CompSci_colour.pdf

Investigating Computational Responsibility

William Wallis (2025138)

April 7, 2017

ABSTRACT

Currently, models are produced for responsibility modelling which have their roots in logic. These models, while sophisticated, suffer from a lack of pragmatism: for guiding agent behaviour in sociotechnical simulations, logical models are not always ideal. In the similar field of trust modelling, algorithmic models which emulate social behaviour produce useful results while being easier to understand, implement, and reason about. In this paper , a proof-of-concept responsibility modelling platform adopting the algorithmic formalism style employed by trust modelling is produced, and its utility evaluated.

paper? report? project?

1. INTRODUCTION

A growing area of research lies in the formalism of human traits into computational representations. These algorithms make computers more human-like; for that reason, they are referred to here as "Anthropomorphic Algorithms" . A similar term, "human-like computing", has also risen in popularity lately. Human-like computing does not strictly focus on the implementation of formalisms of human traits, however, which is the area of interest for this report.

improve the introduction of the Anthropomorphic Algorithms term

This implementation interest, realised in the study of anthropomorphic algorithms, presents an interesting sociotechnical problem. They present an opportunity to alter the behaviour of actors in a sociotechnical system, and to do so in a way that is easy to reason about. This alternation of behaviour is done by the algorithmic implementation of a *formalism*. Formalisms present a concrete definition — by process, mathematical definition or semantic description — which can be used to construct an anthropomorphic algorithm. These formalisms tend to attempt to model in one of two ways:

1. Modelling the trait as a useful metaphor
   These models tend to be inaccurate with regards the social science surrounding the trait that they model. However, they make a trade-off between this accuracy and the model's utility. For example, the notion of trust as a metaphor for a type of behaviour might be useful in information security research, but what matters in the formalisms implemented for this research is the formalism's utility in information security — *not* whether the formalism accurately represents human trust.

2. Modelling social science directly
   These models attempt to accurately model the traits they concern. This can be useful for fields such as sociotechnical modelling, as well as social sciences research. There are also interesting applications for these models in interaction study: making interfaces interact with users in a human-like way, and representing the states of these traits to the users, are valuable research areas which are more applicable to these type-2 formalisms than to type-1 formalisms.

In reality, most formalisms and their implementations lie somewhere on the spectrum that these two types define.

2. STATEMENT OF PROBLEM

Computational formalisms of human traits are a growing field of research, with applications in lots of different areas. A problem with these anthropomorphic algorithms is that there is limited breadth to the scope of existing research in the field (as is demonstrated during the background survey in section 3). The metaphor of the human trait in these algorithms remains largely unexplored.

Breadth in the application of the metaphor is important, however. The importance stems from the utility in the human metaphor when designing systems:

- Human-Computer Interaction can make use of behavioural metaphors to relay complicated internal states to a user. Storer et al. [2] demonstrated methods by which a mobile device might dissuade certain user actions by expressing its "discomfort" or lack of "trust" in its interaction design.

- Information Security can make use of behavioural metaphors in order to increase difficulty of access when negative system states are encountered. A system might allow access on a graded scale, dependant on internal states of trust, comfort, and confidence.

- Theoretical advancements in smart city technology[1] might increase a city's resilience by integrating notions of responsibility into public services and the environment on a community scale.

While similar results can often be achieved using regular techniques, the human metaphor allows for a better communication between a human user and complicated system states. All of the above examples enter around this notion; however, the applications extend beyond Human-Computer Interaction research.

3. RELATED WORK

3.1 Trust Modelling

3.1.1 Castelfranchi & Falcone

3.1.2 Marsh

3.1.3 Eigentrust

3.2 Sociotechnical Systems

3.2.1 FIRE

3.2.2 Ian Sommerville

3.3 Philosophical Literature

3.3.1 P.F. Strawson

3.3.2 Thomas Scanlon

3.3.3 Sloman

4. A FORMALISM OF RESPONSIBILITY

4.1 Design Decisions and Philosophy

4.2 Feature 1: Acting on Responsibilities

4.3 Feature 2: Judgement of Responsibility

5. EVALUATING THE FORMALISM

6. FUTURE WORK

6.1 Advancing the formalism

7. DISCUSSION

8. REFERENCES

[1] author names. talk title. conference name, year.

[2] V. T. Tim Storer, Karen Renauld. Find this find this. 2020.