Title	エージェント間の連携ネットワークにおける信じられる論証の集合の特定
Author(s)	清水, 颯太郎
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18923
Rights	
Description	Supervisor: Teeradaj Racharak, 先端科学技術研究科, 修士(情報科学)

Abstract

In communication, we each hold various ideologies and assertions, contemplating whether we can ultimately convince others of our own claims. At times, we may ignore or conceal arguments that are disadvantageous to our own claims in order to persuade others to accept them. We analyze a formalization of which arguments are hidden or ignored in order to get someone to accept a certain claim.

This behavior is particularly prevalent in interactive multiplayer games, where actions are often taken to achieve one's victory conditions. One well-known example of such a game is the Werewolf game [1,2], where players are divided into Werewolf and Villager factions. All players engage in discussions, with the Villager faction aiming to identify and eliminate Werewolves from the village, while the Werewolf faction seeks to deceive the Villagers and eliminate them. Among these two factions, the werewolf faction, in particular, often ignores or conceals arguments unfavorable to their claims in order to make them more acceptable. This is because, fundamentally, the werewolf faction finds themselves at a disadvantage in debates. While the villager faction moves closer to victory by honestly sharing the information they possess, the werewolf faction must progress discussions in a distorted manner, such as by telling lies, in order to navigate the situation. Hence, the werewolf faction frequently engages in such behavior.

To address this issue, we formalize the act of ignoring or concealing unfavorable arguments to persuade acceptance using Abstract Argumentation Framework [3]. Then, we propose an implementation according to our proposed formalization and verify through algorithms how this affects multi-agent systems. Our algorithmic procedure can be briefly explained as follows: First, each agent provides arguments through debate, creating a set of arguments. Then, we examine from the debates which arguments are critical of each argument and form a set of relationships based on this. Through an algorithm, arguments are removed based on the set of arguments and their relationships. The sets of arguments and relationships are then reduced. And then, the process involves identifying the set of arguments that have been reduced by eliminating unfavorable arguments to ensure the acceptance of the desired claim.

By identifying the set of arguments being reduced, it becomes possible to understand which arguments and relationships of arguments within the arguments need to be removed in order to gain acceptance of the desired

claim by others.