Bachelor's Project: The Plan

Dominik Klepl

Research Question (RQ)

1. Does interpersonal coordination of hand movements provide more predictive power towards the diagnosis of schizophrenia than information from single interlocutor?
2. Is one of the approaches really superior in informing the model about the diagnosis in the sense that using information from the other approach would not bring any new information to the model?
3. What are the patterns that distinguish the classes (patient vs. control)?

Hypothesis

1. YES
    a. Coordination is what underlies the social function therefore it contains more information than individual signal which is only one component of an interpersonal system. Therefore, it does not matter whose signal's power is compared to coordination's power, predictive power of coordination is always higher.
2. Combination of features from both individuals and their coordination will result in the best predictive model
    a. Synergetic/alignment model predicts that information from interpersonal coordination is crucial
    b. Classic psycholinguistics predicts that information from individual is crucial
    c. Both approaches are supported by some evidence but not much research combines both. I think that features combined from both will yield best predictions as both systems contain different information about the interaction and therefore diagnosis as well. In other words, they are 2 individual system with very little to no overlap.
3. It might be better to inspect the patterns post-hoc from the final model

Theoretical point of departure

- Based on the synergetic model (Fusaroli, Rączaszek-Leonardi, & Tylén, 2014).
- Assuming that analysis of interpersonal coordination is crucial (and possibly necessary) to understanding of the interaction (including the goal) => In this case a decision whether the interviewee is a schizo => the decision can be made also from coordination of the signals
- The hand movement of individual (system) contains information about the diagnosis BUT information from relating 2 signals to each other is superior in terms of predictive power

Methods

Data

Information about the participants is reported in (Simonsen et al., 2018). I should rephrase the info from that paper for my final paper

Data preprocessing

- Temporal alignment of actigraph and turn-taking data is realized using psychiatrist clapping 3-times at the beginning of the interview. Therefore, the claps were localized in both datafiles: audio recording from which the turn-taking was automatically derived, and data recorded by wrist movement sensor, actigraph.
    - Claps in audio files were found using an algorithm that was trained to find sudden bursts of acoustic energy in the signal. The algorithm was pretrained for separating syllables

from audio signal as part of R package (find the name) but was found to perform well on our data.
    - Claps in actigraph signal were found using algorithm that searched for a typical shape of the spikes in the data that represented the claps. The top of the spike was defined as point whose neighboring point from the left has less than 80 % of its value and 5 points from the right have lower value than the top of the spike.
- The parts of the actigraph signal where the wearer does not speak will be removed. It is assumed that during listening one does not gesture very much, movement would be rather accounted to i.e. scratching.
- Split data into training and validation datasets – 4:1 ratio – never use validation data for training!!

Feature engineering

All features will be extracted from signals from dominant hand (DH) of both interlocutors. To increase the amount of data, every datafile will be split into several time-windows. Each window is one utterance i.e. one question.

One open question: the ML project is comparing interpersonal system and 2 individual systems – should all systems have the same number of features?

- Statistical – descriptive stats, the exact features yet to be decided
    - Mean, median, min, max, etc.
    - Each of these features will be calculated 3 times because 3 systems are analyzed
        - Interviewer
        - Interviewee
        - Signals of interviewer and interviewee merged into 1 timeseries
    - For the sake of finding powerful features the 2 signals could also be summed, subtracted, etc
        - It could be argued that if the interlocutors form a synergy and the schizophrenic behaves strangely the interviewer might be trying to compensate => difference between signals would be bigger for schizo-doctor pair than control-doctor.
- RQA
    - Windowed RQA
        - 1 utterance-long windows for individuals (DH patient vs. DH patient)
        - 2 subsequent-utterances-long windows for interpersonal coordination (i.e. 1 question + 1 answer)
- Time feature – sequence stating the order of windows
    - Should have no predictive power by itself, however might interact with other features
    - Also, for statistical models it might be used as random slope

Machine learning (ML) for 1st hypothesis testing

1. Select several algorithms trained only on features either from interlocutor 1, 2 or from their coordination => several models
2. Use repeated k-fold cross-validation (CV) (look out for participants leaking into different folds!!!) to select the best performing model for each category => ending with 3 models
3. Construct a data with outputs of the CV of the 3 final models. Do linear regression with this data: $Performance\ measure \sim \beta * model + \varepsilon$ with the coordination model as baseline (intercept) so that the slopes of the 2 effects give back a measure of how their performance differs from the coordination model => results support hypothesis if coordination model outperforms both individual models.

ML for 2nd hypothesis testing and getting the best classifier

1. Keep the best model from 1st ML session
2. Train a model on all features extracted from the data and run repeated CV.
3. Compare performance of the two models again with linear regression.

4. Do hyperparameter tuning for the best model if possible (if logistic regression is the best algorithm than obviously skip this)
5. Train the model on all training data and get final performance measure on the validation data

Interpreting the model's behavior

It is very likely that the best performing model will be a black-box model that is not just exploiting the average of all training samples but exploits the individual samples to get non-uniform prediction.

- 2 options
    - Plot the data as if doing exploratory analysis – boxplots, interactions of features and report everything that shows a significant difference between controls and patients
        - Not my favorite option: lot of work, not really showing what the model's doing and not so cool as option 2 BUT it's easy, easy to produce nice plots, easy to interpret and understand by reader
    - Use specialized ML tools for insights – permutation importance, partial dependence plots or SHAP values (and aggregating them to get model-level insights rather than just prediction-level)

References

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32*, 147-157. doi:https://doi.org/10.1016/j.newideapsych.2013.03.005

Simonsen, A., Fusaroli, R., Skewes, J. C., Roepstorff, A., Campbell-Meiklejohn, D., Mors, O., & Bliksted, V. J. S. b. (2018). Enhanced automatic action imitation and intact imitation-inhibition in schizophrenia.