Comparing Shape Features with Statistical Features for Classification Tasks on Physiological Data

Anonymous Author(s)
Affiliation
Address
email

Abstract

Research in psycho-physiology derives information about physical and emotional state of people based on their physiological data (such as heart rate, skin temperature). In this regard, many use machine learning techniques, especially classification techniques, to find patterns in the data. The importance of features on the results of a classification task is well-known. However, most past studies in this area, use a common set of statistical attributes (e.g. min, max etc.) extracted from a physiological signal as classification features. But those statistical features only capture partial information about a signal which can lead to high classification error in some cases. In this project, I describe another type of feature, shape features, and extract them using K-means clustering algorithm. I compare shape and statistical features by looking at their error rate for gender classification on a publicly available physiological dataset called DEAP [1]. According to the results, shape features with appropriate parameter setting lead to slightly lower error rates. However shape features are much more computationally expensive. Thus, using statistical attributes might be a more reasonable choice in many cases.

1 Introduction

Past research has shown that physiological data, such as heart rate, blood volume pressure, skin conductance and temperature, are valuable source of information about a person's physical and psychological state [2, 3, 4]. Yet physiological data are usually noisy, considerably affected by everyday activities, person-specific and day-dependent. These characteristics make it difficult to use them without substantial processing and make for a big unknown problem space. The huge number of unknowns regarding physiological signals makes it even difficult to define the extent of the research problem properly. Some of the open questions in this area include: What type of information resides in physiological data? What sort of questions can be addressed using these data?

Previous work in this area has extracted various types of information from physiological data using machine learning techniques. As an example, [2] use classification to determine drivers' stress level based on their physiological signals. Most classification work in this area use physiological data as the feature space to predict information about a subject. For example, in a gender classification task, the gender (female/male) labels are determined by using physiological data. Also valence (like/dislike) classification, uses information heart rate etc. to find whether a person likes a music track or not. Interestingly, most previous work use a somewhat similar set of statistical features (e.g. max, min etc.) extracted from the whole or portions of a signal. In some cases, these statistical features lead to low classification accuracy especially for like/dislike labels.

Although the impact of features on classification results are well known, most previous work use a common set of statistical features (e.g. max, min etc.) and just vary the classification technique or even the study condition to achieve better results. To the best of my knowledge, no systematic analysis is available that compares the appropriateness of different types of features for physiological data. In this project, I describe two types of features and compare them based on their resulting classification error. The two types of features include: 1) statistical attributes extracted from signals such as max, min etc. (similar to past studies) 2) shape features extracted by applying K-means clustering algorithm to the segments of physiological signals.

Section 2 presents the related work in classification of physiological data and available psycho-physiological datasets. The methods section includes the description of the dataset used, followed by an explanation of the two feature extraction methods. In the result section, first I show the shape features extracted from the signals. Further, I evaluate the two feature sets by comparing their resulting accuracy on a gender classification task. The discussion and conclusion section comes at the end.

2 Related Work

Rsearchers at SPIN and CARIS labs at UBC have investigated physiological signals for estimating emotions during the last couple of years. M. Pan[1] used Unscented Kalman Filters to estimate the valence states (continuous dislike to like ratings) of a subject in response to music tracks as stimuli. The input to the model were music features as well as the physiological data gathered from the subject's body including heart rate, skin conductance, skin temperature, blood volume pressure, respiration, and muscle movements on the front head. The predication accuracy was highly varied depending on the dataset and the subject. In an earlier work, S. Zoghbi[2] used the kNN algorithm to classify music tracks into three labels of "dislike", "neutral", and "like". Statistical attributes of physiological signals such as mean, min, and max formed the feature space for the classifier. Again the algorithm yielded different classification accuracy for different datasets. Also, the most predictive features selected by the algorithm were not consistent for different datasets. Others have worked on other aspects of using physiological data such as emotional tagging of movies or determining stress level in drivers [5, 2, 1, 3]

2.1 Available Datasets

Despite growing interest in psycho-physiological studies, only a few datasets are publicly available. Most studies use their own collected dataset which makes it difficult to compare the results across different studies.

Local Datasets at CARIS and SPIN: Mutual efforts in CARIS and SPIN labs at University of British Columbia have lead to a number of datasets for music stimuli. These datasets have been a valuable source of research. However, they mostly contain data from one participant, since there are evidence on physiological signals being subject-specific.

