Schnupp, Thomas; Heinze, Christian; Groß, Horst-Michael; Golz, Martin:

Long short-term memory training for the assessment of vigilance

DOI: 10.22032/dbt.40021

URN: urn:nbn:de:gbv:ilm1-2019210164

Zuerst erschienen in: Biomedizinische Technik = Biomedical Engineering. - Berlin [u.a.] : de Gruyter. - 55 (2010), Suppl. 1, E, p. 122-125.

Erstveröffentlichung: 2010-10-26

ISSN (online): 1862-278X

ISSN (print): 0013-5585

DOI (Sammlung): 10.1515/BMT.2010.705

[*DOI:* *10.1515/BMT.2010.307*]

[*Zuletzt gesehen:* 2019-08-19]

Biomed Tech 2010; 55 (Suppl. 1) © 2010 by Walter de Gruyter • Berlin • New York. DOI 10.1515/BMT.2010.307

Long Short-Term Memory Training for the Assessment of Vigilance

Schnupp T.[1], Heinze C.[1], Groß H.-M.[2], Golz M.[1]
[1] University of Applied Sciences Schmalkalden, Faculty of Computer Science, Schmalkalden, Germany
[2] Ilmenau University of Technology, Neuroinformatics and Cognitive Robotics Labs, 98684 Ilmenau, Germany

t.schnupp@fh-sm.de

Abstract

The assessment of vigilance is of increasing importance within our 24/7 working society. Posturography is one candidate for a quick, mobile and cost-efficient vigilance assessment. Nevertheless classification accuracy is yet insufficient. This contribution aims at improving classification accuracy of posturographical vigilance assessment by utilizing the information hidden within temporal dynamics of feature sequences. For this purpose a Recurrent Neural Network, Long Short-Term Memory (LSTM), is applied. In order to evaluate whether temporal dynamics offer additional information, results from LSTM training are compared to non-recurrent approaches. Results indicate that there is no significant gain in accuracy achieved by learning temporal dynamics.

1 Introduction

1.1 Vigilance

Vigilance is the ability for sustained attention in monotonous situations poor in activation. For many monitoring tasks at work it is necessary to react properly to even weak stimuli. Therefore, the assessment of vigilance has gained importance in modern 24/7 societies. Well-known examples are air traffic controllers during their duty and truck drivers. Monotony and sleepiness have been identified as major factors for decrements in vigilance, oftentimes occurring suddenly and unanticipated. Vigilance monitoring and detection of such events progresses [1]. Vigilance tests have to evaluate quick and robustly whether a person is able to perform an upcoming task or not. An important example is the Pupillographic Sleepiness Test [2] for which normative values have been investigated. The test duration is 11 minutes, which might limit user acceptance for every day operations. Therefore, test durations have to be as short as possible. Posturography may gain this objective, but its accuracy and reliability has to be investigated.

1.2 Posturography

Posturography is a method of quantitative balance assessment and is used as a diagnostic tool in Neuro-Otology. Body sway is measured during sitting or upright standing. For the latter subjects are instructed to stand upright and remain as motionless as possible. Even then human equilibrium is maintained by continuously regulating muscle activity. The gravitational force of the body mass and the regulatory forces are distributed across feet's sole to the ground. Utilizing a a three- or four-point force sensor platform the center of pressure (COP) is calculated. Data analysis quantifies features of the medio-lateral and the anterior-posterior components of the COP time series. During the last decade at least seven research groups have examined posturography for vigilance testing, showing that both components are impaired by hypovigilance [3, 4].

Therefore, posturography is a new candidate for vigilance assessment which may be cost-efficient, short-lasting, mobile, and easy-to-administer. Recent studies in our lab however resulted in relatively high error rates when classifying between two extremes: "alert" vs. "strong fatigue" [5]. One possible explanation is the kind of data analysis. For each signal segment a set of different features were extracted. This contribution focuses on the question whether the learning of temporal dynamics between different segments may improve testing accuracy. The application of Recurrent Neural Networks (RNN), for example Long Short-Term Memories (LSTM) [6], provides means to answer this question. Results are compared to those obtained by the analysis of static feature sets. For both, methods of Computational Intelligence were applied.

1.3 Long Short-Term Memory

RNN are Artificial Neural Networks (ANN) characterized by feedback connections between artificial neurons. These connections conduct output variables from one neuron to its neighbors' and predecessors' inputs [7]. In contrast to non-RNN, they are able to process sequences of feature vectors such that the temporal localization of feature characteristics within the sequence gains relevance. RNN are capable to adapt to temporal dynamics within the input sequences of feature vectors. Among others RNN have been applied to time series prediction [8] and speech recognition [9]. LSTM are regarded as fully RNN trained by a variant of Real-Time Recurrent Learning [6]. An alternative approach is the application of Evolutionary Strategies to adapt input weights of network cells [10].

