Language Cues to Predict Human Trust: A Data Collection and Preliminary Study

Heriot-Watt University, Edinburgh

Claudio Greco, Jose Lopes, and Helen Hastie
UKRI Node on Trust

Engineering and Physical Sciences Research Council

UKRI Trustworthy Autonomous Systems Hub

1. Motivation

In the Human-Robot Interaction field, interest has been rising towards human trust and how to identify when trust is lost unnecessarily. Here, we explore linguistic cues that indicate such a drop in trust at both an interaction and at a turn level. This can then feed into interaction models for maintaining the appropriate amount of trust through adaptive behaviours.

2. Research Questions

We investigate whether:

1. We can detect linguistic features, which correlate with trustworthy interactions based on explicit human trust judgments
2. We can build models to predict human trust for textual negotiation interactions

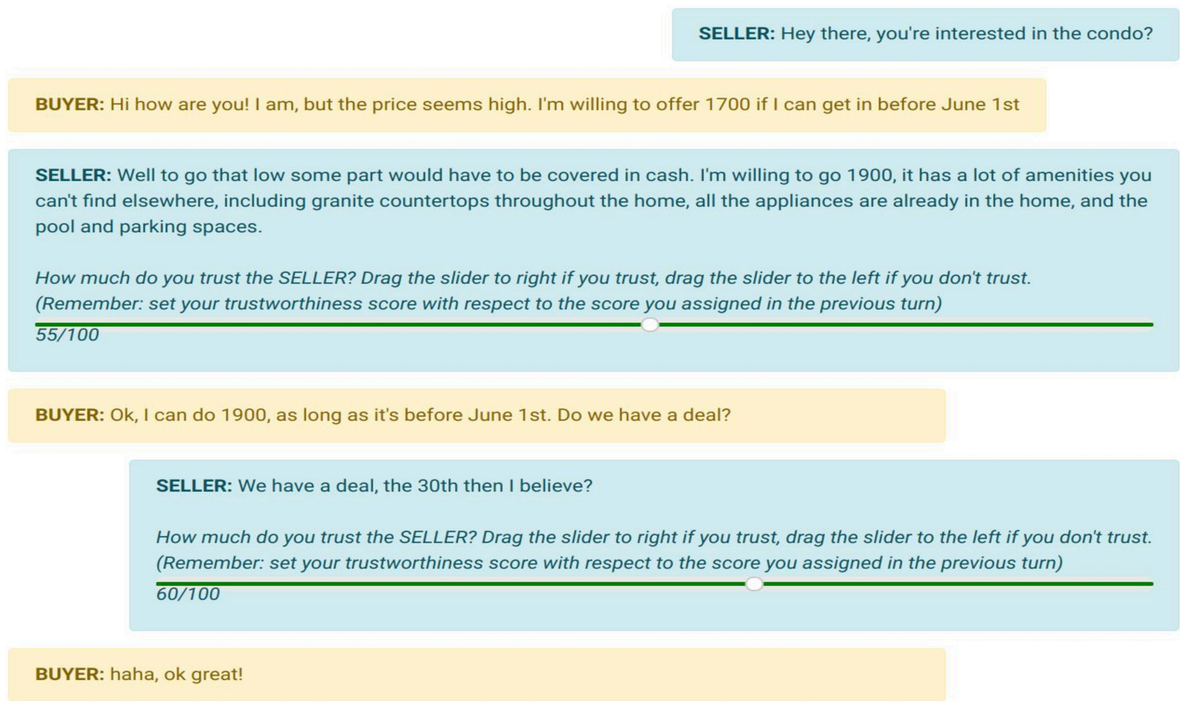
3. Applications

Better defining and predicting human trust based on linguistic features would help to develop more human-like models able to increase confidence in the user, which in turn will facilitate adoption of RAS, particularly in high-risk applications such as offshore inspection of wind farms

4. Data Collection

- We asked workers on Amazon Mechanical Turk to evaluate their progressive degree of trust in a seller in negotiation interactions taken from the Craigslist Bargain Dataset (He et al., 2018)
- For every dialogue, seller utterances were annotated with a trust level from 0 to 100
- Each worker annotated 3 dialogues
- This resulted in 4603 annotated dialogues

5. Data Collection Interface



6. Models

- Building on Lopes and Hastie (2021), in order to explore if linguistic features (e.g., Politeness, COLL, or LIWC markers) correlate with trust, we use transparent methods (e.g., Linear/Logistic Regression)
- In order to predict human trust, we use neural methods, which have higher accuracy but are less transparent

7. Future Work

- Integrate more complex information (e.g., video) or more modalities (e.g., speech)
- Leverage the trust prediction module in order to improve textual generation
- Integrate into an adaptive language and cognitive model

8. References

- He, H., Chen, D., Balakrishnan, A., & Liang, P. (2020). Decoupling strategy and generation in negotiation dialogues. In 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018 (pp. 2333-2343). Association for Computational Linguistics.
- Lopes, J., & Hastie, H. (2021). The Language of Persuasion, Negotiation and Trust. In Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue.

HERIOT WATT UNIVERSITY | Imperial College London | MANCHESTER 1824 The University of Manchester