collect the data. However, it is not more reliable with this data since it is easy for people to control their physical signals. Like it is easy for a person to control his negative emotions and can just smile. The second one is based on the internal signals called physiological signals such as temperature, electrocardiogram, electroencephalogram, galvanic skin response, electromyogram, respiration and many others[1]. The later features change when people are exposed to certain situations accordingly. Since these signals are in response to ANS and CNS of a human body, which changes based on emotion as per Cannon's theory [2]. So, the benefit with the later is all these signals are involuntarily activated and there is no control on these emotions.

2. METHOD

Our paper discuss on system for automatic detection of pain as if one wounded or experience any form of pain. This system classifies pain based on physiological signals which can be collected from wearable devices and then train and test this data on our system. We have used python to implement the algorithm which is discussed more in detail in the later section.

2.1. DATASET

The dataset consists of 140 subjects of which 82 were female and 58 were male. The dataset is divided into dataset1 and dataset2, each of them has 8 physiological data (Left atrial pressure (LAP) systolic blood pressure, Mean Arterial Pressure, Electrodermal Activity, Blood pressure Diastolic, Pulse rate, Respiration rate, Respiratory Voltage, Blood pressure) which are recorded under different situations/tasks. Dataset1 has two emotions(tasks) in which one is pain and the other is no pain. Dataset2 has 10 emotions(tasks) in which one is pain and all the rest are no pain, thus dataset2 is highly imbalanced.

2.1. PROCEDURE

The mean, maximum, minimum, variance, and entropy for each physiological data was calculated and thus the calculated 40 values for each task along with the subject_id and Label(pain/no pain) is stored as a row in python dataframe. Any extra physiological data for certain tasks can be disregarded. Once the data is fully processed, we will have 280 rows(instance) for dataset1 and 1400 rows (instance) for dataset2.

Data splitting was done manually, in this the data frame is split as 80% for training and 20% for a test, the data in the test set remained the same throughout the process explained below.

The SVM and Random Forest classifiers are used for classifying data as pain or no-pain. Random Forest algorithm is commonly referred to as an ensemble learning mode which works by constructing a multitude of decision trees and it has larger training time. The SVM(Support Vector Machines) algorithm is a supervised learning model which works for two class problems, the algorithm works by plotting the examples in space and creating a vector that separates two classes in the given space such that the examples lying on the same side of the support vector belongs to one class.

The data in the training set and testing set are used for training and testing the classifier. Once the accuracy for each classifier for both the data set is calculated, the correlation for each data in the testing dataset with all the data in the training set is then calculated. In the process of calculating the correlation coefficient of the data, we take the average of all the mean, variance, entropy, minimum and maximum of all task, so this results in an data having 5 features for a particular class/label. The prediction for a data in the testing set is calculated based on the label of highest correlation training data, similarly all testing prediction is calculated. The accuracy for the correlation prediction is calculated by considering the classifiers(SVM and RF) prediction as ground truth at each instance as 'Classifier vs Correlation Accuracy score'. The misclassified data from the previous comparison are found and an accuracy score for those with actual ground truth is calculated as 'Misclassified Correlation vs ground truth Accuracy score'.

.

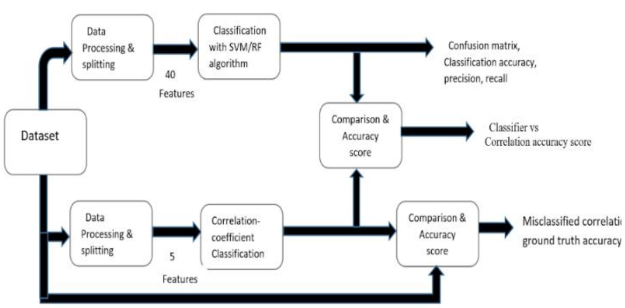

Figure 1: Procedure flow

The examples in the test set should remain the same throughout the entire process.

