Estimating intentions to speak using lexical information

TU Delft

1. Background

- Artificial intelligence is seeing a rise in its use in social settings
  - Chairperson
  - Virtual assistant
- To perform well in social settings, the AI should be capable of understanding and estimating social dynamics such as intentions to speak
- With a good understanding of when a person intends to speak, the AI will know when to give a person room to talk
- 2023 Research Paper "Inferring intentions to speak using accelerometer data in-the-wild." [1]
- How can the lexical information be used with regards to estimating intentions to speak?

2. Research Question

- Is a machine learning approach trained on lexical information viable to estimate speaking intentions?
- Is a rule-based approach viable to estimate speaking intentions?
- Which is better? And why?

3. Method and process

- Gather relevant literature and resources
- Annotate Rewind Dataset
  - Dutch social gathering
  - Annotate unrealized intentions to speak
  - 77.4% of cues involved a filler word
- Generate samples for realized intentions
  - Positive case is speech
  - Negative cases are backchannels
- Train model by extracting lexical information from samples
  - Audio transcription [2]
  - Word embedding
- Create a rule-based approach of estimating intention to speak
  - Dictionary of keywords that indicate speaking intention
  - Filter samples for keywords

Actually... But... However...

Umm... Well... Like...

4. Evaluation and discussion

- The machine learning model seemed to achieve its best performance when testing on failed intentions to continue speaking (0.73 AUC)
- The Rule-Based approach had unremarkable performance resembling that of a random guessing baseline
- ML model looks more promising

5. Future work

- Consider multiple modalities at once, for example non-verbal audio cues as well as lexical information
- Further research on the correct set of keywords

6. Conclusion

- This research provides key insights on predicting intention to speak using lexical information
- ML model had especially good results when estimating unsuccessful intentions to continue speaking
- Could lead to real-world use-cases

References

[1] Litian Li et al. Inferring intentions to speak using accelerometer data in-the-wild. TU Delft, 2023.

[2] Alec Radford et al, Robust speech recognition via large-scale weak supervision, 2022.

Author: Ferhan Yildiz
Email: f.yildiz@student.tudelft.nl

CSE3000
Research Project

Supervisor: Hayley Hung
Email: H.Hung@tudelft.nl