From Joachim Hornegger et al. (Eds.), ***3rd Russian-Bavarian Conference on Biomedical Engineering*** **, Erlangen 2007, pp. 47–51**

Fast Recursive Data-driven Multi-resolution Feature Extraction For Physiological Signal Classification

Florian Hönig, Anton Batliner, and Elmar Nöth*

University of Erlangen-Nuremberg,
Institute of Pattern Recognition (Informatik 5),
Martensstraße 3, 91058 Erlangen, Germany
hoenig@informatik.uni-erlangen.de

Abstract. This study presents a new approach to feature extraction for real-time classification of physiological signals. By using multiple resolutions for the analysis of the signal, the stability of large analysis windows is combined with the capability of small windows to reflect quick changes. A large number of generic features is extracted from each signal for each resolution. These are calculated recursively for each sample which makes them very efficient in terms of computation time; a version with low memory requirements is also provided. A labelled dataset is utilised to convert the generic features into task- and signal-specific features by means of a data-driven transform. The performance of the approach is evaluated on a database containing different stress levels collected in a simulated driving context. A recognition date of 89.8 % is achieved for online, user-independent classification of stress.

Key words: Biosignals, Recursive Calculation, Online Stress Recognition

1 Introduction

Research on Human-Computer-Interaction has recently turned a strong focus on the affective state of the user. Knowledge of this affective user state could lead to more pleasant, safer and more effective user interfaces [1]. For example, an in-car infotainment system as the one developed in the SmartWeb [2] project could respond to a stressed user state by retaining non-vital information in order not to further increase the user's cognitive load.

Affective states are known to have bodily correlates, which can be measured with suitable sensors. Most of the resulting *physiological signals*, e.g. skin conductivity or heart rate, are not under voluntary control and hence not subject to masking like e.g. speech and gesture. Physiological signals are therefore a valuable source of information for affective user state. Several studies have shown the feasibility of recognising affective states using physiological signals [3] [4].

* This work was funded by the EC within HUMAINE (IST-2002-507422) and by the German Federal Ministry of Education and Research (BMBF) within SmartWeb (Grant 01 IMD 01 F). The responsibility for the content lies with the authors.

2 Physiological Signal Processing

A number of problems arise when trying to recognise the affective user state with physiological signals. First, there is a large intra- and interpersonal *variability* of the signals. For a reliable classification, it seems therefore advisable to use large analysis windows in order to smooth out some of the variability. Another difficulty are *artefacts* in the physiological signals produced by motion, pressure, etc. which can render a signal useless for whole passages. Signal analysis and classification should therefore be able to cope with a dynamically varying number of input channels. For many conceivable applications of user state classification, at least a near *real-time* capability is required. On the one hand, this means that the feature extraction must be fast enough for a high classification frequency (i. e. small analysis step size) also for the above-mentioned large analysis windows; on the other hand it means that large analysis windows alone do not suffice because they can hardly provide the information necessary for a quickly reacting classification system.

Our approach seeks to address these issues. It is assumed that some detection algorithm for pronounced artefacts is available that marks passages in each signal which are probably corrupted as unusable. Furthermore, we discuss the case of online classification which means that all analysis windows are causal, i. e. using only samples from the past.

For the present study, six physiological signals are used: electrocardiogram (ECG), electromyogram measured at the neck (EMG), skin conductivity between index and middle finger (SC), skin temperature at the little finger (Temp), blood volume pulse at the ring finger (BVP) and abdominal respiration (Resp). Before computing the features, four additional signals are derived from the actually recorded signals: The heart rate acquired from the ECG channel (HR-ECG) and from the BVP channel (HR-BVP), the lag between ECG and BVP (Lag), which can be regarded as a surrogate parameter of the systolic blood pressure, and the respiration rate (Resp-rate). This has the considerable advantage that no signal-specific algorithms have to be included into the feature extraction. Furthermore, is makes sense to treat these parameters separately with respect to artefact detection. For example, if the heart rate computed from the BVP cannot be used, there might still be useful information in the raw BVP signal.

3 Feature Extraction and Classification

In order to deal with a variable number of input channels, feature extraction is performed separately for each signal. First, it is decided whether the signal is corrupted. Currently, this artefact detection is only a simple rule disqualifying signals with unplugged sensors or physically implausible values for the derived signals. Then, multiple analysis windows of different length (1, 5, 20 and 60 seconds) are extracted. Signals containing a sample marked as corrupted in any of the analysis windows are excluded from further processing for the current point in time. This multi-resolution approach aims at combining the stability of large analysis windows and the capability of small windows to reflect quick changes which is needed for real-time classification.

