Mitigating Cowardice for Reinforcement Learning Agents in Combat Scenarios

Steve Bakos, Heidar Davoudi

2022 IEEE Conference on Games (CoG).

Purpose

- To remove **Cowardice** without negative affecting performance.

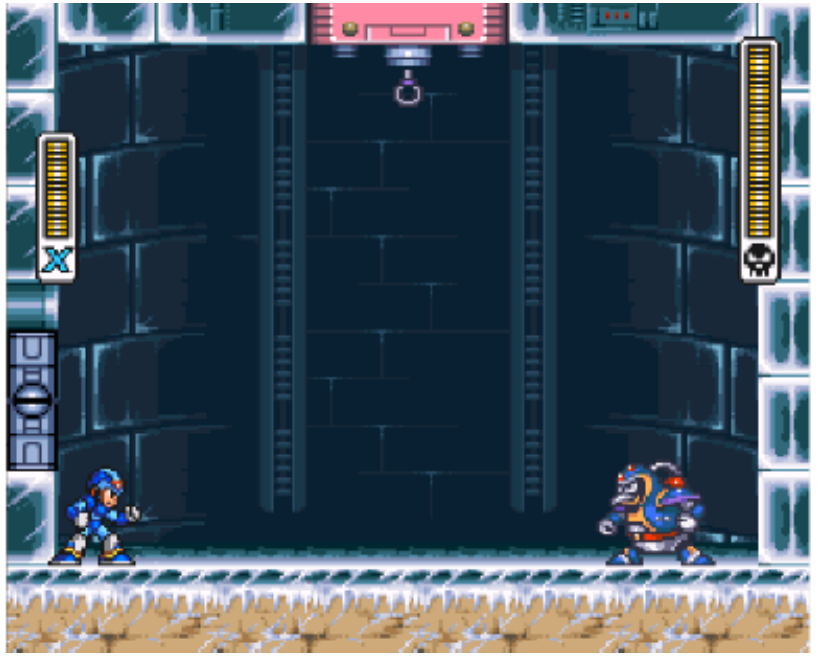
  - Cowardice leads the agent to develop non-aggressive strategy.

Method

- Reducing the fear by decaying the punishment the agent receives at the terminal state.
    - Showing it leads to an increase in the agent’s performance, stability, reduction in training time.

Method

- **Legend of Zelda**
- **Megaman X**
- **M.U.G.E.N**



Method

- Parameters
  - agent's health, opponent's health, match time remaining, ect.

$$V_{net} = \pm V_{ter}^{\frac{P_{obs}}{P_{max}}}$$

- P_obs: the value this parameter has at the terminal state
- P_max: the maximum value this parameter can take.
- V_ter: the static reward given at the terminal state, resulting in V_net: it given to the agent.

