Automating Response Evaluation: A Tool for Automatic Coding Responses to Open-ended Questions

Shih-Chieh Dai (MSIS), **Field Supervisor**: Aiping Xiong (Pennsylvania State University)

BACKGROUND

The open-ended question is a widely used method in human-subject surveys. Open-ended questions provide opportunities for discovering human subjects' spontaneous responses to the questions. However, it is labor-intensive and time-consuming to analyze the responses.

In light of the challenge, I aim to leverage NLP methods to create a tool for assisting the analysis of open-ended questions.

WORK SUMMARY

To design a tool that can help code the responses, I first review the literature and existing work to learn the limitations and needs. Secondly, I discussed the issue with my field supervisor for the workflow. Finally, I implemented the coder generation module with the transformer model and the code classification model with GPT.
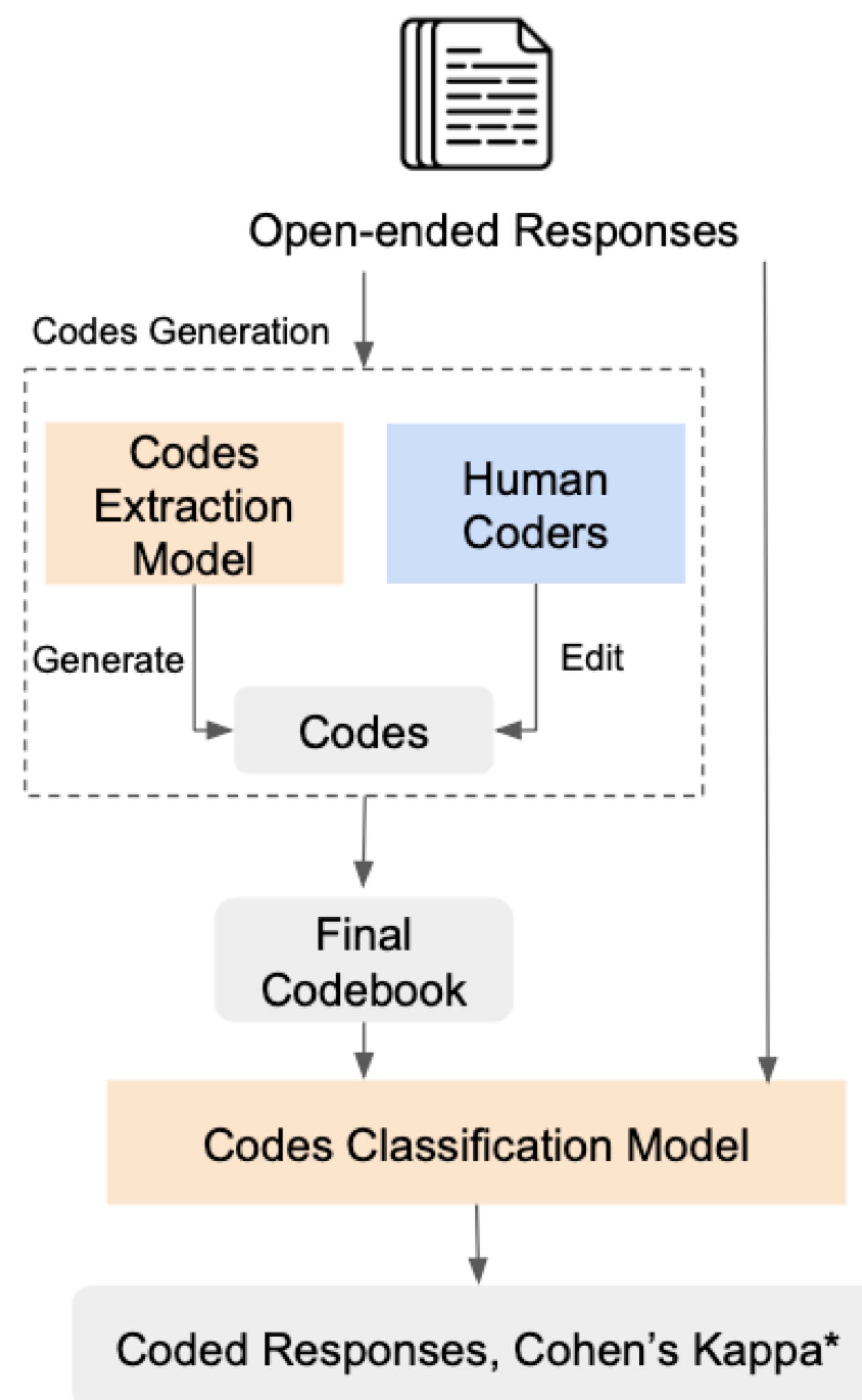
TECH / SOFTWARE

Python   PyTorch   Hugging Face

WORKFLOW

*If the user provided his/her coded result

LIMITATIONS

The result may be biased
Since the user can edit the codes, the codes will inherently lean on the editor's perspective. Therefore, the result may be biased.

Non-experience annotator
If the user is not familiar with the coding process, the user might rely on the result of the tool. As a result, the coding quality might be pooled if the model did not perform well.

TAKEAWAY

- We developed a tool for thematic coding.
- Future work on debias needs to be explored.

Final Deliverables

https://github.com/sjdai/open-ended-question-eval

References

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Sage.

Liew et al., (2014). Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis. ACL.

The University of Texas at Austin
School of Information