DEAP dataset: [1] provides a publicly available multimodal dataset of physiological signals called DEAP (A Database for Emotion Analysis using Physiological Signals). The dataset contains 40 bio-channels recorded from 32 participants (50% female, 19-37 years old) as they watched 40 excerpts of music videos. The participants rated each video on different scales such as valence (like/dislike) and arousal (boring/exciting). The bio-channels include EEG (brain signals) and peripheral physiological signals. The duration of each video clip (stimulus) was 60 seconds with a 3 second baseline data collection before each video. The physiological signals were down sampled to 128Hz. [1] points to some other publicly available physiological datasets such as the enterface 2005 emotional database [4] and MAHNOB-HCI [5].

3 Methods

In this section, first I briefly explain the dataset used in this project. Then, the techniques for extracting two feature sets and their comparison are discussed in more details.

[1]M. Pan is affiliated with CARIS lab, Department of Mechanical Engineering, UBC

[2]S. Zoghbi is affiliated with CARIS lab, Department of Mechanical Engineering, UBC

3.1 Dataset

In this project, I use the peripheral physiological data of the first ten subjects from the DEAP dataset [1]. The eight captured peripheral physiological signals including horizontal EOG(1) and vertical EOG(2), EMG signal from Zygomaticus (3) and Trapezius (4) muscles, skin conductance (5), respiration(6), blood volume pressure(7), and skin temperature (8). EOG signals indicate eye activity/blink rate of the subject, while EMG signals capture muscle movements on chicks and jaw line. Respiration shows the breathing rate and blood volume pressure is an indicator of heart activity. The number of subjects (10 participants, 6 female, 4 male) is chosen for computational purposes and the signals are downsampled to 8 points per second. In the following, a trial refers to all physiological signals captured while a subject was watching one video clip. So there is a one to one correspondence between trials and video clips for each subject.

3.2 Machine Learning Techniques

Statistical Features: The first set of features (statistical attributes) has been the primary choice of researchers in many signal classification tasks. These features are efficiently extracted and usually result in good approximation of the shape of a signal. However, they sometimes fail to capture differences between especially long signals. As a result researchers tend to use additional features (e.g. in frequency domain) or extract the statistical attributes from multiple segments of a signal. This can lead to large numbers of features.

The statistical features typically used in psycho-physiological literature include max, min, average, standard deviation, skewness, and kurtosis. To form the feature vector, first I divide each of the eight physiological signal from a trial into a number of segments (1, 3, 5, 7, or 14 segments are used for cross-validation). Then the seven abovementioned statistical attributes are computed for each segment. The extracted attributes for all segments of the eight physiological signals in a trial compose the final feature vector for that trial (size of feature vector= number of segments*7 attributes*8 signals).

K-means Features: The second set of features (shape bases) is motivated by the works of [6] and [7]. The idea is to extract local shapes representing signals and classify based on the existence or location of those shapes in a signal. [6] proposes the idea of using local shapes for classification of time series. [7] shows that K-means can extract local shape bases and achieves high classification accuracy for images. Since the main shortcoming of statistical features is its deficiency in modeling the shape of the signal, shape bases extracted by K-means algorithm form a potential feature set to investigate.

In this project, I divide each physiological signal in the dataset to N segments by moving a window over the signal (see Figure 1). I also use a pre-defined overlap between consecutive windows (segments) which is a common technique for dividing signals into segments. The resulting set of signal segments is input to the K-means algorithm. K-means algorithm divides the signal segments to K clusters with K centeroids. Those K centeroids show the average shape of signals in their clusters and can be considered as shape features (bases). Thus, we have K shape features after applying K-means to the signal segments.

To form the feature vector for a trial (a video clip), each signal is decomposed to its segments again by using the overlapping window. Then each segment is replaced by the corresponding cluster number (Figure 1). The final feature vector for a trial (a datapoint) includes the result of the last step for all eight physiological signals (See figure 1).

Using cross-validation on window/segment size, I set the window size to 16 signal values representing two seconds of the trial. Also, changing the window overlap between 1/4, 2/4, and 3/4 of the window size and using cross-validation lead to an overlap of 8 signal values or one second (2/4 of window size). Figures 3, 4, 5 show the cross-validation graphs for the number of shape features, window size, and overlap size repectively.

Comparison of the two feature sets: In order to compare the value/utility of these two feature sets, I use their error rate on the gender classification task. For both feature sets, I use the kNN algorithm to perform the classification.

