An Unsupervised Method for Head Movement Detection

Yu Wang, Hendrik Buschmeier

Digital Linguistics Lab, Bielefeld University

Introduction

Background

- Head movements are important signals in human-human interaction and head gesture detection has been done quite successfully.
- Most of the proposed models for head movement detection, however, use supervised learning methods.

Goal

Developing unsupervised method for head movement detection. At current stage we are interested in detecting the movement in general, not specific head movement categories (nod, shake, …).

Methodology

Work 1: Data collection

We are building a multimodal corpus (MUNDEX) where an explainer explains a board game to an explainee

(See presentation "MUNDEX: A multimodal corpus for the study of the understanding of explanations" on Friday)

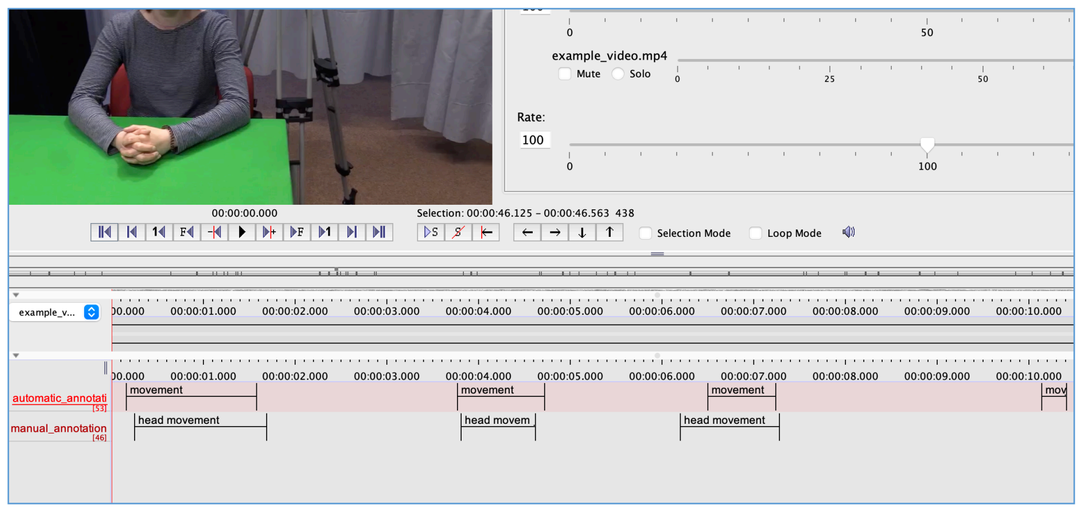
Work 2: Data processing

- We use the software OpenFace to create a spatial model of the head and compute the features velocity, acceleration and jerk
- An 18-dimensional feature vector at each point of time as our training data

Work 3: Algorithm design

- Shannon entropy reflects the variance of a probability distribution function
- Low entropy values mean a large variance. An interval with a large variance is usually the place where head movement is located
- We designed an algorithm which dynamically searches a combination of the sub-sequences whose sum of the entropy can be minimized and consider those head movements

Current results



- The method is a good starting point for unsupervised automatic head movement detection that can be improved upon these points
- There are differences in the position of the boundaries of some annotations
- The automatic method is more sensitive to small movements, as it only looks at the change of information gain value

Future work

- Dive deeper to the task of head movement categories detection, i.e., detecting head nods, head shakes, with an unsupervised method
- Integrate other modalities to discover how participants show engagement and understanding during the interaction
- Develop a dynamical model to predict participants' level of engagement and understanding during interaction

References

- Baltrusaitis, T., Zadeh, A., Lim, Y.C., & Morency, L-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. doi:10.1109/FG.2018.00019
- Deldari, S., Smith, D.V., Sadri, A., & Salim, F. (2020). ESPRESSO: Entropy and ShaPe awaRe timE-Series SegmentatiOn for Processing Heterogeneous Sensor Data. doi:10.1145/3411832
- Jongejan, B. (2016). Classifying head movements in video-recorded conversations based on movement velocity, acceleration and jerk. Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016), 10–17.
- Morency, L-P., Quattoni, A., & Darrell, T. (2007). Latent-dynamic discriminative models for continuous gesture recognition. doi:10.1109/CVPR.2007.383299
- Paggio, P., et al. (2020). Automatic detection and classification of head movements in face-to-face conversations. Proceedings of ONION 2020, 15–21.
- Sadri, A., Ren, Y., & Salim, F. (2017). Information gain-based metric for recognizing transitions in human activities. doi:10.1016/j.pmcj.2017.01.003
- Lücking, A., Ptock, S., & Bergmann, K. (2012). Assessing agreement on segmentations by means of Staccato, the segmentation agreement calculator according to Thomann. doi:10.1007/978-3-642-34182-3_12

Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

TRR 318 CONSTRUCTING EXPLAINABILITY

UNIVERSITÄT BIELEFELD

LAB