Abstract

In this research, we are talking about the methods that develop an AI agent for solving the werewolf game just like human beings. Werewolf game is party game at first. That means the information you get from other players is not only the conversation. In werewolf game of the real world, the emotions, the movements(body languages), the sound when you close your eyes and even the tones of voices, these are essential information to win the game. Moreover, the werewolf game is an incomplete information game. The deficiency of the information makes it is difficult to be resolved. And the werewolf game is a social game that players play the game by communication. It makes the werewolf game more difficult than others just like Majiang. However, this is also the charm point of the game that attracts many researchers to study it.

Here we introduce several pieces of research of werewolf agent. The first one, we talk about the AI Wolf Project. This is a group based on several researchers who are interested in the werewolf game. And they developed a platform that could let Agents play werewolf game together. They hold the competition of werewolf agents every year. This platform provides two types of agents to develop. One is the protocol type, and another one is natural language branch type. To consider that communication by natural language is too difficult to achieve, developers who want to concentrate on inference could choose the protocol type. The rules of the protocol provide a limited language option. The natural language type agent uses natural language(Japanese) to communicate with each other. So to reduce the difficulty of this type of agent, the rules of the game is more straightforward than protocol type.Next, we introduce a piece of research that use an extend BDI model to describe the werewolf game. They extended the BDI model by adding probabilistic intentions into the model to solve the problem that the beliefs of the agent are uncertain. Next, there is a piece of research that developed a werewolf game model based on play log from werewolf bbs. So then we introduce a study developed a werewolf game agent based on deep reinforcement learning. As a result, the win rate of the agent is better than the agents exist.

The most common method to make a werewolf agent is using machine learning, which is famous in solving perfect information games. Also, according to the results, machine learning and deep learning are useful to improve the win rate of werewolf agent. However, the data of human beings is not enough to make the agent powerful as human beings. Moreover, the log of the game is too difficult to analyze because it has a large amount of redundant information. Moreover, this research introduces several logic methods and show how to use these methods to solve the werewolf game. The first method is the modal logic. Modal logic is

developed in the 1960s. It is an extending of the first-order logic by adding two symbols, must and may. So the modal logic has several different axioms. The elementary axiom of modal logic is axiom K, and by extending the axiom K, we received the other axioms. Possible world theory is based on modal logic. In this theory they definite that W is a set of possible worlds, R is a set of connections and v is the value of the propositions which is in possible worlds. This theory could describe while the agent cannot be sure about the values of propositions by connecting the possible worlds.In the next part, we introduce Dynamic epistemic logic and Public Announcement Logic. Dynamic epistemic logic (DEL) is a logical framework for dealing with changes in knowledge and information. Public Announcement Logic (PAL) is a study of modal logic of knowledge, belief, and public communication. Combine the Public Announcement Logic with possible world theory, and we can show how the beliefs of agents change. This is considered that suitable to solve the werewolf game. But even you can describe the werewolf game by possible world completely. The uncertainty of this game and the less information let that making a decision is too complicated. While the werewolf agent model based on BDI logic also facing this problem. There still need an algorithm to resist the uncertainty of the werewolf game. The next part is the Mental Spaces. Mental spaces theory is an excellent way to analyze some difficult natural language. We can use mental spaces to describe the thinking of the agents. And it also could be dynamic to represent the changing of the thought of agent. But it is difficult to reason the real mind of the agent.

The next part is about a study that develops an Observation Model of Werewolf Game. In this study, they propose a model of belief and intention change throughout a dialogue. They use Situation Calculus to model to analyze the evolution of the world and an observation model to analyze the evolution of intentions and beliefs. This model is an observation model and it could describe dialogues combined with actions. The goal of this study is to model the interaction between beliefs, intentions, and utterances. By using this model they can predict decisions resulting from the dialogue is used as a performance measure.The Observation Model provides a new view of the werewolf game. By using this model, the agent could predict the effect of his act. In that example, the seer predicts the result of voting after his coming out. While the agent could predict the result of every act he doses, he could choose the most benefit act to reach the intention of the games. This is allowed the agent to persuade other agents and win the game.

In the past studies, researchers absorbed in finding the werewolves to prove the win rate. However, here is a new idea that not only exposes the werewolf but also persuades other agents to achieve success. The Observation Model concentrates on the observe so it can not describe the werewolf game complete. In future work, we could extend the model entirely and build an agent that good to persuade others.