



Hi Francisco,

Thank you for helping me with my thesis.

I just have a few quick questions. I want to know what you thought about the survey, rubric, and method we used to evaluate your explanations.

I would also like to know if you felt that the survey results made it easier to improve your explanations

Regards,

Charlotte

Section A: SurveyRubricMethod

A1. How did you feel about the survey?

The survey was the combination of the rubric questions, demographic questions, and the comprehension questions we asked the MTurk participants to answer.

Overall, I feel the survey was successfully performed. The data collected allowed us to better understand what kind of explanations the participants favoured but also how the explanations were presented to the participants in terms of quality. Moreover, the demographic questions gave us a general view about the profile of the participants showing variability in most of the aspects maybe with the exception of the country where they lived.

A2. How did you feel about the rubric?

The rubric was part of the survey. It consisted of 10 multiple choice questions about the quality of the explanation.

As mentioned, the rubric questions were quite useful to assess the explanations in terms of their quality. Most of these questions were very useful in order to improve each explanation in the following step, therefore, this is a crucial part of the survey in order to create better quality explanations. Particularly, the question "can the explanation be acted upon", was unclear to me in terms of how to interpret both the question and the obtained results. My guess is that this question may be more useful for other contexts or other kinds of explanations.

**A3. How did you feel about the method we followed?**

The method included the survey, but also the process of using MTurk and creating comprehension and demographic questions.

I think the method is the right one to use. There are different opinions about the use of MTurk in terms of the quality of the answer provided, but on the other hand, it allows to collect enough data to process possible outliers next. The comprehension and demographic questions, in my opinion, allow establishing a general framework to be applied to most (or any) of explanations used in machine learning contexts.

Section B: Other**B1. Did the survey help you improve your explanation?**

If yes, was it easier to use the survey to improve the explanation than trying to improve it by yourself?

Yes, definitively. After reviewing the answers provided by the participants, new (and hopefully better) explanations were created. Indeed, using the survey to improve the explanations instead of trying by myself was a better approach since allowed me to identify gaps that would be very difficult to find. For instance the use of technical terms in the explanations. There were a few terms I assumed were well-known, however, they were still considered too technical by some participants.

B2. Do you have anything else to say about the experiment or your experiences of using the survey, rubric, or method to evaluate your explanation?

I just want to say I am happy to have been involved in this project. It is a very interesting rubric to be used in explainability. Count on me for any possible follow up.

If you have any other comments please let me know,

Thank you for filling out this survey

Regards,

Charlotte.