3. RESULTS

Table 1 and table 2 displays accuracies of RF and SVM classification algorithms on both Dataset1 and Dataset2. Also, they show XAI accuracy and accuracy of Misclassified correlation with the ground truth.

RF	Accuracy score	Classifier vs Correlation Accuracy score	Misclassified Correlation vs ground truth Accuracy score
Dataset1	87.50%	57.14%	20.83%
Dataset2	91.42%	95.71%	33.33%

Table 1: Accuracy of RF and XAI.

SVM	Accuracy score	Classifier vs Correlation Accuracy score	Misclassified Correlation vs ground truth Accuracy score
Dataset1	83.92%	58.92%	8.69%
Dataset2	90.71%	98.57%	25%

Table 2: Accuracy of SVM and XAI.

Table 3,4,5 and 6 shows the Confusion matrices of RF and SVM on both datasets.

RF on Dataset1	Pain	No Pain
Pain	24	4
No Pain	3	25

Table 3: Confusion Matrix of RF on Dataset1

RF on Dataset2	Pain	No Pain
Pain	8	20
No Pain	4	248

Table 4: Confusion Matrix of RF on Dataset2

SVM on Dataset1	Pain	No Pain
Pain	21	7
No Pain	2	26

Table 5: Confusion Matrix of SVM on Dataset1

SVM on Dataset2	Pain	No Pain
Pain	3	25
No Pain	1	251

Table 6: Confusion Matrix of SVM on Dataset2

4. DISCUSSION AND CONCLUSION

Initially, the dataset is processed to have 40 features for each subjects under different labels, the data is then split in ratio 8:2 for training and testing and the subjects in both the sets remained same throughout the process. The result of the classification using SVM and Random Forest shows that Random Forest has better accuracy on both the datasets (87.50% & 91.42%).

The Misclassified vs ground truth accuracy scores shows the amount of test data which correlation predicted right and the classifier predicted wrong. In other terms, we can say it shows that how close the feature correlation is well aligned and utilized with classifier's algorithm, if score is low then classifier has better correlation with correlation prediction or features or vice versa. This accuracy score of correlation prediction corresponding to SVM classification prediction has low percentage of miscorrelation (8.69% & 25%) on both dataset1 and dataset2. We can also see that out of the conflicting values, Random Forest went wrong more than the SVM when compared with the ground truth. Even though Random Forest has better overall accuracy than SVM by little margin, which can be reason of imbalanced data or even as negligible percentage, Correlation accuracy score presents SVM as classifier with better correlation with dataset/feature which might improve with increased test data.

We can see from above discussion that XAI gives an idea about which kind of machine learning algorithm to use based on the available dataset, which helps in building a better model corresponding to the particular dataset in hands. In future with more amount of test dataset on the same model or building different models and their correlation the certainty of XAI significance can be established well.

5. REFERENCES

1. Shu, Lin & Xie, Jinyan & Yang, Mingyue & Li, Ziyi & Li, Zhenqi & Liao, Dan & Xu, Xiangmin & Yang, Xinyi. (2018). A Review of Emotion Recognition Using Physiological Signals. Sensors. 18. 2074. 10.3390/s18072074.
2. 2. Cannon, W.B. The James-Lange theory of emotions: A critical examination and an alternative theory.Am. J. Psychol. 1927, 39, 106–124.
3. 3. Vallerand, A. H., Cosler, P., Henningfield, J. E., & Galassini, P. (2015). Pain management strategies and lessons from the military: A narrative review. Pain research & management, 20(5), 261-8
4. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0024124.
5. https://ieeexplore.ieee.org/abstract/document/7590729.
6. https://ieeexplore.ieee.org/abstract/document/1617286.
7. https://www.sciencedirect.com/science/article/pii/S109038010400148X.
8. https://towardsdatascience.com/data-correlation-can-make-or-break-your-machine-learning-project-82ee11039cc9.