Empathy assessment from eye fixations

January 13, 2023

1 Project Description

You're a data scientist in a start-up that wants to sell empathy assessments to other businesses which can be used during recruitment by HR. The goal is to use an eye-tracker while people who are being interviewed perform some tasks, and to come up with an assessment of their empathy based on features from the eye-tracker only.

To assess the viability of this, the CSO organised a partnership with a university to collect some data in a trial (including the traces from the eye tracker for multiple participants and tasks and their answers to a questionnaire from which empathy can be calculated as your target) [1]. As the only data scientist in the start-up, you have been given access to the data set and asked to investigate whether an assessment of empathy is possible from the eye tracker data.

You need to show that you understand the signals that are being recorded, how to properly process them and find out whether and how they can be used to determine an empathy score for HR so that you can make a recommendation to the company and present your results **concisely**, paying special attention to how the algorithm is making the decision and which signal/s can and cannot be used for this task.

Due to GDPR regulations, your model should be explainable, i.e., it should be possible to tell a person why their score is high/low if this score was used by HR to not hire them.

Read the following tasks in detail and make sure you understand the project.

2 Tasks

2.1 Assignment 1

1. Load and explore the data set. Ensure that you leave a subset of data separate from the exploration to avoid overfitting. **You cannot use the code given by the authors of the paper for this data set and present it as your own.**

2. Clean and preprocess the data set (including feature extraction if appropriate).

3. From your exploration only, find out which features might be better candidates for the task of predicting an empathy score. You can run some early modelling tests, but this is not strictly necessary for this assignment. During your demo, we will check that you understand the different signals and have a plan for how to analyse the data for Assignment 2 that is reasonable and feasible.

2.2 Assignment 2

1. The modelling is up to you, as long as it follows good practice in data science (i.e., using cross-validation and train/test splits properly and as appropriate for the problem and the chosen modelling approach). You are encouraged to check out the available literature and find out what has worked in the past, but you need to present something new/different as part of your project. **You cannot use the code given by the authors of the paper for this data set and present it as your own.**

2. While you're modelling, reflect on the project description given above and make sure you pay special attention to the requirements set out by your manager. The future of the start-up depends on you!

3. For your report, make sure that all your decisions are justified and that you present your findings clearly and concisely following the template given to you on Moodle. As a data scientist, you must reflect on your results. Talk about your findings and what they indicate. Finally, you must answer the question: how good is your system at determining a person's empathy? Which guarantees can your model offer to an HR department that is considering hiring the services of the start-up? Do you have any other insights from the data that can help the start-up in the future (e.g., for future features that can be commercialised)?

References

[1] P. Lencastre, S. Bhurtel, A. Yazidi, S. Denysov, P. G. Lind, et al. Eyet4empathy: Dataset of foraging for visual information, gaze typing and empathy assessment. *Scientific Data*, 9(1):1–8, 2022.