Human Factors and Systems Interaction, Vol. 52, 2022, 123–129
https://doi.org/10.54941/ahfe1002147

Time Series Representation using TS2Vec on Smartwatch Sensor Data for Fatigue Estimation

Adria Mallol-Ragolta[1], Michail Maniadakis[2], George Papadopoulos[2], and Björn Schuller[1,3]

[1]Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
[2]Institute of Computer Science, Foundation of Research and Technology – Hellas, Greece
[3]Group on Language, Audio, & Music, Imperial College London, UK

ABSTRACT

This work investigates the use of TS2Vec time series representations in an end-to-end approach to detect the fatigue levels perceived by workers of the transportation and logistics industry from the analysis of the accelerometer and the heart rate measurements sensed using a Garmin Vivoactive 3 device. The experiments are conducted using the dataset collected during a pre-pilot study with a total of 1 h 22 min 20 sec of data available. The results obtained support the use of TS2Vec representations for the task at hand, as the binary model trained using this approach and exploiting the heart rate modality obtains the best performance with an Unweighted Average Recall of 67.1 %.

Keywords: Artificial intelligence, Time series representation, Fatigue estimation, Wearable devices, Ubiquitous computing

INTRODUCTION

Fatigue is a condition that affects the performance of employees in a wide range of occupations. Fatigue in the workplace is critical, as it can lead to a poor reasoning and decision making, or to a prominent deterioration in creative problem solving. To prevent risky situations in industrial environments, it is therefore important to early detect physical strain and fatigue, as tired workers are more prone to being involved in work-related accidents. The detection of fatigue has been systematically investigated for more than two decades (Van Dongen 2004). Typically, bio-mathematical models are used to predict performance changes over the day caused by fatigue (Flynn-Evans et al. 2020). Daily sleep time and the so-called sleep debt have been shown to play a very important role in employee's fatigue. Circadian disturbances are also critically correlated to fatigue events (McCauley et al. 2013).

The widespread use of smartwatches provides new opportunities for a more accurate and real-time fatigue monitoring. These devices allow the

© 2022. Published by AHFE Open Access. All rights reserved.

unobtrusive monitoring of workers while they operate in work environments without any physical discomfort or distraction from their work-related duties. The present work focuses on the use of smartwatches to estimate the fatigue of the workers in the transportation and logistics industry – specifically, from the analysis of accelerometer and heart rate information sensed with a Garmin Vivoactive 3 device, and using the self-reported fatigue levels perceived by the workers at different points in time throughout their working days as the ground truth information.

Herein, we investigate the performance of binary end-to-end models targeting the detection of workers' perceived fatigue levels mono- and multimodally. To extract embedded representations from the available modalities, we explore the use of TS2Vec, a universal framework for learning time series representations (Yue et al. 2022). The current study is part of a broader system which aims to provide personalised recommendations that support the employment, safety, and health of aging workers in occupational contexts (Mallol-Ragolta et al. 2022). By estimating the fatigue of users in real-time, we envision the composite system to be able to issue fatigue-related recommendations that allow minimising risky and stressful conditions and situations for both individual workers and worker groups.

The rest of this paper is laid out as follows. We first describe the dataset analysed. The next section details the methodology followed, and we then present and interpret the experimental results obtained. The last section concludes the paper and suggests research directions for future works.

