Persuasion by Shaping Beliefs about Multidimensional Features of a Thing

Extended Abstract

Kazunori Terada
Gifu University
Gifu, Japan
kazunori.terada@acm.org

Yasuo Noma
Gifu University
Gifu, Japan
yasuo.noma@ai.info.gifu-u.ac.jp

Masanori Hattori
Chubu Electric Power Co.,Inc.
Nagoya, Japan
Hattori.Masanori3@chuden.co.jp

ABSTRACT

Research has demonstrated the effectiveness of personalization in persuasive agents, recommendation agents, and nudge agents. Ultimate personalization targets the presentation of information tailored to an individual's nuanced beliefs and utilities, rather than relying on broad attributes such as personality traits, age, or gender. Multi-attribute utility theory suggests that the utility of a thing is determined by the sum of the utilities given to its various features. In our research, we developed a method to enhance the personal utility of a thing by addressing and manipulating people's beliefs about the features of a thing. We conducted an experiment (n=197) to verify whether the proposed method can increase the participants' utility of a fully autonomous vehicle, as a target of persuasion. Among 13 propositions (features) that constitute the concept of fully autonomous vehicles, in a semi-structured dialog, a virtual agent presented counter-propositions to the top propositions that each participant assigned the most negative utilities. Before and after the dialog, the monetary value of fully autonomous vehicles, the desire to ride them, and the social obligation to accept them were measured. The results showed that the proposed method improved the social obligation to accept fully autonomous vehicles more than the baseline method and the non-personalized method, but had no effect on the monetary value and the desire to ride them. This suggests that personalized belief manipulation may not be effective in enhancing the "want to" desire or utility of a thing, but may only improve the thought of "ought to do".

KEYWORDS

Persuasive agent; Personal value; Personalized information presentation; Belief manipulation

ACM Reference Format:
Kazunori Terada, Yasuo Noma, and Masanori Hattori. 2024. Persuasion by Shaping Beliefs about Multidimensional Features of a Thing: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

Recommendation agents, persuasive agents, and nudge agents have been designed to alter human decision-making through information presentation. In these agents, personalization of the presentation of information is essential to increase the effects on human decisions. Personalization sometimes utilize demographic profiles such as individual preferences, age, gender, and income [3], personal character tendencies [10], and user history [2, 8]. The common assumption of these personalization is that people with similar attributes or tendencies will make similar decisions. Extending this idea, the more detailed the personalization, the greater the potential influence on decision-making. Therefore, in this study, we propose a method for presenting more detailed personalized information in persuasion, using fully autonomous vehicles as the target of persuasion. We acquired individual value on multi-dimensional features (propositions) related to fully autonomous vehicles in advance, and in dialogues with an agent, the agent selectively presented positive counter-propositions for features individual highly concerned. We validated the proposed method in an experiment with human participants.

Enhanced personalization in information presentation requires intricate individual modeling across multiple dimensions. The efficacy of multidimensional feedback that mirrors individual activities [9], as well as explanations tailored to the user's mental state (including beliefs and goals) [1], has been demonstrated. These studies show the effectiveness of explaining the target concept for persuasion from different aspects. Although multidimensional information presentation has not been well studied in persuasion, multidimensional information presentation is used in recommendation [4, 6, 7].

Things are composed of numerous features. The final evaluation of thing is determined as the sum of the values attributed to these features [5]. Suppose that a concept of a thing is composed of an N-dimensional feature vector, where each dimension represents a proposition expressing a characteristic that composes the thing. For example, let $A \rightarrow B$ denote a certain dimension. An individual's belief about a particular thing can be expressed as follows.

$$Bel = \{\mathbf{x} \in \mathbb{R}^N \mid -1 \leq x \leq 1\} \tag{1}$$

, where $\mathbf{x}$ denotes the degree of belief in each proposition. $x_{A\rightarrow B}$ is the degree of belief that $A \rightarrow B$ is true for positive values, and $A \rightarrow \neg B$ is true for negative values. Zero implies the irrelevance of A and B. The utility u is expressed by the inner product of the value (weight)

$$Val = \{\mathbf{w} \in \mathbb{R}^N \mid -\infty \leq w \leq +\infty\} \tag{2}$$

of each proposition constituting the thing and *Bel* as follows.

$$u = \mathbf{w} \cdot \mathbf{x} \tag{3}$$

Consider, for example, the thing "self-driving car" represented by a two-dimensional proposition (feature) $[self\text{-}driving\_car \Rightarrow employment\_decline, self\text{-}driving\_car \Rightarrow traffic\_efficiency]$. For individuals who believe that the causality between the self-driving car and employment declines as true and netagively evaluate, and believe that the traffic efficiency caused by the self-driving car is true and positively evaluate, the utility of the self-driving car is calculated as $u = [-1, 1] \cdot [1, 1] = 0$, implying that the two utilities cancel each other out, resulting in zero.

2 METHODS

Figure 1: Interface for persuasion.

We conducted a study to determine whether the attitudes of participants towards fully autonomous vehicles would change after being exposed to tailored positive counter-propositions, negating the perceived negative aspects of these vehicles, through semi-structured persuasive dialogue with an agent. First, we conducted manual text mining of 540 sentences concerning perceptions of fully autonomous vehicles, from 100 participants to obtain 66 features (propositions) that constitute the concept of fully autonomous vehicles. Subsequently, through a secondary experiment involving 94 participants, we selected propositions that were associated with the negative perception of fully autonomous vehicles. Of the 66 features, the features such as "Reducing traffic accidents" was not selected because many people evaluated it positively. On the other hand, the features such as x_1 = "Automobile accident fakers who take advantage of malfunctions of AI algorithms emerge." and x_2 = "Murder and terrorism due to system hacking" were selected because most people valued it negatively. As a result, 13-demensional features $\mathbf{x}$ were obtained. These were then considered as candidates for the target propositions during the persuasion experiment.

