RLereWolf – Reinforcement Learning Agent Development Framework For The Social Deduction Game Werewolf

Georgi Ventsislavov Velikov – u01gvv16@abdn.ac.uk

University of Aberdeen

Werewolf Rules

- Werewolf is a game where players, **from 5 to 75**, are randomly assigned specific roles, at the start of a game.
- The game is split into turns which consist of ***day*** time and ***night*** time.
- Roles can be one the two factions – ***evil*** and ***good***. The goal of the ***good*** faction is to kill off all ***evil*** members and vice-versa:
  - ***Villager*** – Part of the *good* faction; can vote during the *day* to execute a *Player*.
  - ***Guard*** – A member of the *good* faction; can protect players, including himself, during the *night* from being attacked.
  - ***Seer*** – A member of the *good* faction; can find out the role of a player during the *night*.
  - ***Werewolf*** – *Villagers* that turn into *Werewolves* during the *night*; member of the *evil* faction and have the ability to attack other players during the *night*.

Research Question

- Can we offer a **platform** for both multiple non-expert *Agents* to play the game and *developers* to **train** *Agents*?
- How can we improve the existing ***communication protocol*** for the *Agents*?
- How do different *Agent* behaviour models affect the winning rate?

Conclusion

- Insufficient time in order to complete the project. Resulted in unfinished *Agents*.
- Built-in *Agents* do not reach the performance of *AiWolf's Agents*.
- Potential for the development of *Agents* whose performance is not entirely bound on the *Game* implementation.
- Metrics and analysis tools the framework provides are not currently provided by any existing framework.

Tool
tinyurl.com/RLereWolfFramework

Dissertation
tinyurl.com/RLereWolfPaper

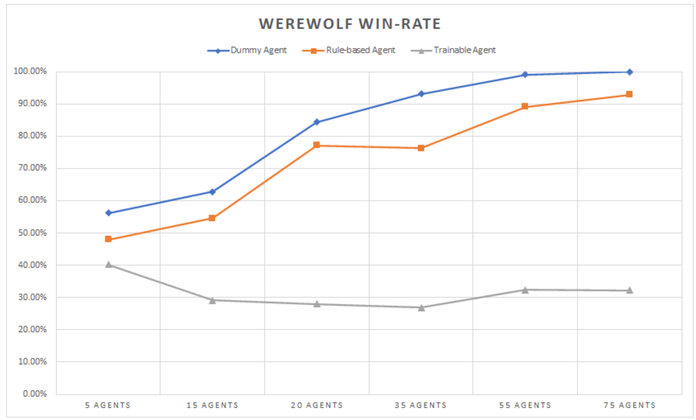
LinkedIn
linkedin.com/in/georgi-velikov/

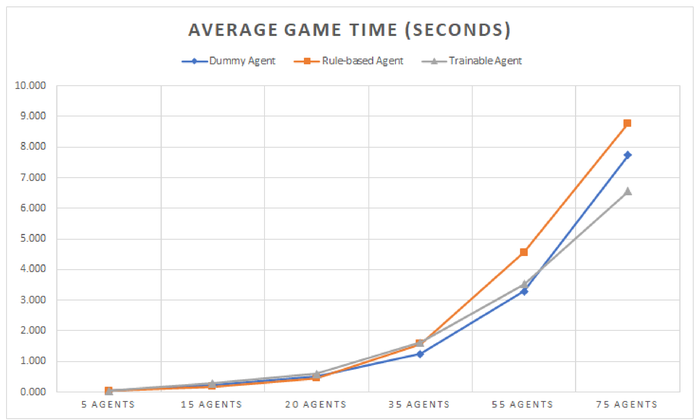
Contribution 1: Client

- Easily expandable with the developed framework GUI pipeline which employs Pygubu builders and custom built renderers.
- An access point for users to play the game Werewolf with other humans or *Agents* on the targeted RLereWolf *Server*.

Contribution 2: Built-in *Agents*

- ***Dummy Agent*** – A stochastic *Agent* that does random *valid* actions.
- ***Rule-based Agent*** – An *Agent* with an *honesty* factor who votes for the least *trustworthy*, according to them, *Player*.
- ***Trainable Agent*** – An *Agent* that can learn from playing multiple games of the current Werewolf *Game* implementation. Has no pre-existing knowledge of the game and needs to *train* in order to learn the game's rules and how to optimally play it.





Contribution 3: Development Framework

The development framework consists of:

- The Werewolf game implementation.
- The analytic utilities provided by the built-in training *Environment*.
- Comprehensive *Server* & *Game* activity logging.
- The modular implementation of the four subsystems: *Client*, *Server*, *Game*, and *Environment* (see graph below).