Gradual Restriction with SC2

201911050 김승하
201911050 도담록
201911185 추성재

Index

- Backgrounds
- Motivation: APM restriction
- Methods: Gradual restriction
- Results
- Discussion
- References

Motivation: APM Restriction

- Starcraft2

Figure 1. A screenshot of DeepMind AlphaStar vs. StarCraft 2 Pro Player MaNa

Motivation: APM Restriction

- APM(Actions per minute): the number of actions in a minute.
- High-level Pro Players can perform 800 APM in a moment until their physical limits.
- AI models can make actions until their tremendous computing power.

Motivation: APM Restriction

- APM 750
- Blue line: maximum action/sec is 50
- Orange line: maximum action/sec is 15
- **How to impose a restriction on the well-trained model to meet conditions**

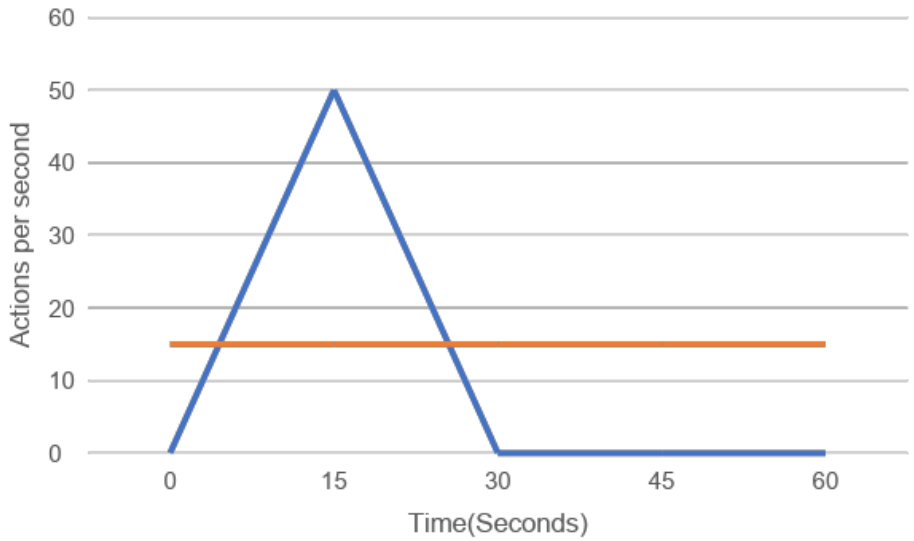


Figure 2. Both Sudden Bursting actions and Steady actions are APM 750

Backgrounds

- Starcraft2 and Pysc2 – StarCraft II Learning Environment
- Low-level human interface action space
- Too big Action Space

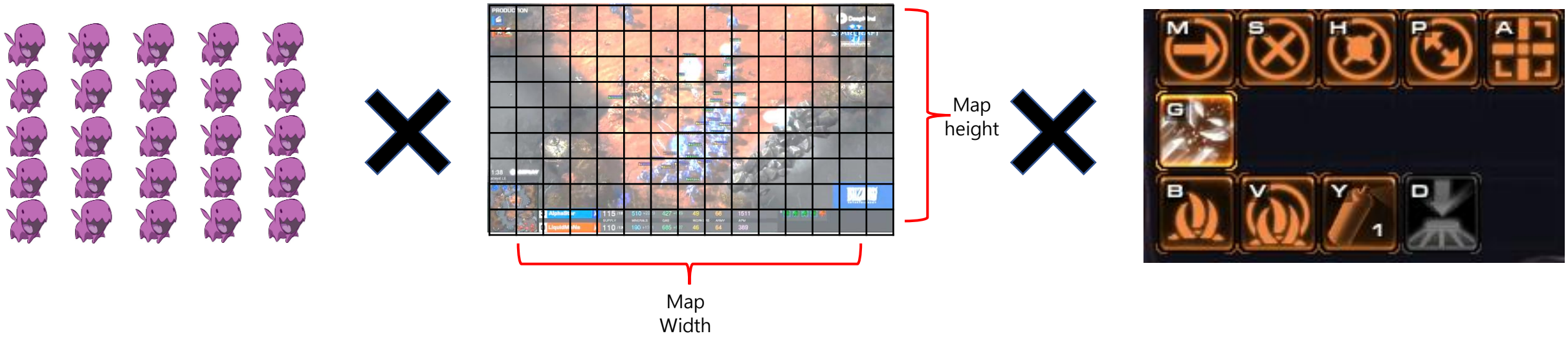


Figure 3. pysc2 library has a large action space

Backgrounds

- TStarBots wraps Action into 165 macro actions.