No attempt was made to design special features for each of the recorded and derived signals or the different window lengths. Instead, a large number of multi-purpose features like mean, standard deviation or slope is extracted from each

of these analysis windows. A labelled dataset is then utilised to create features specialised to the set of states to be recognised and signal at hand by means of a data-driven transform, the Fisher linear discriminant analysis (LDA): the generic features from all analysis windows of a signal are stacked into a single feature vector which is then projected into a lower-dimensional space.

Two different feature sets are provided. The *moving features* are computed recursively for each new sample and thus have a constant computational complexity with respect to the length and step size of the analysis windows. A ring-buffer is used to store the necessary sample history. In effect, these features can be computed very quickly even for the large required window sizes and are well-suited for a possible implementation on limited hardware. The recursive calculation is illustrated by the update rule for the mean value μ_n of a window containing w samples at the n-th sample x_n: $\mu_n = \mu_{n-1} - x_{n-w}/w + x_n/w$. If floating point numbers are used, and unless w is small, errors due the numerical instability of adding and subtracting small values accumulate and render the result useless with time. This can be solved by periodically providing a mean value calculated anew; substituting the recursively calculated value every w samples results in a reasonable degree of numerical stability and only increases the computational effort by a constant factor of about 2. With similar techniques, also mean values as would result from a triangle- and bell-shaped window can be computed recursively, e. g. $\mu_n^{\mathrm{tri}} = \mu_{n-1}^{\mathrm{tri}} - m_{n-w_1}/w_1 + m_n/w_1, m_n = m_{n-1} - x_{n-w_2}/w_2 + x_n/w_2, w_1 = \lfloor w/2 \rfloor, w_2 = w - w_1$. Further recursively computed features are e. g. the slope of the regression line, a smoothed derivative, energy, variance, mean absolute or squared rise, fall and change and approximations of minimum, maximum, median and the amplitude. The variance $\sigma_n^2 = e_n - \mu_n^2, e_n = e_{n-1} - x_{n-w}^2/w + x_n^2/w$ is given as another example. Furthermore, the square, the square root or the absolute value of the computed features is added where applicable, e. g. the absolute value of the slope or the square root of the variance, yielding the standard deviation. In total, 50 moving features are calculated for each analysis window.

The *sliding features* drop the need for a sample history, resulting in a memory requirement independent of the window length. This is favourable for a possible implementation on hardware with small memory. The recursive calculation is illustrated by the update rule for the sliding mean $\mu_{\alpha,n}$ with a parameter $\alpha < 1$: $\mu_{\alpha,n} = \alpha \cdot \mu_{\alpha,n-1} + (1-\alpha) \cdot x_n = (1-\alpha) \sum_{i=0}^{\infty} \alpha^i x_{n-i}$, i. e. $\mu_{\alpha,n}$ is the mean value of the signal multiplied with an exponentially decaying window function. The parameter α determines how quickly the window function approaches zero. The standard deviation is used to characterise this by assigning a nominal window length $w = 2\sqrt{3}/(1-\alpha)$ which is the length of a rectangular window that has the same standard deviation as the exponential window. This rectangular window contains approx. 97 % of the mass of the exponential window. Depending on the desired length of the analysis window, α is computed from the nominal window length. Due to the fact that the window function never actually reaches zero, large outlier values of a signal can corrupt the mean value for a long time. Therefore, $\mu_{\alpha,n}$ is periodically substituted by a value that would result if the exponentially decaying window function was set to zero after 99 % of its mass. Again, the computational effort is only increased by a constant factor of about 2. With similar techniques, "sliding" equivalents of most of the moving features can be calculated, e.g. the mean value over a decaying bell-shaped window, $\mu_{\alpha,n}^{\mathrm{bell}} = (a-1)^2/(\epsilon(1-a-\epsilon)) \cdot \mu_{\alpha+\epsilon,n} - (1-a)/\epsilon \cdot \mu_{\alpha,n}$, and a sliding smoothed derivative, $\delta_{\alpha,n} = (\alpha-1)^3/(2/\epsilon^2) \cdot (\mu_{\alpha,n} - 2\mu_{\alpha+\epsilon,n} + \mu_{\alpha+2\epsilon,n})$ with $\epsilon \to 0$ (for

practical purposes, $\epsilon = (1 - \alpha)/100$ suffices). In total, 44 sliding features are calculated for each analysis window.

The final feature vectors of each valid signal resulting from the LDA transformation of the generic features are scored with a Gaussian Mixture Model consisting of 10 mixture components. The resulting probabilities are, assuming statistical independence between the different physiological signals, combined by multiplication, yielding a final score for each class.

4 Experiments and Results

We evaluate our approach on the Drivawork (Driving under Varying Workload) database which contains audio, video and physiological recordings of different stress levels in a simulated driving context. The six above-mentioned physiological signals have been digitised at 256/2048 Hz with the Mind Media NeXus-10 device. Relaxed and stressed states are elicited by giving the participant different tasks; subjective and objective measures support the effectiveness of this approach. The structured design of the experiment can be used to obtain a preliminary "ground truth"; a fine-grained manual annotation of the perceived stress level is currently being conducted. The database contains recordings of 24 participants and amounts to 15 hours or 1.1 GB of physiological data alone.