Method

$$V_{net} = \pm V_{ter}^{\frac{P_{obs}}{P_{max}}}$$

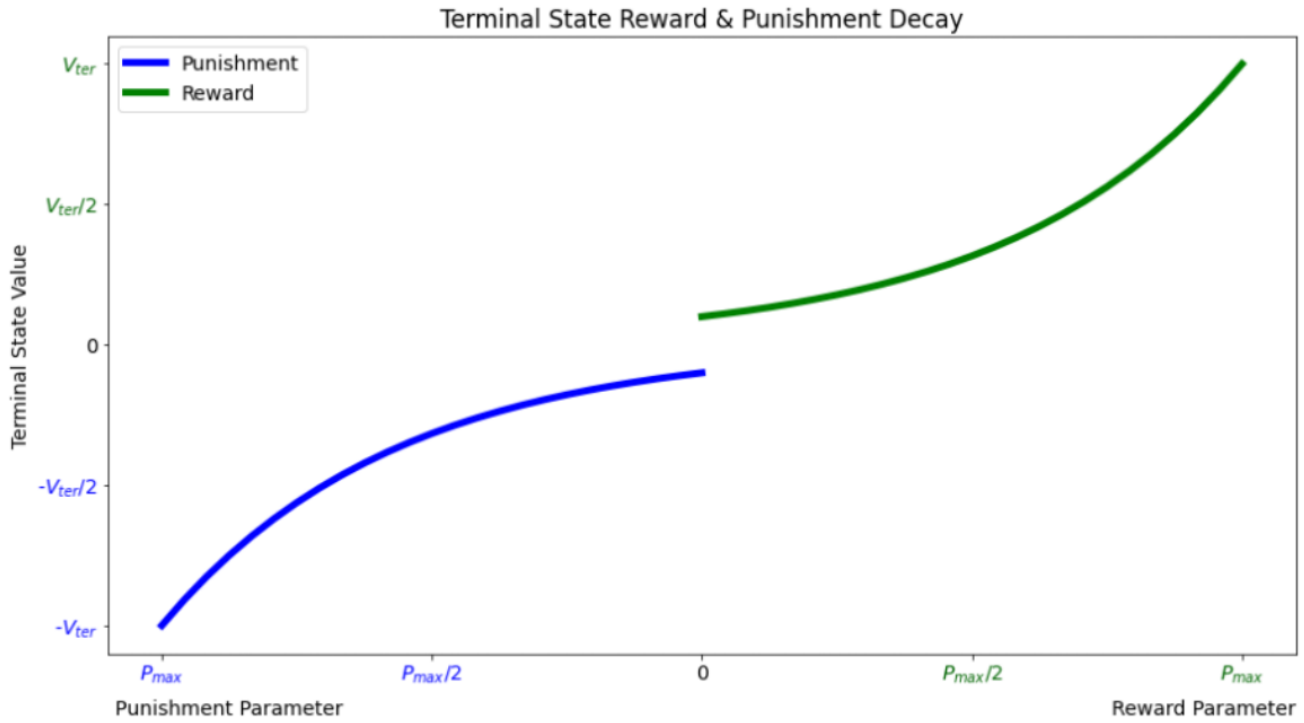


Fig. 1: V_{ter} is the value to be decayed. The reward and punishment parameters are chosen to encourage mastery and mitigate cowardice respectively. Their P_{max} values are the maximum value these parameters can take with P_{obs} being a position on the x-axis depending on win or loss.

Method

Rewards

Legends of Zelda

- doing damage to the boss: +1
- Taking damage to the health: -1
- Hit by fireball: -0.5
- Colliding with the boss: -1
- Closing the distance: +0.001
- Increase the distance: -0.001

Terminal states

- Leaving the room: -10
- Defeating the boss: +10
- Dying to the boss: -10

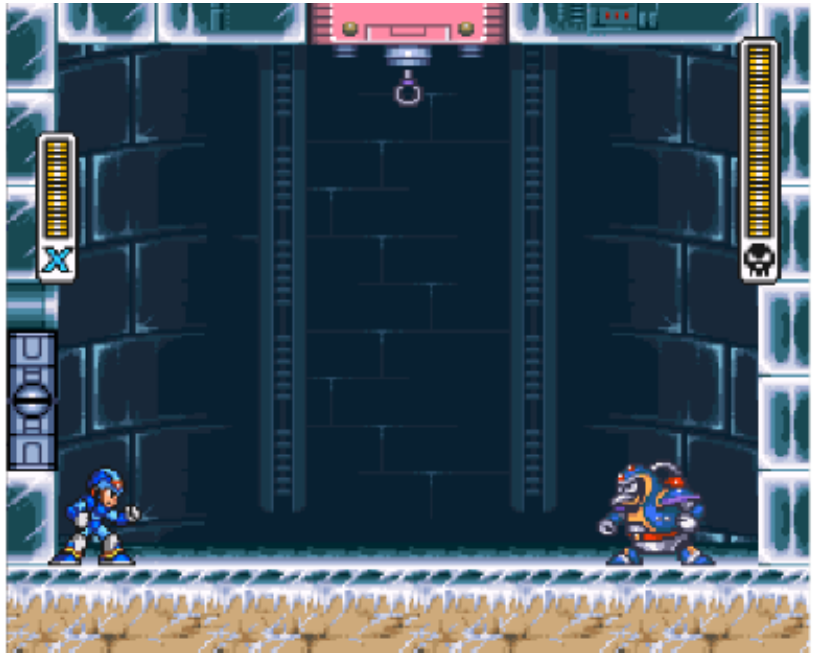
Method

Rewards

Megaman X

- doing damage to the boss: +1
- Taking damage to the health: -1

Both the agent and boss are capable of dealing multiple points of damage in a single hit.



