DTU Library

End-to-End Deep Learning Model For Automatic Sleep Staging Using Raw PSG Waveforms

Olesen, Alexander Neergaard; Peppard, P. E.; Sorensen, H. B.; Jennum, P. J.; Mignot, E.

Link back to DTU Orbit

accelerometer data (vector magnitude) were made assessible. Seventeen sleep clinic patients, including patients diagnosed with OSA, undergoing PSG wore on their non-dominant wrist both an Arc device and a research-grade actigraph (Actiwatch Spectrum, Philips, Bend OR). Time-stamped minute-to-minute data from each participant were aligned with scored PSG studies. The Cole-Kripke algorithm was used to determine sleep or wake for each 60s epoch on both the Arc and Actiwatch.

Results: Compared to the gold-standard PSG, Arc has an accuracy of $89.5 \pm 1.1\%$ (SEM), sleep sensitivity (or wake specificity) was $95.3 \pm 1.1\%$, and sleep specificity (wake sensitivity) was $50.1 \pm 5.0\%$; while Actiwatch has an accuracy of $89.9 \pm 1.1\%$ (SEM), sleep sensitivity (or wake specificity) was $95.5 \pm 0.9\%$, and sleep specificity (wake sensitivity) was $52.0 \pm 4.2\%$.

Conclusion: Preliminary results indicate that, as compared to minute-to-minute PSG, sleep and wake estimates generated by a consumer-grade wearable (Arc) were comparable to those generated by a clinical-grade actigraph (Actiwatch Spectrum). Further studies may allow for optimization of sleep/wake algorithm to improve accuracy of Arc.

Support (If Any): Stanford CTSA UL1 TR001085, Klarman Family Foundation, NIH T32 HL110952, NIH K23 NS101094.

0315

INTER- AND INTRA-EXPERT VARIABILITY IN SLEEP SCORING: COMPARISON BETWEEN VISUAL AND AUTOMATIC ANALYSIS

Muto V[1], Berthomier C[2], Schmidt C[1], Vandewalle G[1], Jaspar M[1], Devillers J[1], Chellappa S[1], Meyer C[1], Phillips C[1], Berthomier P[2], Prado J[2], Benoit O[2], Brandewinder M[2], Mattout J[3], Maquet P[1]
[1]GIGA-Cyclotron Research Centre-In vivo Imaging, University of Liège, Liège, BELGIUM, [2]Physip, Paris, FRANCE, [3]CRNL, Brain Dynamics and Cognition Team, INSERM, U 1028, UMR 5292, Bron, FRANCE

Introduction: Visual sleep scoring (VS) is affected by inter-expert (difference in scoring between several scorers working on the same recording) and intra-expert variability (evolution in the way to score of a given expert when compared with a reference). Our aim was to quantify inter and intra-expert sleep scoring variability in a group of 6 experts -working at the same sleep center and trained to homogenize their sleep scoring- by using the validated automatic scoring (AS) algorithm ASEEGA, which is fully reproducible by design, as a reference.

Methods: Data were collected in 24 healthy young male participants (mean age 21.6 ± 2.5 years). 4 recordings (data set 1, DS1) were scored by the 6 experts (24 visual scorings) according to the AASM criteria, and by AS, which is based on the analysis of the single EEG channel Cz-Pz. Other 88 recordings (DS2) were scored a few weeks later by the same experts (88 visual scorings) and AS. The epoch-by-epoch agreements (concordance and Cohen kappa coefficient) were computed between all VS, and between VS and AS.

Results: Inter-expert agreement on DS1 decreased as the number of experts increased, from 86% for mean pairwise agreement down to 69% for all 6 experts. Adding AS to the pool of experts barely changed the kappa value, from 0.81 to 0.79. A systematic decrease of the agreements was observed between AS and each single expert between DS1 and DS2 (-3.7% on average).

Conclusion: Inter-expert differences are not restricted to a small proportion of specific epochs that are difficult to score, even when the expert team is very homogeneous. Intra-expert variability is highlighted by the systematic agreement decrease across datasets, and can be interpreted as a scoring drift over time. Even if autoscoring neither provides any ground truth, nor can improve the inter-scorer agreement,

it can efficiently cope with the intra-scorer variability, when the AS used is perfectly reproducible and largely insensitive to experimental conditions. These properties are mandatory when dealing with large dataset, making autoscoring methods a sensible option.

Support (If Any): None.

0316

END-TO-END DEEP LEARNING MODEL FOR AUTOMATIC SLEEP STAGING USING RAW PSG WAVEFORMS

Olesen AN[1,2,3], Peppard PE[4], Sorensen HB[1], Jennum PJ[3], Mignot E[2]
[1]Department of Electrical Engineering, Technical University of Denmark, Kgs Lyngby, DENMARK, [2]Stanford Center for Sleep Sciences and Medicine, Stanford University, Stanford, CA, [3]Danish Center for Sleep Medicine, Rigshospitalet, Glostrup, DENMARK, [4]Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI

Introduction: Deep learning has seen significant progress over the last few years, especially in computer vision, where competitions such as the ImageNet challenge have been the driving factor behind many new model architectures far superior to humans in image recognition. We propose a novel method for automatic sleep staging, which relies on current advances in computer vision models eliminating the need for feature engineering or other transformations of input data. By exploiting the high capacity for complex learning in a state of the art object recognition model, we can effectively use raw PSG signals to detect and classify sleep stages in a robust and reliable way.

Methods: A total of 2322 PSG studies from the Wisconsin Sleep Cohort were used in this study. Central and occipital EEG, left and right EOG, and chin EMG signals were extracted from all PSGs and subjected to initial pre-processing of zero-phase Butterworth bandpass filters with AASM-specified cutoffs. The raw signals were then segmented into 30 s epochs and fed as inputs to a novel deep neural network model based on the ResNet-50 architecture. The model was optimized over cross-entropy loss with respect to annotated scorings using the Adam optimizing algorithm and trained on a subset of 1858 PSGs. Hyperparameters were tuned using 40 iterations of random search in relevant hyperparameter intervals. Best performing model was selected based on performance measured by overall accuracy on a hold-out validation set of 232 PSGs.

Results: Training accuracy, precision and recall were 84.93%, 97.42% and 97.02%, respectively. Evaluating on the validation set yielded an overall accuracy of 85.07% and overall precision/recall of 98.54% and 95.72%, respectively.

Conclusion: Preliminary results indicate that state of the art deep learning models can effectively be used to classify sleep stages using untransformed PSG signals. We will perform further testing on independent datasets to enhance the model's utility.

Support (If Any): Klarman Family Foundation, Technical University of Denmark, University of Copenhagen, Reinholdt W. Jorck Foundation, Knud Hoejgaard Foundation, Otto Moensted Foundation, Vera & Carl Michaelsens Foundation, Augustinus Foundation, and Stibo Foundation.

0317

NORMAL VALUES AND REPEATABILITY OF A NOVEL SLEEP DEPTH SCALE; THE ODDS RATIO PRODUCT

Penner C[1,2], Gerardy B[2], Williams M[2]
[1]University of Manitoba, Winnipeg, MB, CANADA, [2]Cerebra Health Inc, Winnipeg, MB, CANADA

Introduction: The Odds Ratio Product (ORP) has recently been described as a method of objectively measuring sleep depth using the