Schmidhuber et al. [10] demonstrated that combinations of RNN and non-linear classifiers have an improved ability to generalize a sub-symbolic temporal memory in the feature space.

2 Experiments

2.1 Study

10 young and healthy volunteers participated in partial sleep deprivation experiments. The cross-over design included two nights with at least one week of recreation between them. Driving simulation experiments were performed in either the first or second study night. The other night, additional vigilance tests were performed instead of driving simulation.

During registration process applicants filled out a set of questionnaires including the Pittsburgh Sleep Quality Index (PSQI) [11] and the German translation of the Morningness-Eveningness Questionnaire (MEQ, German: DMEQ) [12]. Applicants with PSQI ratings greater than 7 indicating notable sleep disturbances were rejected. In order to achieve hypovigilance during early night hours applicants evaluated by DMEQ to be of eveningness chronotype were rejected, too. The remaining applicants were invited for interview and training sessions. Moreover, subjects experiencing simulator sickness during the training sessions in our real car driving simulator were rejected. The first ten randomly selected applicants meeting all requirements took part in the study.

After subjects completed training sessions for each test and questionnaire, baseline measurements were examined for reference purposes. They were performed during morning hours with time-since-sleep (TSS) between 2 and 4 hours, usually between 10:00AM and 12:00AM, and were assumed to be in a fully alert state. Each subject had to perform three posturographical reference measurements. In both nights eight experimental sessions lasting one hour each were conducted between 8:00PM and 4:00AM. Within each session one posturographical measurement was requested.

Figure 1 Posturography was performed in uniform, upright pose with hands folded ventral.

2.2 Measurements

During posturographical measurements subjects are instructed to maintain a specified upright pose (fig. 1) in order to avoid random movements. In addition subjects were instructed to remain as motionless as possible. One posturographical measurement consisted of two trials with 130 seconds duration each. The first trial was performed with eyes open (EO) and the second one with eyes closed (EC). During EC visual feedback, as one of the three sensory inputs to the postural control system, is disabled. This way the visual perception is blocked and the vestibular as well as the proprioceptive senses provide feedback. In consequence the postural regulation becomes more instable. The adherence to our instructions, especially EO and EC instructions as well as instructions on the pose, was monitored online using video cameras.

3 Methods

3.1 Pre-Processing

From the three force sensors of our platform COP time series of the medio-lateral and the anterior-posterior component were calculated. The data set was divided into two classes. The first class ("alert") contained test runs of the morning (TSS < 5 hours) and the second class ("hypovigilant") contained test runs of the late night (TSS > 14 hours). Sessions between these extremes were not regarded within this contribution.

In order to increase the number of patterns available for the methods of computational intelligence, the acquired trajectories are segmented. In earlier examinations segment lengths of 20 seconds have shown to be optimal. The optimal segment length was evaluated empirically resulting in 20s for time domain features and 30s for spectral domain, respectively.

3.2 Feature Extraction

3.2.1 Sequence Generation

RNN process sequences of feature vectors (FV) instead of single FV. In order to generate these sequences each trajectory segment was processed using a sliding window. Window length and the percentage of overlapping were varied trying to find an optimal sequence setup. For each window FV were extracted as described below.

3.2.2 Time Domain

A set of cost-efficient time-domain features was extracted. Extraction methods of several authors were applied, including amplitude, range, measures of velocity and sway area. For features highly sensitive to noise, like sway path and maximum velocity, a 50 Hz low-pass filter was applied. The final set consists of 15 features for each trajectory, leading to a total of 30 features when both EO & EC

trials are fused. An optimal feature set was obtained using window lengths of 5 seconds overlapping 50%.

3.2.3 Spectral Domain

Power spectral densities (PSD) were estimated using Weighted Overlapped Segment Averaging (WOSA) This method aims at reducing variance of PSD estimations at the cost of a raised bias and reduced spectral resolution. This effect is amplified by averaging logarithmically scaled PSD within equidistant bands. Parameters of the band averaging are the lower cut-off frequency f_{lower}, the upper cut-off frequency f_{upper} and the band width f_{Δ}. The basic setup relies on earlier findings. Empirical parameter optimization in order to gain best classification performances were performed selectively. An optimal feature set was obtained using windows of 10 seconds length with 75% overlapping. Optimal band averaging parameters are $f_{lower} = 0.1$ Hz, $f_{upper} = 20$ Hz and $f_{\Delta} = 0.5$ Hz.