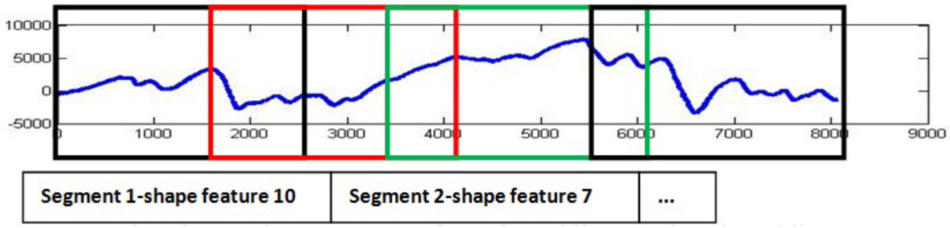

Figure 1: Dividing the signal to segments using a moving window with overlaps- different colors show different segments and are used to better illustrate the overlap- Also Showing the encoding of the segments into shape features (cluster numbers)

k-Nearest Neighbours (kNN): Having the feature vector for a data point, the kNN algorithm assigns the class label for that data point based on the majority vote of its k closest neighbours. For statistical features, I use Eucledian distance as a similarity (proximity) measure. Hamming distance is used for shape features (cluster numbers corresponding to shape features). In all classification tasks, the number of neighbours (k) varies between 1 and 10.

4 Results

4.1 Shape features extracted from signals:

Figure 2 shows the shape features extracted from 198400 signal segments (10 subjects* 40 trials* 8 signals* 62 segments). The extracted shape features have different range of values. For example, the first shape in figure 2 represents signal segments with large negative values, while the second one represents large positive signal segments.

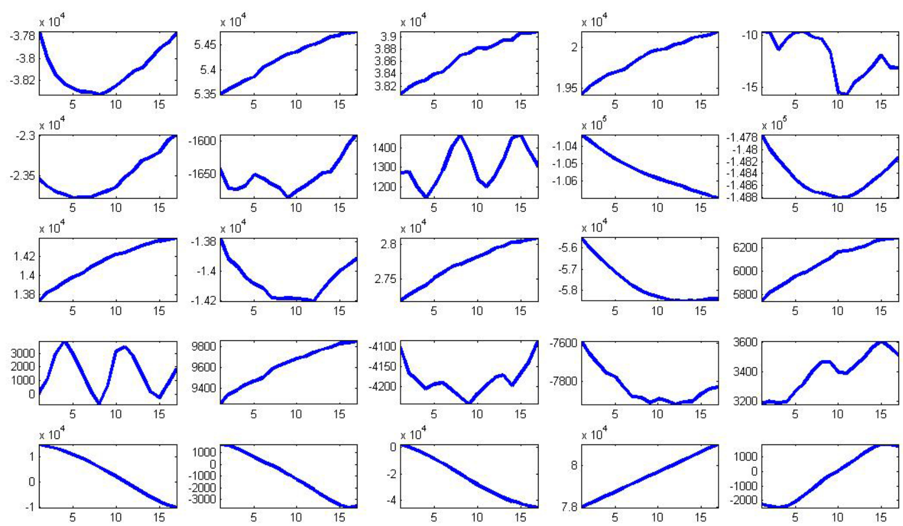

Figure 2: 25 Shape features extracted by K-means algorithm

4.2 Gender classification:

According to figure 3 and 4, the error rate for shape features drops by increasing the number of shape features as well as by using smaller window size. This, on the other hand, increases the computation time. After setting the parameters for shape features with cross-validation, classification with shape features achieves slightly lower error rates compared to the statistical features. The error rate of 12.83% is achieved for 50 shape features, window size= 2 seconds (16 signal values), overlap size= 1 second (8 values). The lowest error rate with statistical features is 13.5% (Figure 4). Despite slightly higher error rate, statistical features are less computational-intensive and lead to almost similar error rates.