DATASET

This work targets the automatic detection of workers' perceived fatigue by analysing the accelerometer and heart rate data collected using a smartwatch – specifically, using a Garmin Vivoactive 3 device, which senses the accelerometer information at a sampling rate of 25 Hz, and the heart rate information at a sampling rate of 1 Hz. We conducted a pre-pilot study with 3 male workers from the transportation and logistics industry, providing a total of 1 h 22 min 20 sec of data. The workers involved in this pre-pilot study were asked to record their perceived fatigue level via an Android-based smartphone app 6 times a day: before starting the shift, before and after the lunch break, after finishing their shift, and at two random times in the evening. In the app, workers were asked to select the statement that best matched their perceived fatigue level according to the 7-point fatigue scale defined by Samn-Perelli (Samn and Perelli 1982), which is widely used nowadays in fatigue research: 1) fully alert, wide awake; 2) very lively, responsive, but not at peak; 3) okay, somewhat fresh; 4) a little tired, less than fresh; 5) moderately tired, let down; 6) extremely tired, very difficult to concentrate; 7) completely exhausted, unable to function effectively. The initial exploration of the ground truth information provided directly by the workers revealed the scarcity of samples in terms of representing the whole spectrum of the Samn-Perelli fatigue scale. Consequently, in this work, we investigate the problem as a binary classification task, clustering the levels 1) and 2) in the Samn-Perelli fatigue scale into the *no-fatigue* class, and the levels 3) to 7), into the *fatigue* class.

After this clustering, the dataset investigated in this work contains 50 and 197 samples corresponding to the no-fatigue and the fatigue classes, respectively.

METHODOLOGY

This section describes the methodology followed in this work. We first detail the pre-processing applied to the collected data. Next, we describe the models implemented, and summarise their training conditions.

Data Preparation

Two different parameters are logged in our system when workers provide the ground truth information via the Android app: the self-reported fatigue level perceived, and a timestamp, identifying at which point in time the annotation was reported. During the data collection, we assume that workers' perceived fatigue does not change drastically in short periods of time. Based on this assumption, we select the smartwatch sensor data received from the current worker in the 20 min prior to the self-report and annotate this batch of data with the self-reported fatigue level. To benefit the extraction of salient information from the sensed measurements with neural networks in an end-to-end system, we segment the available data into non-overlapping windows of 20 sec length, which are then used for training the models and assessing their performance.

Two different modalities are available from the collected data: triaxial accelerometer and heart rate measurements. A dedicated processing is applied to each modality. The accelerometer measurements in the x-, y-, and z-axes are filtered using a 1-dimensional Gaussian filter with a standard deviation of 1 to smooth them and reduce noise (Zhuang and Xue 2019). We then compute the first and the second order derivatives of the filtered measurements in the x-, y-, and z-axes separately in order to model the velocity and the acceleration of the measurements over time. The raw measurements in addition to their first and second order derivatives form the 9-dimensional input traces to be fed into the networks. In terms of the heart rate measurements, we also compute their first and second order derivatives to model their dynamics over time. The actual heart rate measurements are then normalised by a factor of 220 Beats Per Minute (BPM), which is widely considered as the maximum heart rate of a human being, disregarding the age-dependency in the computation of the maximum heart rate of an individual from a theoretical point of view (Fox and Naughton 1972). The resulting 3-dimensional traces form the input of the networks to model the heart rate information.

Time series representation is a challenging problem, and, as a consequence, the characteristics extracted from the original measurements, defined in this work as the baseline, can be suboptimal. To overcome this issue, we investigate the performance of TS2Vec (Yue et al. 2022) for this task. Specifically, we explore two Transfer Learning (TF) approaches to train TS2Vec models: i) using the GesturePebbleZ2 and the ECG200 datasets from the UCR time series classification archive (Dau et al. 2019) to generate a time series representation of the individual raw triaxial accelerometer measurements and the normalised heart rate information, respectively (results encoded as TF in our experiments); and ii) using the corresponding modalities and analogous representations of the harAGE dataset (Mallol-Ragolta et al. 2021) to generate time series representations of the characteristics extracted from the original measurements (results encoded as harAGE in our experiments). In all scenarios, the TS2Vec models are trained to produce a 16-dimensional representation of the input time series. Furthermore, we aim to compare the performance differences of the TS2Vec representations when encoding the sequential data into a single or sequential representation, which results are encoded as TS2Vec and TS2VecSeq in our experiments, respectively.