For the main experiment, 197 participants were recruited via crowdsourcing. Initially, participants were asked to evaluate the values of the 13 propositions. Following this, participants were engaged in semi-structured persuasive dialogue with an agent (see Figure 1) on one of three topics: personalized, non-personalized, and baseline. In the personalized topic, participants were presented with five to ten positive counter-propositions that negated their *top* five negatively evaluated propositions out of the 13 propositions. Counter-propositions were such as "Everyone now has access to dash cams for recording, so I don't see a problem." for x_1 and "Even with today's computers, if appropriate measures are taken, hacking can be prevented, and I believe that these measures will continue to improve in the future" for x_2. In the non-personalized topic, participants were presented with five to ten positive counter-propositions that negated their *bottom* five negatively evaluated propositions out of the 13 propositions. In the baseline topic, propositions about hobbies, food, reading, and sports, which are unrelated to fully autonomous vehicles, were presented. The interface used in the experiment is shown in Figure 1. We used an anime-style female agent with large eyes included in the Live2D Cubism software from Live2D Inc. We quantified changes in participants' attitudes by measuring their perceived monetary value of fully autonomous vehicles, their desire to utilize them, and the sense of societal obligation to accept them, both before and after the dialogue.

3 RESULTS AND DISCUSSION

The experimental results showed that the presentation of positive counter-propositions to the propositions that the participants were more concerned about increased the social acceptability of fully autonomous vehicles, suggesting that the proposed method of persuasion may be effective to improve the thought of "ought to do". However, there was no change in monetary value before and after the dialogue, furthermore, the desire to drive increased after the dialogue, regardless of the personalization of the dialogue and the content of the topic, suggesting that personalized belief manipulation may not be effective in enhancing the "want to" desire or utility of a thing. These suggest that there are limitations and room for improvement in the proposed persuasion method.

The goal of this study was to change people's attitudes and decision-making through interaction with agents. However, there is a concern that the change in attitudes and decision-making is unintended and unwanted by the person. To address this issue, guidelines outlining informed consent procedures are needed to help people understand the potential impact of their interactions with AI agents on their attitudes and disicion-making.

The ultimate personalization in persuasion and recommendation involves shedding light on the multi-dimensional feature that constitute a person's beliefs and values in detail, and offering pinpoint information to enhance these beliefs and values. To our knowledge, our study is the first to experimentally identify multi-dimensional features forming a single concept and to realize an agent that offers tailored information to dispel concerns. We found that presenting counter-propositions to the aspects that people are more concerned about improves the notion of "ought to do", but it does not affect "want to do" or monetary value. Enhancing the persuasive effect is possible by presenting an agent with a higher level of animacy or anthropomorphic appearance, a more knowledgeable presentation, and features that are easier to convert into monetary value. There is also the potential for different persuasive effects to be seen in different persuasive targets such as healthcare.

ACKNOWLEDGMENTS

This work is supported by JSPS KAKENHI Grant Number JP21H03782, JST CREST Grant Number JPMJCR21D4 and JST-Mirai Program Grant Number JPMJMI22J3, Japan.

REFERENCES

[1] Amal Abdulrahman, Deborah Richards, and Ayse Aysin Bilgin. 2022. Exploring the influence of a user-specific explainable virtual advisor on health behaviour change intentions. *Autonomous Agents and Multi-Agent Systems* 36, 1 (apr 2022). https://doi.org/10.1007/s10458-022-09553-x

[2] G. Adomavicius and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, 6 (jun 2005), 734–749. https://doi.org/10.1109/tkde.2005.99

[3] Yae Dai, HongWu Ye, and SongJie Gong. 2009. Personalized Recommendation Algorithm Using User Demography Information. In *2009 Second International Workshop on Knowledge Discovery and Data Mining*. IEEE. https://doi.org/10.1109/wkdd.2009.156

[4] Michael D. Ekstrand. 2011. Collaborative Filtering Recommender Systems. *Foundations and Trends® in Human–Computer Interaction* 4, 2 (2011), 81–173. https://doi.org/10.1561/1100000009

[5] Ralph L. Keeney and Howard Raiffa. 1993. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge University Press. https://doi.org/10.1017/cbo9781139174084

[6] Michael J. Pazzani and Daniel Billsus. 2007. Content-Based Recommendation Systems. In *The Adaptive Web*. Springer Berlin Heidelberg, 325–341. https://doi.org/10.1007/978-3-540-72079-9_10

[7] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative Filtering Recommender Systems. In *The Adaptive Web*. Springer Berlin Heidelberg, 291–324. https://doi.org/10.1007/978-3-540-72079-9_9

[8] Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence* 2009 (oct 2009), 1–19. https://doi.org/10.1155/2009/421425

[9] Max J Western, Dylan Thompson, Oliver J Peacock, and Afroditi Stathi. 2019. The impact of multidimensional physical activity feedback on healthcare practitioners and patients. *BJGP Open* 3, 1 (feb 2019), bjgpopen18X101628. https://doi.org/10.3399/bjgpopen18x101628

[10] Mohan Zalake, Alexandre Gomes de Siqueira, Krishna Vaddiparti, Pavlo Antonenko, and Benjamin Lok. 2021. Towards Understanding How Virtual Human's Verbal Persuasion Strategies Influence User Intentions To Perform Health Behavior. In *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*. ACM. https://doi.org/10.1145/3472306.3478345