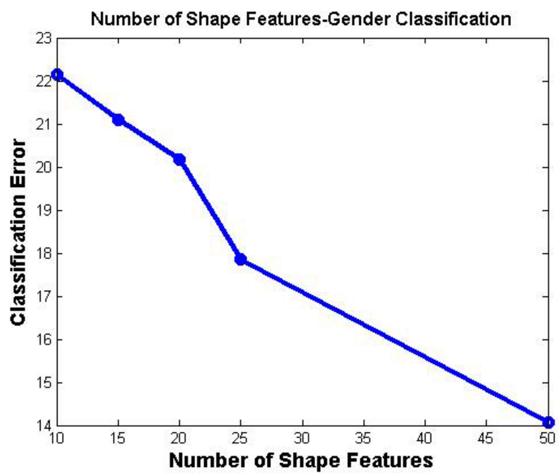

Figure 3: Number of shape features(bases) vs. classification error

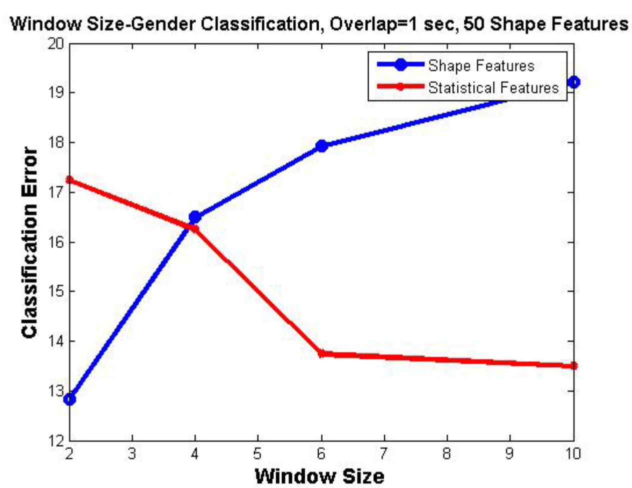

Figure 4: Error rate for shape features and statistical features vs. window size-horizontal axis shows window size in seconds- Overlap size is 1 sec, Number of features is 50

5 Discussion and Conclusion:

In this project, I compare two feature sets for classification tasks on physiological data. The first feature set is statistical attributes extracted from the signals and is commonly used in psycho-physiological studies. The second feature set, motivated by the work of [6] and [7] is composed of shape features extracted by K-means clustering algorithm. The error rates of these two feature sets on gender classification provides a measure of the appropriateness of them for physiological data.

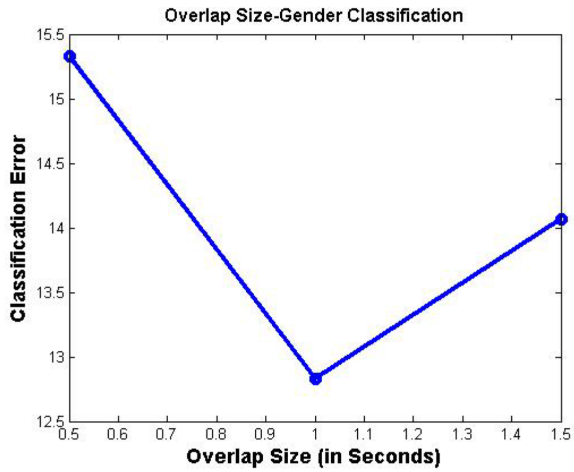

Figure 5: Error rate for shape features vs. overlap size (horizontal axis shows overlap size in seconds)

According to the results, the shape features can achieve similar or slightly lower error rates by carefully setting the parameters (e.g. number of shape features, overlap size, window size) with cross validation. However, these settings lead to higher computational cost compared to statistical features. Comparing the gained classification rate with the computational cost of the two feature sets, statistical features seem to be a more reasonable choice compared to the shape features extracted using the abovementioned method.

In addition, As mentioned before the generated shape features by K-means algorithm represent different value ranges of the signals. Some shape features represent high positive signal segments, some low positive signal segments etc. Thus, they include similar information to more efficient statistical features.

In terms of future work, research on other possible sets of features for time series data, and specifically physiological data, is valuable and can shed light onto the properties of these signals. This is especially true because statistical features achieve unacceptably high error rates for valence (like/dislike) classification tasks. Finally, further work is needed on sophisticated filtering of physiological signals before feeding them into a classification algorithm. Such filtering can lead to better comparison of features and also better classification results.

References

[1] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: a database for emotion analysis using physiological signals. *Affective Computing, IEEE Transactions on*, (99):11, 2011.

[2] J.A. Healey and R.W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156166, 2005.

[3] C.L. Lisetti and F. Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing*, 2004:16721687, 2004.

[4] A. Savran, K. Ciftci, G. Chanel, J. Cruz_Mota, L.H. Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut. Emotion detection in the loop from brain signals and facial images. In *eINTERFACE'06-SIMILAR NoE Summer Workshop on Multimodal Interfaces*, 2006.

[5] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A Multi-Modal affective database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, (99):11, 2011.

[6] Z. Xing, J. Pei, P. Yu, and K. Wang. Extracting interpretable features for early classification on time series. *the Proceedings of SDM*, 2011.

[7] A. Coates, H. Lee, and A.Y. Ng. An analysis of single-layer networks in unsupervised feature learning. *Ann Arbor*, 1001, 2010.