Models Description

The networks implemented in this work are composed of two main blocks: the first block extracts deep learnt representations from the input traces, while the second block is responsible for the actual classification. For the feature extraction block, we implement a 1-dimensional convolutional layer with 16 filters, a kernel size of 2, and a stride of 1. Following the convolutional block, we use batch normalisation, and the output is transformed using a Rectified Linear Unit (ReLU) function. A 1-dimensional adaptive average pooling layer is included at the end of the convolutional block, so it generates a 16-dimensional embedded representation of the monomodal input traces. When training the multimodal models, the input traces from each modality are analysed using a dedicated feature extraction block. Following an inner-stage fusion, the outputs of both blocks are concatenated before being fed into the section of the network responsible for the classification. We use this feature extraction block to process the baseline representations of the input modalities.

The following adjustments are performed to the aforementioned feature extraction block when modelling the TS2Vec representations. For the TF TS2Vec representation from the triaxial accelerometer measurements, the feature extraction block is replaced by a Fully Connected (FC) layer, acting as a feature reductor from the 48-dimensional into a 16-dimensional representation. For the TF TS2Vec representation from the normalised heart rate measurements and the harAGE TS2Vec representations, this block is fully removed. For the TF TS2VecSeq and the harAGE TS2VecSeq representations, the feature extraction block is preserved as in the baseline model.

The classification block implements two FC layers, preceded by a dropout layer with probability 0.3. The number of input neurons in the first FC layer depends on the number of modalities to be fused during the training process. The first FC layer produces a 4-dimensional representation, which is transformed using a ReLU function. This transformed representation is fed into the second FC layer to perform the final binary classification. Network outputs are transformed using a Softmax function, so they can be interpreted as probability scores.

Networks Training

For a fair comparison among the models, these are all trained under the exact same conditions. We employ the Categorical Cross-Entropy as the loss to

Table 1. Descriptive statistics (mean- μ, standard deviation - σ) of the UAR scores computed from the nested 3-fold cross-validation approach followed to assess the performance of the binary monomodal and multimodal models trained.

	Baseline		TF TS2Vec		harAGE TS2Vec		TF TS2VecSeq		harAGE TS2VecSeq	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Accelerometer	51.5	2.6	45.5	4.9	**51.6**	8.5	50.0	0.0	41.8	10.5
Heart Rate	62.2	15.2	50.0	0.0	50.0	0.0	**67.1**	13.0	63.4	9.6
Fusion	50.0	0.0	54.2	15.9	**56.1**	4.4	55.4	10.1	55.8	8.9

minimise, using Adam as the optimiser with a fixed learning rate of 10^{-3}. To compensate for the imbalanced instances in the no-fatigue and fatigue classes, we implement a weighted random sampler to select the training samples at each epoch. Model performances are assessed following a nested 3-fold Cross-Validation (CV) approach. The Unweighted Average Recall (UAR) is employed as the evaluation metric, and, consequently, we define (1 – UAR) as the validation error to monitor during the training process. Network parameters are updated in batches of 64 samples, and trained during a maximum of 100 epochs. We implement an early-stopping mechanism to stop training when the validation error does not improve for 20 consecutive epochs.

EXPERIMENTAL RESULTS

The results obtained from the experiments conducted are summarised in Table 1. As it can be seen, the TS2Vec-based representations surpass the baseline representations in most of the cases investigated. The TF TS2VecSeq representation from the heart rate information scores the highest UAR of 67.1 %. The best performances when exploiting the accelerometer information and when fusing the accelerometer and the heart rate measurements are achieved using the harAGE TS2Vec representations, with a UAR of 51.6 % and 56.1 %, respectively. Hence, the results obtained support the use of TS2Vec representations to detect workers' fatigue levels from the analysis of smartwatch sensor data.

Comparing the performance of the TS2Vec-based models when encoding the sequential data into single representations, we observe that the TS2Vec representations trained using the harAGE dataset surpass those trained using the GesturePebbleZ2 and the ECG200 datasets in 2 out of the 3 scenarios investigated. On the other hand, when encoding the sequential data into sequential representations, the TS2Vec representations trained using the harAGE dataset only obtain a better performance than those trained using the GesturePebbleZ2 and the ECG200 datasets in 1 out of the 3 scenarios compared.

