Pattern Recognition Methods – A Novel Analysis for the Pupillographic Sleepiness Test

Jarek Krajewski
krajewsk@uni-wuppertal.de

University Wuppertal
Experimental Industrial Psychology
Gaußstraße 20, 42097 Wuppertal
Germany

Thomas Schnupp
t.schnupp@fh-sm.de

University of Applied Sciences Schmalkalden
Neuroinformatics and Signal Processing
Blechhammer, 98574 Schmalkalden, Germany

Sebastian Schnieder
s.schnieder@uni-wuppertal.de

University Wuppertal
Experimental Industrial Psychology
Gaußstraße 20, 42097 Wuppertal
Germany

David Sommer
d.sommer@fh-sm.de

Christian Heinze
c.heinze@fh-sm.de

Martin Golz
m.golz@fh-sm.de

University of Applied Sciences Schmalkalden
Neuroinformatics and Signal Processing
Blechhammer, 98574 Schmalkalden, Germany

ABSTRACT

The aim of this paper is to improve the information gained by the most commonly applied fit-for-duty sleepiness test (Pupillographic Sleepiness test, PST) by using pattern recognition approaches. The pupil diameter based sleepiness detection is enriched by several new features and machine learning methods. Using all newly computed pupil diameter features we achieved on the two-class detection problem (moderate sleepiness vs. high sleepiness) an accuracy of 83.03% on participant-dependent data with a Random Forest classifier. This result suggested that the PST-standard feature set should be enriched by the here proposed enlarged feature set.

Author Keywords

Pupil diameter, sleepiness detection, machine learning.

ACM Classification Keywords

H.5.m Information interfaces and presentation: Miscellaneous

INTRODUCTION

Measuring sleepiness has been recognized as an important factor for the prevention of a broad range of traffic accidents. Hence, many efforts have been reported in the literature for developing sleepiness detection systems. One of the most promising fit-for-duty tests – the pupillographic sleepiness test (PST) focuses on instability of pupil size [1-3]. The background of this method is that in an alert participant, the pupil remains dilated in darkness with amplitude of change below 0.3 mm and a frequency of approximately 1 Hz. In sleepy subjects, the pupil shows spontaneous oscillations with predominantly low-frequency components and amplitudes reaching several millimeters. Furthermore, the pupil diameter (PD) decreases with time. These changes are measured by infrared video pupillometry during a recording session of usually 11 min. To quantify the sleepiness-induced changes the Pupillary Unrest Index (PUI), a feature reflecting spontaneous oscillations of pupil diameter, is most often applied (even if a few other features as e.g. interpolation rate, lnPUI, relative PUI are sometimes proposed).

Nevertheless, little empirical research has been done to examine the benefit of signal processing and pattern recognition based "brute-force" methods in addition to the original approach. Thus, the aim of this study is to apply multiple state-of-the-art pattern recognition methods [4-7] on the PST based detection of sleepiness. Attention is drawn particularly on the comparison of several commonly applied classifiers. The rest of this paper is organized as follows: Section 2, 3, and 4 describes the experimental design, feature extraction and classification. After the results of the sleepiness detection are provided in Section 5, the paper closes with a conclusion and a discussion of the future work in Section 6.

METHOD

Twenty-seven students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. Initial screening through a questionnaire excluded those having sleep disorders or sleep difficulties (PSQI). The participants were instructed to maintain their normal sleep pattern and behaviour. Due to recording and communication problems, the data of the 6 participants could not be analyzed.We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized self-report fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the reference value of sleepiness. On this scale, a score of 1 point indicates “feeling active and vital, alert, wide awake” and a score of 7 points indicates “almost in reverie, sleep onset soon, losing struggle to remain awake”. For training and classification purposes, the records were further divided into two classes according to the informative values of the PST: low and moderate sleepiness (LS) and high sleepiness (HS) samples with the boundary value SSS $\geq$ 5.0 (5 samples per subject; total number of samples: 111 samples; 77 samples LS, 34 samples HS). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions, they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Feature Extraction. Due to measurement errors, e.g. during eye lid closures, the PD is preprocessed by deleting these segments and correcting them through linear interpolation. The following pupillometric variables (PST standard features; PST-SF) are computed: mean pupil diameter, the PUI, the interpolation rate (due to blinking) and the square root of the power within the frequency band 0–0.8 Hz in pupillary oscillations (fatigue waves). Furthermore, we calculated ‘brute-force’ signal processing based features of pupil diameter within the time, frequency and state space domain (PST-BF). This assignment follows the first processing step of computing frame level descriptors, independent of feature characteristics of the second, contour describing, functional based processing step.