We investigate the task of user-independent, online classification of a relaxed or stressed user state using a subset of the Drivawork dataset: due to the fact that the actual user state is unknown, the classification accuracy is only evaluated during the most unambiguous segments. For those 3.4 hours, it is assumed that the affective state of the person is the one intended by the experimental design. Classification is done with a frequency of 1 Hz, so the number of used feature vectors is about 15600. Note that the chosen online classification task is more difficult than the task of discriminating previously defined, relatively large segments in an offline manner as studied e. g. in [4] in the following sense: the context of 60 seconds available to the classification module is relatively small, in addition, it is only taken from the past. So, a considerable fraction (28 %) of the feature vectors is computed from intervals that are not completely contained within the unambiguous segments. However, the task is still artificially simplified by the fact that the studied segments are well separated.

All evaluations are done using person-independent 10-fold cross validation, i. e. each pair of train and test set is disjoint with respect to the participants. The class-wise averaged recognition rates are reported. Table 1 lists the results obtained using different input features, for the individual signals as well as for the combination of all signals. Using the moving features from a single analysis window of length 60 seconds, recognition rates between 48.5 % (Resp-rate) and 80.5 % (ECG) were obtained for the single signals. The combination of all signals yielded an accuracy of 88.1 %. Using the multi-resolution approach with the four analysis windows of 1, 5, 20 and 60 seconds length (i. e. a total of 200 generic features) was better than using only the single window of 60 seconds in all but one case. Again, ECG was the best single channel with 83.8 % recognition rate. For the combination of all signals, 89.8 % resulted in this case. The sliding feature behaved similarly; however, the gain from the multi-resolution approach is not so marked. For the combination even a slight decrease from 89.6 % to 89.5% was observed. Combining moving and sliding features (i e. using a total of 376

Table 1. Class-wise averaged recognition rates in % for recognising stress using single channels or the combination of all signals. For feature extraction, either one analysis window of length 60 seconds ("single") or multiple windows of length 1, 5, 20 and 60 seconds ("multi") are used. The used feature set per analysis window is either the moving or sliding set or both ("All").

Features	*ECG*	*EMG*	*SC*	*Temp*	*BVP*	*Resp*	*HR-ECG*	*HR-BVP*	*Lag*	*Resp-rate*	*Comb.*
Moving single	80.5	67.4	64.7	**77.1**	76.3	75.6	66.3	68.9	54.3	48.5	88.1
Moving multi	**83.8**	**67.6**	**71.4**	76.0	**79.5**	**77.5**	**68.2**	**69.6**	**54.7**	**49.1**	**89.8**
Sliding single	80.9	**73.0**	72.4	76.4	77.7	79.5	**67.8**	68.7	54.7	50.5	**89.6**
Sliding multi	**84.2**	71.3	**75.2**	**76.9**	**78.4**	**79.8**	67.6	68.7	**55.9**	**50.6**	89.5
All multi	**84.3**	72.6	75.0	**77.3**	78.7	**80.1**	67.7	69.5	**56.0**	49.2	88.8

generic features) gave an additional gain only in some cases, but not for the combination. Simulating user adaption by normalising the mean and variance of all features per participant (before estimating the LDA transform), the best recognition rate was obtained for the moving features from multiple resolutions. Here, an accuracy of 96.0 % resulted (not contained in the table).

5 Conclusion

This study presents a unified and efficient approach to feature extraction and classification for physiological signals. No signal- or task-specific knowledge is used to define the features; instead, a labelled dataset is utilised by means of a data-driven transform to convert a large number of generic features into specialised features. The approach is evaluated on the task of online, person-independent classification of relax vs. stress. The results of up to 89.8 % are quite satisfactory and prove that the approach works well. Further research will be devoted to the evaluation of the real-time capability of the system in terms of a reaction speed to user state transitions. It is expected that the multiple analysis windows will be especially useful in this respect. Further studies will investigate a sophisticated artefact detection, recursively calculated spectral features and an un-supervised adaption to the user.

References

1. Nass, C., Jonsson, I.M., Harris, H., Reaves, B., Endo, J., Brave, S., Takayama, L.: Improving automotive safety by pairing driver emotion and car voice emotion. In: CHI '05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM Press (2005) 1973–1976
2. Reithinger, N., Herzog, G., Blocher, A.: Smartweb - mobile broadband access to the semantic web. KI Zeitschrift (2) (2007) 30–33
3. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(10) (2001) 1175–1191
4. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems **6**(2) (2005) 156–166