3.3 Classification

Since the acquired data set was unbalanced all classifiers are trained using 25-fold delete-d cross validation with a 90:10 training set size to testing set size ratio. For each fold balanced training and test sets are selected randomly. Training set based Z-Transform was applied to the extracted features within each fold, which was beneficial for all classifiers.

3.3.1 Non-Recurrent Classifiers

The outcome of RNN-based pattern recognition was compared to the output of non-RNN-based algorithms. The optimized feature extraction was comparable between the non-RNN and RNN approach.

As non-RNN approaches the k-Nearest-Neighbor (kNN) algorithm, the pure feedforward ANN Learning Vector Quantization (LVQ) and Support-Vector Machines with Gaussian kernel function (SVM) were utilized. From the family of LVQ Algorithms the Optimized Learning Vector Quantization 1 was chosen. Gaussian kernel was chosen, since it has proven to be a versatile and robust kernel function. Basically all utilized classifiers are able to regularize between local and global decision finding, whereas SVM have the finest granularity of regularization.

It has been proposed that temporal dynamics within signals with fixed lengths can be formulated using a sliding window-based feature extraction. Features from all windows are concatenated to a single FV. This approach focuses on the absolute temporal localization, disregarding contextual relations

3.3.2 LSTM

LSTM training was performed in online and batch mode. Neither showed any significant advantages. The bias of LSTM Gates were chosen without two memory blocks sharing the same bias setup. The number of memory cells was equal for all blocks. Further parameters of network topology including, but not limited to the number of memory blocks, the number of cells in each block and the number of hidden units, were optimized empirically. Due to the complexity of the parameter space the findings within this paper can not be regarded as final optima.

Basic LSTM setup utilizes one output neuron (LSTM-I). Sequences from the class “hypovigilant” were annotated with the target value $t_{I,p} = 1$, sequences from the class “vigilant” with $t_{I,n} = 0$. Classification was obtained by comparing LSTM output on the test set to the threshold value .5.

In order to generate an output space with more degrees of freedom LSTM were trained with two output neurons (LSTM-II). Sequences from the class “hypovigilant” were annotated with the target vector $t_{II,p} = (0; 1)$, sequences from the class “vigilant” with $t_{II,n} = (1; 0)$. Classification was obtained by evaluating which LSTM output neuron showed highest activation.

3.3. LSTM + SVM

Based upon the LSTM-II setup the combination of LSTM and SVM as proposed by Schmidhuber et al. [10] was implemented. The combination of LSTM and SVM required the generation of a valid data set for SVM training which we will call svm-set for simplicity. After an optimal LSTM setup was established for both time and spectral domain, 25 balanced training sets (lstm-sets for simplicity) were generated using delete-d strategy. For each set LSTM was trained using a 5-fold cross validation strategy. This way we obtained a valid balanced svm-set of LSTM test responses for each sample in the lstm-set. The svm-set is annotated using the corresponding labels $t_{I,p}$ and $t_{I,n}$. The classification performance of each SVM run is evaluated using fast leave-one-out validation estimation which is an almost unbiased estimation of the true classification error [11]. The classification performance of the combined meta-classifier is obtained as the mean of all 25 runs.

4 Results

Results (fig. 1) show that for all classifications approaches feature sets from time domain were inferior to spectral domain features. Results of non-RNN classifiers show that SVM outperform kNN and LVQ. Both findings are in concordance to earlier results [5].

Since kNN and LVQ were inferior in classification accuracy, only SVM were applied to the sliding window feature extraction sets (SVM:window). Comparison between static (SVM) and sliding window features (SVM:window) shows no clear difference between both approaches.

Comparison between both LSTM setups LSTM-I and LSTM-II shows no significant difference, too. Comparing recurrent to non-recurrent approaches indicates that the classification performance of LSTM is higher than the performance of kNN and LVQ. Nevertheless static SVM outperform LSTM

Error rates of combined LSTM and SVM classifier did not result in the expected increase of classification performance. In time domain the combined classifier performed slightly better than the single SVM classifier in terms of

mean and standard deviation of classification error. In spectral domain however performance was worse compared to both single SVM and single LSTM. Visualizing the svm-set shows a complementary relation of both output neurons' activation. In order to benefit from additional degrees of freedom this symmetry needs to be broken. One approach could be to include further variables of LSTM in the svm-set, e. g. using the sequence of LSTM responses instead of the single response, i.e. final output vector.
Results show that LSTM was successful in order to discriminate both vigilance states. Nevertheless there is no significant benefit of LSTM, especially when regarding the increased computational cost of sequential feature extraction. SVM proved to be the most versatile and best performing classifier. When SVM are not applicable due to any circumstances LSTM are an alternative to feedforward ANN.

