Fast recursive data-driven multi-resolution feature extraction for physiological signal classification—Abstract

Florian Hönig, Anton Batliner, Elmar Nöth
Institute of Pattern Recognition, University of Erlangen-Nuremberg

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, or even to completely new applications. 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 therefore not subject to masking like e.g. speech and gesture.

Physiological signals are therefore a valuable source of information for acquiring the affective user state. Several studies have shown the feasibility of automatically recognizing at least some few, application-dependent affective states using physiological signals.

2 Physiological Signal Processing

A number of problems arise when trying to recognize the affective user state with physiological signals:

Variability: There is a large intra- and interpersonal variability of the signals. While different affective states exhibit specific reactions in physiological signals *on average*, it can be difficult to assign a specific affective state to a single realization of a physiological signal. For the classification of physiological signals, it seems therefore beneficial to use large analysis windows in order to smooth out some of the variability.

Artefacts: Physiological signals are easily corrupted by motion, pressure, muscle activity or other external influences. These artefacts can be large in magnitude compared to the body function actually measured and render signals useless for whole passages. This conflicts with the need for large analysis windows stated above.

Real-time requirement: 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.

Different Signals: The different physiological signals each exhibit individual properties. This makes the engineering of a dedicated set of features necessary for each physiological modality—and each application-dependent set of states.

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

3 Feature Extraction

For each signal, the length of the primary analysis window is chosen, from a predefined set of lengths, as large as possible without containing an artefact. Then, all smaller predefined window lengths are used to extract further, shorter analysis windows. As we are using causal windows, each of the analysis windows ends at the current time. Signals that do not possess an artefact-free history even for the smallest of the predefined window lengths are ignored for current classification. This multi-resolution approach aims at combining the stability of large analysis windows and the capability of small analysis windows to reflect fast changes.

From each of these sub-windows, a large number of multi-purpose features like mean, standard deviation etc. are extracted. We present two versions for these features: The "moving" features can be computed recursively for each new sample and thus have a computational complexity that is *independent* of the length of the analysis window *and* the step size. A ring-buffer is used to store the necessary sample history; thus, the method has a memory complexity of the largest of the predefined window lengths. In effect, these features can be computed very fast for all sub-windows. The "sliding" features go further and drop the need for a sample history, thus resulting in a constant memory requirement with respect to window length. The resulting features from all sub-windows are stacked into a single feature vector which is then reduced in dimension with the Fisher transform. The large number of different features together with the multiple resolutions aims at creating features specifically adapted to each signal and the task at hand by means of the data-driven transform.

The final feature vectors are scored with a Gaussian Mixture Model. The resulting probabilities are, assuming statistical independence between the different physiological signals, combined by multiplication, yielding a final score for each class. Note that for Fisher transform and Gaussian mixture models, dedicated parameters are estimated as well for each signal as for each primary analysis window length. We evaluate our approach on a stress database collected in a simulated car scenario. For the task of real-time, user-independent classification of stress vs. non-stress, a class-wise averaged recognition rate of 89 % is achieved.