



A Persuasive Virtual Chat Agent Based on Sociolinguistic Theories of Influence

Samira Shaikh (UNC Charlotte; sshaikh2@uncc.edu)

DOI: [10.1145/3098888.3098899](https://doi.org/10.1145/3098888.3098899)

Problem Statement

Machine intelligence breaks new ground with each passing day, achieving feats like winning against humans in the game of Go and Jeopardy!. However, the goal of implementing social competence in machines so that they may achieve specific social goals remains elusive. Persuasion, the ability to tailor messaging to the target audience is closely related to social competence (Gass and Seiter, 2015). In using technology to persuade, we delegate a central function of human communication to a computer. This central function is intentional and goal-directed communication - deliberate, purposeful messaging with a clear outcome in mind. Our goal in this thesis is to automate the very process of persuasive communication, by designing a system which can purposefully communicate, without any restrictions on domain or genre or task, and which has the clear intention of persuading the recipients of its messaging.

We investigate how models of social phenomena - specifically persuasion strategies - may be automated in an artificial autonomous agent, in the form of two overarching research questions:

- Can specific persuasive strategies be automated in a virtual chat agent?
- Can active persuasion by individuals during conversation be detected and counteracted by such an agent?.

Our goal is not to create an artificial agent capable of passing the Turing test or the Loebner prize. Rather, our goal is to define specific human persuasive strategies that can be programmed into an agent who can then persuade participants to its own view.

Methodology

The foundation of our persuasive strategies comes from the summative model of attitude (a well-established model of attitude in social psychology), where belief change leads to attitude change, and, ultimately, behavior change (Fishbein and Ajzen, 2011).

We organized the work in three phases. First, we conducted a belief elicitation study to obtain salient beliefs on a variety of social issues (for example, *Should the minimum legal drinking age be lowered from 21 to 18?*). We had two distinct reasons to undertake such a study. The first motivation was to use the beliefs so elicited to design survey instruments. These surveys are crucial to our work, because the responses on the surveys provide the ground truth from which to measure the agent's efficacy in being persuasive in controlled experiments. Our second motivation was to use the corpus of beliefs elicited as a database of natural language statements for the agent to use. The goal of this thesis was not to undertake the task of natural language generation for the agent. Our solution is to use the responses made by humans on the topics as a database of arguments made in favor of and against the topic of discussion. We pre-programmed these arguments in the agent, with the intention of using them at *opportune* moments in the conversation (as explained below).

Next, we programmed behaviors and strategies in the agent that were aimed at persuading individuals through online conversation as well as counteracting persuasion by the participants. The behaviors programmed in the agents are triggered, in part, by a variety of linguistic cues emerging from the conversation, such as dialogue acts, topic, polarity and communication acts. The annotated context of conversation is used to inform the agent's models by updating the underlying beliefs of participants in real time. It is necessary for the agent to create and maintain a represen-

tation of the mental states of the participants with respect to the topic so as to understand their viewpoints. In this work, the mental state or point of view is the overall attitude towards the topic, and its essential pieces are the belief strength (how strongly a belief is held) and belief evaluation (is the evaluation towards the belief positive or negative) on the beliefs related to the topic of discussion. The agent uses the belief models and annotated utterances and selects appropriate behaviors to perform from a list of pre-determined behaviors (operationalized as dialogue acts e.g. assertions, agreements, disagreements, etc).

In the third phase, we ran controlled experiments. The aim of these experiments was to deploy the agent and validate our persuasion and counter-persuasion strategies in online synchronous conversation environments. We selected the majority-minority influence setting in our experiment design. In social influence research, it has been shown that minorities influence people's thinking, attitudes, and behavior by being consistent in their views and flexible in their negotiation with majority members' (Gardikiotis, 2011). Each chat session in our experiments consisted of four participants, two majority opinion holders, one minority opinion holder and the agent (either wizard or computer agent, who advocated the extreme minority opinion). Procedures for both the Woz and autonomous agent experiments were nearly identical. The only difference was the presence of the wizard or the algorithm acting autonomously. Our main research hypothesis is that the interventions made by the agent would result in attitude change in the participants. Pre- and post-discussion surveys allow us to measure changes in participants belief models, and thus, the shifts in their overall attitude towards the topic of discussion. The system goal in our experiments is multifaceted, attempting to change the belief strength and belief evaluation, and furthermore, the overall attitude of the participants. Accordingly, we used the Wilcoxon signed-rank test (Wilcoxon, 1945) to measure persuasion effectiveness of the algorithm. We observed statistically significant changes in overall attitude from pre- and post-discussion surveys across 10 discussion sessions with 30 participants in total using the Wilcoxon signed rank test ($Wilcoxon T=12, p=0.03$).

Contributions of this Thesis

Upon placing an agent in the midst of conversations, it is able to discern beliefs that are expressed by the participants in the group, and use them to ascertain participant's opinions on topics of discussion. Using this information and drawing upon theories of influence and persuasion from social psychology, cognitive science and communication, the agent aligns participants towards or against a particular issue. We validated that the agent achieved statistically significant changes in the participant's attitudes, thus demonstrating its effectiveness in being persuasive. In doing so, this work makes contributions to the field of AI and human-computer dialogue.

Acknowledgments

This research was supported by a grant from Army Research Laboratory.

References

- Fishbein, M. and Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis.
- Gardikiotis, A. (2011). Minority influence. *Social and Personality Psychology Compass*, 5(9):679–693.
- Gass, R. H. and Seiter, J. S. (2015). *Persuasion: Social influence and compliance gaining*. Routledge.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83.



Samira Shaikh works on computational sociolinguistics, artificial intelligence and natural language processing. She obtained her PhD from the State University of New York - University at Albany in Computer Science in July 2016. She is currently an Assistant Professor of Cognitive Science in the Department of Computer Science department at University of North Carolina - Charlotte.