Table 1 Comparison of Classification Errors

Classifier	Error (Time Domain)	Error (Spectral domain)
kNN	**35.8 ± 8.7 %** *k=44*	**18.8 ± 9.0 %** *k=8*
LVQ	**37.1 ± 8.9 %** *n=2*	**19.0 ± 9.0 %** *n=41*
SVM	**22.4 ± 8.3 %** *C=1e3.75 γ=1e-3.375*	**9.0 ± 4.2 %** *C=1e6.375, γ=1e-2.188*
SVM:window	**21.6 ± 7.4%** *C=1e0.2813, γ=1e-2.325*	**10.4 ± 6.1 %** *C=1e6.875, γ=1e-4.875*
LSTM-I	**28.4 ± 7.5 %** *8 blocks, 6 cells each*	**13.2 ± 5.9 %** *6 blocks, 8 cells each*
LSTM-II	**29.9 ± 8.0 %** *6 blocks, 6 cells each*	**12.0 ±-5.9** *8 blocks, 8 cells each*
LSTM + SVM	**21.1 ± 5.6 %** *6 blocks, 6 cells each* *C=1e1.375, γ=1e-1.75*	**15.3 ± 1.5 %** *8 blocks, 8 cells each* *C=1e-1.625, γ=1e0*

5 Conclusions

Despite the absence of a significant increase in classification performance this contribution successfully introduces the application of RNN to posturography in order to test for vigilance. Preliminary results of a relatively small data set of 10 partially sleep-deprived subjects offered that the mean classification performance is comparable to SVM. This is not sufficient due to the increased computational cost of sequential feature extraction. Furthermore the complexity of optimizing LSTM setup is relatively high due to the variety of parameters involved. In order to test for further performance increases future experiments may apply Evolutionary Strategies for both, optimization of LSTM setup and training of LSTM cells' input weights.
Results of Collins et al. [12] indicate that posture in quiet stance is not explainable by random walk models. There are short-term as well as long-term correlations which should by assessable also by sub-symbolic sequence learning. Thus, other RNN than LSTM should be applied in order to investigate the potential of temporal sequence learning.
Finally additional effort has to be put into the process of feature extraction. In this contribution we focused on a subset of time and spectral domain feature extraction methods. Beneath other methods in these domains, additional domains need to be investigated, e.g. state space. It has been shown, that the fusion of features from different approaches successfully increases classification accuracy [13] – yet it is unclear how LSTM will behave with increased input dimensionality.
With an achieved error rate of 9.0 ± 4.2 % using SVM posturographical vigilance assessment with test durations of less than one minute seems applicable. If further decreases of the error rate can be achieved, posturography may offer a quick, mobile, cost-efficient and easy-to-administer method of vigilance assessment. Nevertheless further experiments with a larger pool of subjects have to verify whether these results are reproducible.

Acknowledgement

This study was funded by the Federal Ministry of Education and Research within the research program "Research at university of applied sciences together with enterprises" under the project 176X08.

References

[1] Golz M, Sommer D (2008) Automatic knowledge extraction for fatigue monitoring. In: Mandic DP et al. (eds) Signal processing techniques for knowledge extraction. Springer, New York, 299-316
[2] Wilhelm B et al. (1996) Pupillographie als objektiver Vigilanztest. Wien Med Wochenschr 146:387-9
[3] Liu Y et al. (2001) Changes in postural sway during a period of wakefulness. Occup Med 51(8):490-495
[4] Forsman P et al. (2007) Daytime changes in postural stability. J Occup Environ Med 49(6):591-6
[5] Schnupp T et al. (2009) Is Posturography a Candidate for a Vigilance Test? World Congr Med Phys Biomed Eng, IFMBE Proc 25/IX: 388-9, Munich, Germany
[6] Hochreiter S & Schmidhuber J (1997) Long Short-Term Memory. Neural Computation, 9(8):1735-1780
[7] Mandic DP, Chambers JA (2001) Recurrent Neural Networks for Prediction. Wiley & Sons, London
[8] Hellbach S et al (2009) Time Series Analysis for Long Term Prediction of Human Movement Trajectories. Proc 15th ICONIP, Springer LNCS 5507:567-574
[9] Skowronski, M. & Harris, J. (2007) Automatic speech recognition using a predictive echo state network classifier. Neural Networks, 20(3):414-3
[10] Schmidhuber, J, Gagliolo, M, Wierstra D & Gomez FJ (2006) Evolino for recurrent support vector machines. ESANN:593-598
[11] Joachims T (2002) Learning to Classify Text Using-Support Vector Machines. Kluwer
[12] Collins JJ et al. (1995) Age-related changes in postural control. Exper Brain Res 104:480-492
[13] Sommer D (2009). Analyse des Mikroschlafs mit Methoden der Computergestützten Intelligenz. Shaker