Predicting and classifying stress-related human sensor data

Marijn Schraagen and Hans Marien

Utrecht University, The Netherlands
m.p.schraagen@uu.nl

Abstract. Low literacy is a growing problem in The Netherlands. We measured stress response during a sentence reading task using physiological sensors in both low literate and high literate participants. We will investigate whether various possible machine learning approaches are able to predict and classify aspects of the collected data in several directions, using sensor readings, recorded speech data, reaction times, literacy level of the participant, difficulty of the sentences in the task and other aspects of interest.

Keywords: Sensor data · Speech processing · Classification

1 Data collection

Based on the Age of Acquisition (AoA) dataset of Ghyselinck et al. [2], a set of 80 sentences in Dutch has been created containing 40 words with AoA ≥ 11 years (difficult) and 40 words with AoA ≤ 5 years (easy). For half of the words a real sentence has been retrieved from OpenSoNaR [3] with a length of 7–14 words. Each real sentence is paired with another sentence of the same length from OpenSoNaR where one word has been replaced with one of the remaining 40 AoA words, see Figure 1. The replacement was done such that the resulting sentence would still be grammatical but no longer semantically valid.

difficulty (AoA)	*word*	*validity*	*sentence*
difficult (15.7)	onyx	valid	Ik heb een aantal ringen gemaakt met zwarte onyx.
difficult (14.7)	atol	invalid	Het atol zelf koken en opeten is geen risico.
easy (4.4)	ijsje	valid	Je mag een keer per week een ijsje eten.
easy (4.5)	ruzie	invalid	Eind mei ging het toen de verkeerde ruzie op.

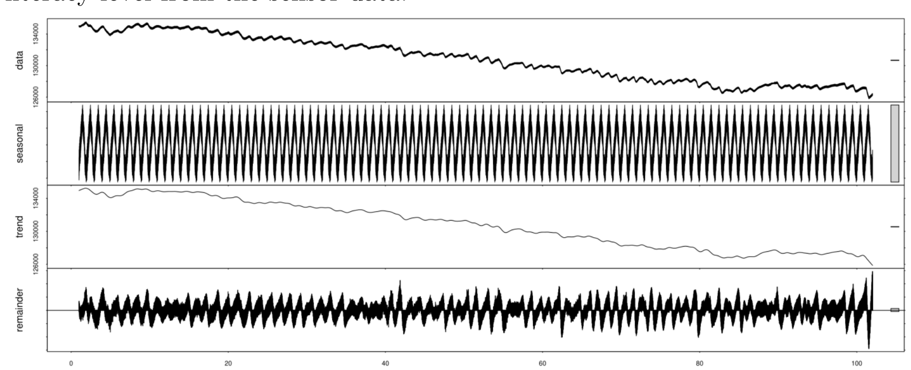
Fig. 1. Example sentence pairs.

An experiment was performed with 37 low literate and 36 high literate participants. Each participant had to read the 80 sentences out loud while their galvanic skin response (GSR) and photoplethysmogram were recorded using sensors on the hand. Half of the participants additionally had to decide if a sentence was valid or invalid by pressing a button before reading the sentence out loud. Sound recordings of the participant reading the sentences were stored, as well as their reaction times in pressing the button and various other timing measurements.

2 Machine learning models

The project is currently in the modelling stage[1]. We use the transformer-based Wav2Vec2 model [1] for the audio classification. Using Wav2Vec2 as a pre-trained ASR model results in a WER of 0.24 for high literate and 0.57 for low literate participants when comparing the target sentence to the ASR transcription. The WER reflects both ASR errors and human reading errors in unknown proportions, but it shows that the model can represent the data adequately. This is a starting point for fine-tuning a task-specific model to identify if a speaker is from the low literate or high literate group based on the audio. Furthermore, given that a majority of the low-literate group consists of non-native speakers of Dutch, we can train a model to recognize non-native speakers and analyze the performance compared to the low vs. high literate model for a more comprehensive interpretation of the performance of both models. Finally we are planning to train models to predict whether the participant judged the sentence to be valid or invalid, as well as a model to predict the trends in sensor readings.

Analysis of the sensor data has revealed various patterns and trends in the sensor data across participants, showing downwards, stable and upwards trends for different participants. An example of a downwards trend for the GSR signal is shown below, as a time series decomposition. We aim to predict the continuation of the signal from initial values, and classify correct decisions, difficulty, and literacy level from the sensor data.



Acknowledgments. This research is part of the project *Estimating literacy levels by recognizing stress using machine learning* with file number NGF.1609.23.025 of the research programme NGF - AiNed 2023 which is financed by the Dutch Research Council (NWO).

[1] If the abstract is accepted for a poster presentation at BNAIC 2024 then we will likely be able to present a number of modelling results on the poster.

References

1. Baevski, A., Zhou, H., Mohamed, A., Auli, M.: wav2vec 2.0: A framework for self-supervised learning of speech representations. In: Proceedings of the 34th Conference on Neural Information Processing Systems. pp. 12449–12460 (2020)
2. Ghyselinck, M., Moor, W.D., Brysbaert, M.: Age-of-acquisiton ratings for 2816 Dutch four- and five-letter nouns. Psychologica Belgica **40**(2), 77–98 (2000)
3. Reynaert, M., van de Camp, M., van Zaanen, M.: OpenSoNaR: user-driven development of the SoNaR corpus interfaces. In: Proceedings of the 25th International Conference on Computational Linguistics: System Demonstrations. pp. 124–128 (2014)