

Agent discription of ddhb 5th international aiwolf contest

1 Overview

Our agent consists of two main parts, Role Estimation part and Decision Making part. In Role Estimation part, we created the agent based on Basket, the winner of the 4th international aiwolf contest. In Decision Making part, we decided how to act with our own strategy based on rule-based and behavioral learning. In particular, we devised the voting method in re-voting in order to avoid as much as possible the randomness of executions.

2 Role Estimation

2.1 Score Matrices

We defined the $\text{ScoreMatrix}[A_i][R_i][A_j][R_j]$ as a score representing the likelihood that agent A_i has role R_i and agent A_j has role R_j . This is almost identical to what takeda proposed. However, we made some changes. In the original code, roles that seemed probable were added with negative values. We modified this so that roles that appear probable now accumulate positive values, making it more intuitive. Additionally, we introduced a finite upper limit of ± 100 for these values. We referenced Basket's code to determine how to update the values in the score matrix based on other agents' statements and the results of our abilities. Here, we highlight two improvements over Basket:

1. We used infinity to distinctly differentiate certain information derived from game rules. For instance, if it's known that a certain agent A_i is a villager, $\text{ScoreMatrix}[A_i][\text{Werewolf}][A_i][\text{Werewolf}]$ is set to negative infinity. This ensures that if there's a contradiction with other agents' statements, our information always takes precedence.
2. We took into consideration the "inverse." For example, when agent A_i divines agent A_j as a werewolf, the value of $\text{ScoreMatrix}[A_i][\text{Seer}][A_j][\text{Werewolf}]$ would increase. How-

ever, there’s also a possibility that agent A_i is not actually a Seer and is falsely accusing someone from the Villager faction. In ddhb, we named this ”inverse” (though it’s not a proper proposition). In this case, the score for $\text{ScoreMatrix}[A_i][\text{Werewolf faction}][A_j][\text{Villager faction}]$ would also increase. To abstract this, we can describe as follows:

- $P \wedge Q$: Seer correctly identifies a Werewolf.
- $\neg P \wedge \neg Q$: Non-seer falsely accuses a Villager (usually a Werewolf faction member).
- $P \wedge \neg Q$: Seer falsely accuses a Villager (a rare troll behavior).
- $\neg P \wedge Q$: Non-seer correctly identifies a Werewolf (either a Werewolf betraying its teammate or a Possessed making an error).

2.2 Behavioral Learning

Behavioral learning was a method used by takeda and implemented based on Basket’s code. For each agent, we record the frequency of specific actions performed in every turn and, in subsequent games, infer their roles from this information. These estimations, represented as probabilities ranging from 0 to 1, are then weighted and integrated into the score matrix. Major modifications from Basket include the addition of new types of behaviors. According to comments in Basket’s code, the actions considered in takeda were 11 types including ”CO (Villager, Seer, Medium)”, ”Divination report (white, black)”, ”Medium report”, ”Vote declaration”, ”Suspicion of being a werewolf”, ”SKIP”, and ”OVER”. In ddhb, we grouped ”SKIP”, ”OVER”, and others into ”Others”, further divided the ”Medium report” into white and black, added ”Possessed estimation”, ”Villager estimation”, ”Guard report”, and ”Voting request”. Also, behaviors like ”Divination without CO” and ”Medium result without CO” were added to detect the Werewolf faction easily. Another addition was the action ”Contradictory divination from the previous one”, which was tried to detect the Werewolf faction.

2.3 Role Estimation

This is also based on the codes of takeda and Basket. We represent the roles assigned to agents in a list format, such as Agent1 being a Werewolf, Agent2 being a Villager, ... and AgentN being a Seer, which is expressed as Werewolf, Villager, ..., Seer. In a 5-player game, the likelihood of each role assignment being correct is calculated from the score matrix. The most probable role for each agent is then inferred from the probabilities of these assignments. In a 15-player game, considering all assignments is impractical, so we start with random assignments. During moments when the server waits for our response, like TALK or DIVINE,

we slightly modify the most probable assignment and continue our search.

3 Decision Making (5-player game)

- VILLAGER

The villager agent makes random statements each turn based on role estimation. This agent votes for the agent with the highest probability of being a werewolf.

- SEER

The seer agent is basically the same as the villager agent. This agent divines the agent with the high winning rate on day 0 and the highest probability of being a werewolf on day 1. If the result is black, this agent reports it as it is, and if it is white, this agent reports it as black to one of the opposing seers.

- POSSESSED

The possessed agent pretends to be the seer agent. This agent reports black results to the agent with the lowest probability of being a werewolf among the opposing seers. If there are less than or equal to 3 alive agents, this agent CO as a werewolf (power play). This agent votes for the agent with the lowest probability of being a werewolf.

- WEREWOLF

The werewolf agent pretends to be the seer agent or the villager agent depending on the situation. This agent estimates the agent with the highest probability of being a possessed agent or a seer agent each turn. If there are less than or equal to 3 alive agents and a possessed agent is alive, this agent CO as a werewolf (power play). If a possessed agent is alive, this agent votes for the same agent as it votes for. Otherwise, this agent votes for the agent who is likely to be executed. This agent attacks the agent with the highest probability of being a villager among the agents with no seer CO in order to avoid attacking a possessed agent.

4 Decision Making (15-player game)

- VILLAGER

The villager agent acts in the same way as in the 5-player game.

- SEER

The seer agent basically acts in the same way as in the 5-player game. Unlike the 5-player game, this agent always reports true divination results.

- POSSESSED

The possessed agent basically acts in the same way as in the 5-player game. Unlike the

5-player game, this agent reports black results to the agent with the lowest probability of being a werewolf at 20% and reports white results to the agent with the highest probability of being a werewolf at 80% after day 1.

- **WEREWOLF**

The werewolf agent basically acts in the same way as in the 5-player game. In pretending the seer agent, this agent reports white result to the one of allies on day 1.

- **MEDIUM**

The medium agent is basically the same as the villager agent. This agent always reports true identification results. This agent decides a vote candidate based on it. If the result is black, this agent chooses an agent among other than agents who voted for the agent who was executed the day before.

- **BODYGUARD**

The bodyguard agent is basically the same as the villager agent. This agent determines the guard target based on the number of seer CO and medium CO.