- Implying Strategy
- Easy Game Rule Learning
- Reducing Trivial Decision

- PPO algorithm

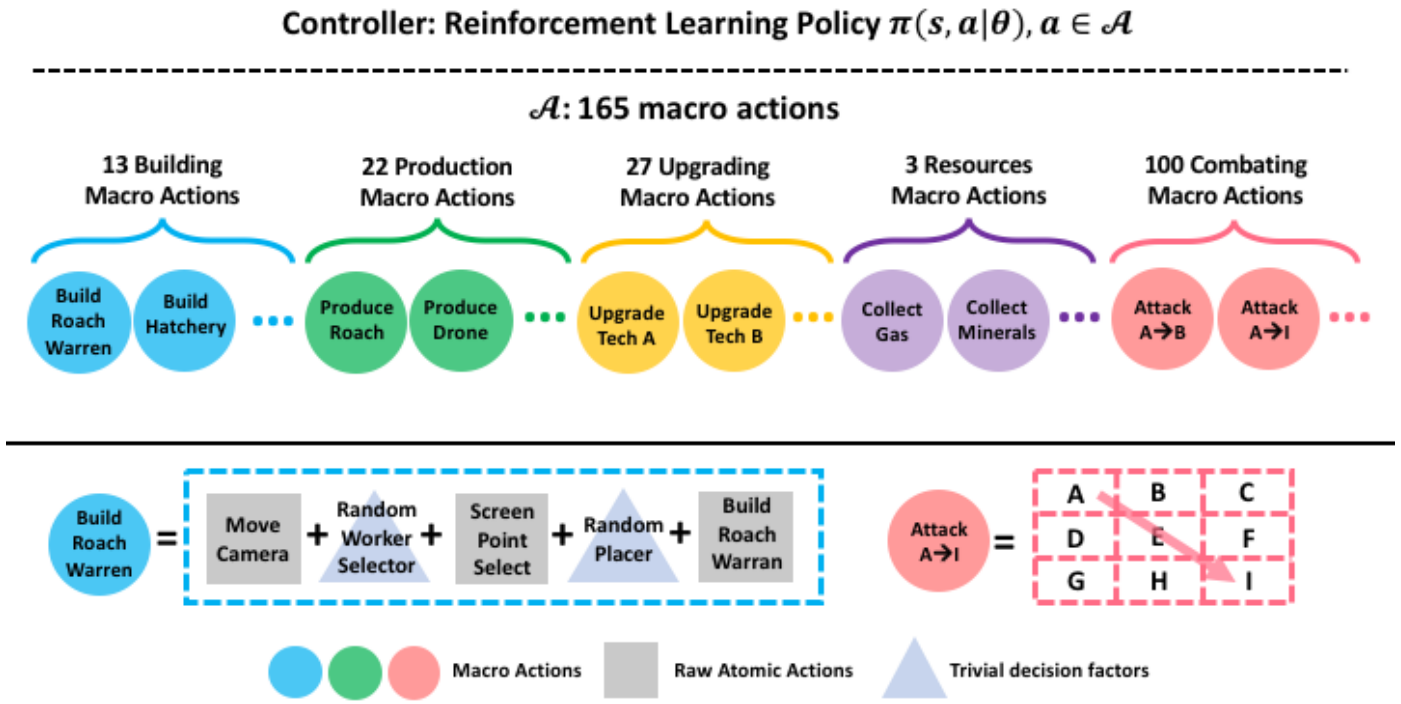


Figure 4. TStarBot has a light action space

Method: Gradual Action Restriction

- **Action Restriction**

- Action burst indicator n

$$n_{k+1} = \begin{cases} \max(n_k - 1, 0) & \mathrm{n_k} \geq 100\alpha \\ n_k + 1 & n_k < 100\alpha \end{cases}$$

- **Choose action== 0 (no-op) when n $\geq 100\alpha$**
- Can restrict “momentary” bursting of APM
- Restriction Condition $\alpha = 0.2$