Time domain features (97#+97Δ#). Within the time domain the following features can be extracted: regression descriptors (e.g. regression slope, intercept, maximum of regression error), class distribution measures (e.g. number of values within pupil diameterbin 0.0-0.1), peak amplitudes and distances (e.g. mean distance of peaks; maximum of peak amplitude), entropy, zero crossing distances and slope (e.g. maximum of distance between consecutive zero crossings; mean velocity of pupil diameterin zero crossings).

Frequency domain features (935#+935Δ#). To capture fine temporal changes of spectral descriptors we performed a framing and windowing of the signal (frame size = 512, frame shift = 256, hanning window), and computed the power spectral density per frame. The resulting frame-level descriptors (FLDs) were aggregated to FLD-contours. The next processing step captures temporal information of the FLD contours by computing functionals. Frequently used functionals are percentiles (quartiles, quartile ranges, and other percentiles), extremes (min/max value, min/max position, range), distributional functions (number of segments/ intervals/reversal points), spectral functionals (DCT coefficients), regression functions (intercept, error, regression coefficients), higher statistical moments (standard deviation, skewness, kurtosis, and zerocrossing-rate), means (arithmetic mean and centroid). This procedure of combining FLDs and functionals results in 935 frequency domain features as e.g.: relative and absolute power spectral density (PSD) of raw and first derivate contours in 30 spectral bands (e.g. minimum of relative PSD of first derivate of 0.6-0.7 Hz spectral band contour), band energy ratios (PSD 1-5hz/ PSD 0-1 Hz), spectral flux (e.g. max of spectral flux = Euclidean distance of PSD between consecutive frames), and long term average spectrum descriptors (e.g. skewness of PSD distribution).

CLASSIFICATION

We conducted a validation experiment to examine whether automatically trained models can be used to recognize the SSS based classification of moderate vs high sleepiness. Our approach can be summarized in four steps: 1. Collect individual PST data and the associated sleepiness ratings for each participant; 2. Extract relevant features from the pupil diameter data (PST-SF vs. PST-BF); 3. Apply a correlation filter for feature selection (correlation > .30; PST-SF 4 features vs. PST-BF 44 features remaining); 4. Build PST-SF and PST-BF based classification models using dichotomized SSS in order to solve the two-class detection problem of LS vs. HS; 5. Test the learned models on unseen PST data. Classifiers typically used within pattern recognition based biosignal analysis include a broad variety of dynamic algorithm (Hidden Markov Models) and static classifiers. When choosing a classifier within this highly correlated and noisy feature space, several aspects might be of importance such as low memory, low computation time, quick converging, and no suffering from overfitting. With respect to these requirements, we applied the following static classifiers from the popular 4.5 RapidMiner [8] software using standard parameter settings: Support Vector Machines (‘LibSVM’, rbf kernel function; ‘JMySVM’, linear kernel function; ‘FastLargeMargin’ [9], linear kernel; ‘W-SMO’, Sequential Minimal Optimization), Kernel Logistic Regression (‘MyKLR’), Multilayer Perceptrons (‘NeuralNetImproved’, 2 hidden sigmoid layer, 5 nodes each), k-Nearest Neighbors (‘NearestNeighbors’; k = 3), Decision Trees (‘RandomForest’, 800 trees), Naive Bayes (‘NaiveBayes’, ‘KernelNaiveBayes’), and Fuzzy Lattice Reasoning (‘W-FLR’). In a participant-dependent validation protocol, we applied a stratified 10-fold crossvalidation. The final classification errors were calculated averaging over both classifications.

RESULTS

In order to determine the detection performance, different classifiers were applied on the 2064 features. The recognition rate (RR) of the different classifiers for the two class prediction problems is computed (see Table 1). Within the applied classification schemes the Random Forest classifier using the enlarged brute-force signal processing feature set reached the highest RR of 83.03%. The average RR benefit using the PST-BF instead of PST-SF feature set. For the PST-SF feature set, the 3-Nearest Neighbor classifier achieved the highest recognition rate of 77.42%. Within all applied classification schemes the PST-SP feature set achieved 72.96%, the PST-BF 79.86% accuracy (average improvement 6.9%).

Classifier	Specification	PST-SF	PST-BF	Δ
SVM	LibSVM, kernel = rbf	73.94	79.24	5.30
SVM	FastLargeMargin, kernel = linear	61.26	80.23	18.97
SVM	JmySVM, kernel = linear	75.68	79.32	3.64
SVM	Sequential Minimal Optimization	73.79	77.58	3.79
KLR	MyKLR	75.76	81.21	5.45
NN	k = 3	77.42	81.97	4.55
MLP	NeuralNetImproved, 2 hidden layer, 5 nodes	73.87	75.76	1.89
NB	NaïveBayes	72.12	81.06	8.94
KNB	KernelNaïveBayes	68.33	81.86	13.53
FLC	Fuzzy Lattice Reasoning	73.86	79.12	5.26
RF	#Tree = 500	74.77	83.03	8.26
	Average Classifier	*72.96*	*79.86*	*6.90*

Table 1. Correct recognition rate (RR) (in %) of several classifiers on the test set using participant-dependent validation schemes. PST-SF: 4 standard PST features; PST-BF: 44 features (4 standard + 40 brute-force PST feature).