Terminal states

- Defeating the boss: +10
- Dying to the boss: -10

Method

Rewards

M.U.G.E.N

①. Receiving static rewards(+10, -10)

②. ① + ticking time penalty

Other three agents

・receive decaying

・opponent’s health remaining as the punishment parameter.

The reward is different

③. Match time remaining. (aggressive)

④. Remaining health. (defensive)

⑤. Simple average of the two. (balanced)

Result

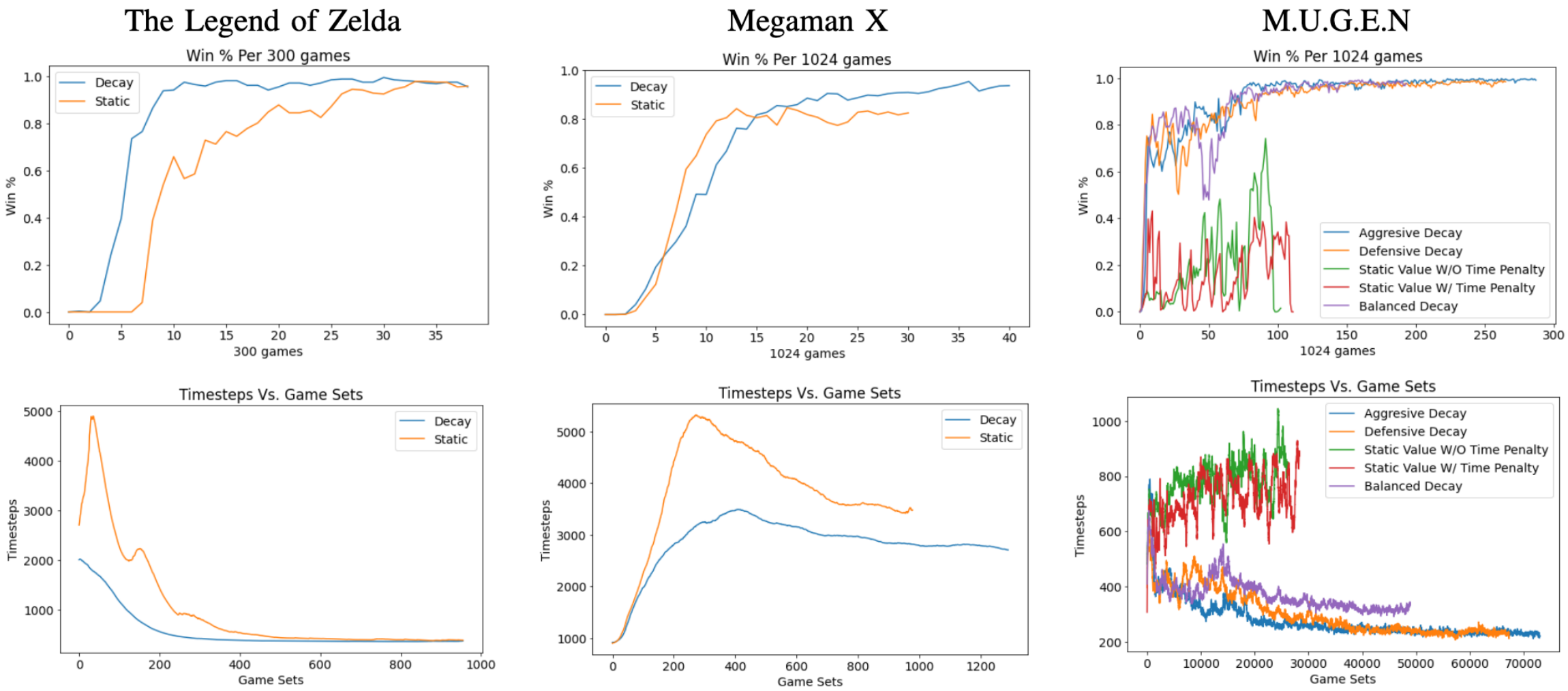


Fig. 6: Results for win rates and timestep counts across The Legend of Zelda, Megaman X, and M.U.G.E.N environments.

Conclusion

- Removing its Cowardice via decaying the terminal state rewards leads to better performance in terms of win rates, average number of timesteps, and stability.

- Allowing creating different play styles for agents through decaying reward via different parameters or combinations.

Thank you for listening!