CONCLUSION

This work presented an end-to-end approach to detect the perceived fatigue of workers in the transportation and logistics industry. The proposed approach exploited the information embedded in the triaxial accelerometer and the heart rate information sensed using a Garmin Vivoactive 3 device. Specifically, we investigated the performance of TS2Vec-based models to extract time series representations from the input modalities. The results obtained supported the use of TS2Vec for the task at hand, and suggested the suitability of the heart rate modality to detect workers' perceived fatigue, as it achieved the best performance with a UAR of 67.1 %, when tackling the task as a binary classification problem.

The main limitation of this work, especially in terms of assessing the performance of the trained models, is due to the limited data available. Hence, as future work, we aim to increase the sample size for a stronger assessment and validation of the trained models. By increasing the sample size, we also aim at improving the performance of the current fatigue estimation models. Further research directions include the investigation of other time series representation techniques, and the exploration of advanced fusion strategies to exploit the complimentary information embedded in the available modalities.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 826506 (sustAGE).

REFERENCES

Dau, H., Bagnall, A., Kamgar, K., Yeh, C., Zhu, Y., Gharghabi, S., Ratanamahatana, C., and Keogh, E. (2019), "The UCR Time Series Archive." *Journal of Automatica Sinica*, 6 (6), 1293–1305, IEEE.

Flynn-Evans, E., Kirkley, C., Young, M., Bathurst, N., Gregory, K., Vogelpohl, V., End, A., Hillenius, S., Pecena, Y., and Marquez, J. (2020), "Changes in performance and bio-mathematical model performance predictions during 45 days of sleep restriction in a simulated space mission." *Scientific reports*, 10, Article ID: 15594, 14 pages, Nature.

Fox, S., and Naughton, J. (1972), "Physical Activity and the Prevention of Coronary Heart Disease." *Preventive Medicine*, 1 (1 – 2), 92–120, Elsevier.

Mallol-Ragolta, A., Semertzidou, A., Pateraki, M., and Schuller, B. (2021), "harAGE: A Novel Multimodal Smartwatch-based Dataset for Human Activity Recognition." In *Proceedings of the 16th International Conference on Automatic Face and Gesture Recognition*, Jodhpur, India – Virtual Event, 7 pages, IEEE.

Mallol-Ragolta, A., Varlamis, I., Pateraki, M., Lourakis, M., Athanassiou, G., Maniadakis, M., Papoutsakis, K., Papadopoulos, T., Semertzidou, A., Cummins, N., Schuller, B., Karolos, I.-A., Pikridas, C., Patias, P., Vantolas, S., Kallipolitis, L., Werner, F., Ascolese, A., and Nitti, V. (2022), "sustAGE 1.0 – First Prototype, Use Cases, and Usability Evaluation." In *Proceedings of the 7th International Conference on Human Interaction & Emerging Technologies: Artificial Intelligence & Future Applications*, Lausanne, Switzerland, 10 pages, to appear.

McCauley, P., Kalachev, L., Mollicone, D., Banks, S., Dinges, D., and Van Dongen, H. (2013), "Dynamic Circadian Modulation in a Biomathematical Model for the Effects of Sleep and Sleep Loss on Waking Neurobehavioral Performance." *Sleep*, 36 (12), 1987–1997, Oxford University Press.

Samn, S., and Perelli, P. (1982), "Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations." *USAF School of Aerospace Medicine – Brooks Air Force Base, Texas*, Report No. 82-21, ADA 125319, 29 pages.

Van Dongen, H. (2004), "Comparison of Mathematical Model Predictions to Experimental Data of Fatigue and Performance." *Aviation, space, and environmental medicine*, 75 (3), A15–A36, Ingenta.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., and Xu, B. (2022), "TS2Vec: Towards Universal Representation of Time Series." In *Proceedings of the 36th Conference on Artificial Intelligence*", Virtual Event, 9 pages, to appear, AAAI.

Zhuang, Z., and Xue, Y. (2019), "Sport-Related Human Activity Detection and Recognition Using a Smartwatch." *Sensors*, 19, Article ID: 5001, 21 pages, MDPI.