DISCUSSION

The main findings of the present study may be summarized as follows. First, using all pupil diameter features and all samples (without pre-selecting prototypical classes out of the whole database) we achieved on the two-class detection problem (moderate sleepiness vs. high sleepiness) an accuracy of 83.03% on participant-dependent data with a Random Forest classifier. This result suggested that the PST-standard feature set might be enriched in prospective studies by the here proposed enlarged feature set including known features as relative PUI, logarithm of PUI (lnPUI), logarithm of relative PUI and PSD of several frequency bands [10]. Our classification performance is in the same as has been obtained for comparable tasks, e.g. for sleepiness classification using speech, steering behavior, EEG, posturographic information, cf. [11-14].

Our results are limited by several facts. The present results are (which is a truism) preliminary and need to be replicated using enlarged databases and a greater diversity of participants. We relinquish to optimize the performance of SVM by a fine grained hyperparameter optimization. Therefore, the performance has to be recognized as lower border of the SVM capabilities. Moreover, it would seem advisable that future studies address the main topics of enriching the pupil diameter feature set with eye blinking and lid behaviour features. These camera based features carry further information and might therefore contribute to a higher detection rate of sleepiness. This performance gain might be probably higher than adding other fancy classifier for this sleepiness detection task (e.g. maximum-likelihood bayes classifiers, fuzzy membership indexing, hidden markov models, gaussian mixture density models).

REFERENCES

1. Wilhelm B, Lüdtke H, & Wilhelm H (1999). Spontaneous papillary oscillations. An objective measure for the level of tonic central nervous activation. In J Kuhlmann & M Böttcher (Eds.), Pupillography: Principles, Methods and Applications (pp. 27–36). München: W. Zuckerschwerdt Verlag.
2. Regen, F (2009). Assoziation zwischen Pupillen-Unruhe-Index (PUI) und Korrelaten des zentralnervösen Aktivierungsniveaus im Wach-EEG. Dissertation, FU Berlin.
3. Krajewski, J, Peters T & Wilhelm B (2009). Standardkennwerte des Pupillographischen Schläfrigkeitstest: Ergebnisse einer Schlafdeprivationsstudie. In G. Mayer, & T. Penzel (Eds.), Somnologie Schlafforschung und Schlafmedizin (p. 49-50). Würzburg: Steinkopff Verlag.
4. Sommer D (2009). Analyse des Mikroschlafs mit Methoden der Computergestützten Intelligenz. Aachen: Shaker.
5. Zeng Z, Pantic M, Roisman G. I. & Huang T S (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1):39–58.
6. Krajewski J, Batliner A, & Wieland R (2009). Multiple classifier applied on predicting microsleep from speech. In ICPR (Eds.), 19th Conference on Pattern Recognition (4 pages, no pagination). Tampa/ Florida: IEEE Computer Society Press.
7. Steidl S (2009) Automatic classification of emotion-related user states in spontaneous children’s speech. Logos Verlag, Berlin,. PhD thesis.

8. Mierswa I, Wurst M, Klinkenberg R, Scholz M, & Euler T (2006) YALE: Rapid prototyping for complex data mining tasks, in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).
9. Fan RE, Chang KW, Hsieh CJ, Wang XR, & Lin CJ (2008) Liblinear: A library for large linear classification, Journal of Machine Learning Research 9.
10. Lüdtke, H (2005). Mess- und Analyseverfahren in der Pupillografie. Aachen: Shaker.
11. Krajewski J, Batliner A, & Golz M (2009). Acoustic sleepiness detection – Framework and validation of a speech adapted pattern recognition approach. Behavior Research Methods, 41, 795-804.
12. Krajewski J, Golz M, & Sommer D. (2009). Detecting sleepy drivers by pattern recognition based analysis of steering wheel behaviour. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), Der Mensch im Mittelpunkt technischer Systeme (*S. 288-291)*. Fortschritt-Berichte VDI, Reihe 22, Mensch-Maschine-Systeme. Düsseldorf: VDI-Verlag.
13. Sommer D, Golz M, Schupp T, Krajewski J, Trutschel U, & Edwards D. (2009). A measure of strong driver fatigue. Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 4, 9-15.
14. Schnupp T, Schenka A, Edwards D, Krajewski J, & Golz M. (2009). Is posturography a candidate for a vigilance test? World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings, 25 (9), 388-392.