Method: Gradual Action Restriction

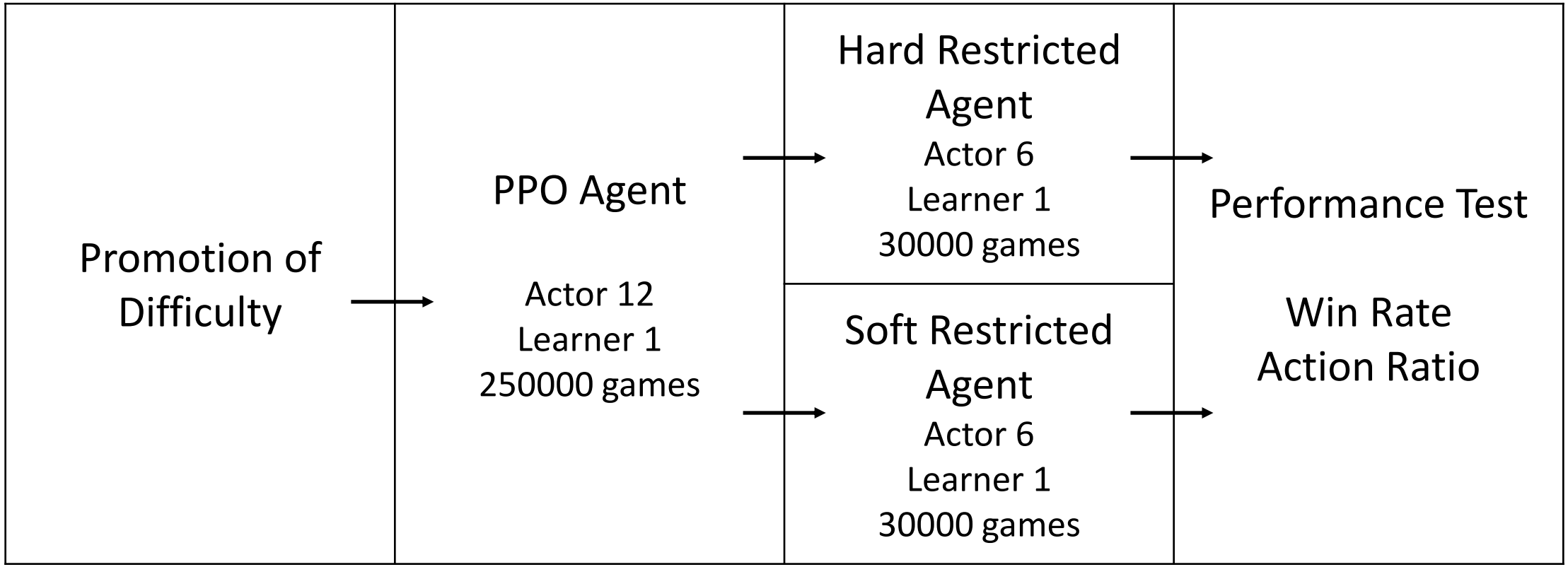
- Train PPO Agent until defeating SC2 AI Bot Difficulty 7
- Retrain PPO Agent imposing Action Restriction
- Hard Restriction: $\alpha = 0.2$
- Soft Restriction: gradually reduced α from 1 to 0.2 linearly

Method: Promotion of Difficulty

- TStarBot: Actor vs. Random Difficulty 1 to 10
- Assumption: Too Easy game, Too Hard game might affect model training.
- Ours: Promotion when the win rate is over 50% between the recent 100 games.

Method

- Train Pipeline



Results

- PPO Agent
- The agent at the start of the training behaves similarly to a random agent by repetitively constructing inexpensive buildings without an apparent strategy.

Video 1. Random Agent

Video 2. PPO Agent Checkpoint_50000

Results

- PPO Agent
- As the training progresses, the agent shows a meaningful abstract-level feature.
- The agent's action is similar to the human build order in PvP game.

Video 3. PPO Agent Checkpoint_250000

Results

- Restriction Agents
- Pattern requiring fewer actions
- Avoiding forward base & Prefer defense towers rather than attack units

Video 4. Hard Restrict Agent Checkpoint_30000

Video 5. Soft Restrict Agent Checkpoint_30000

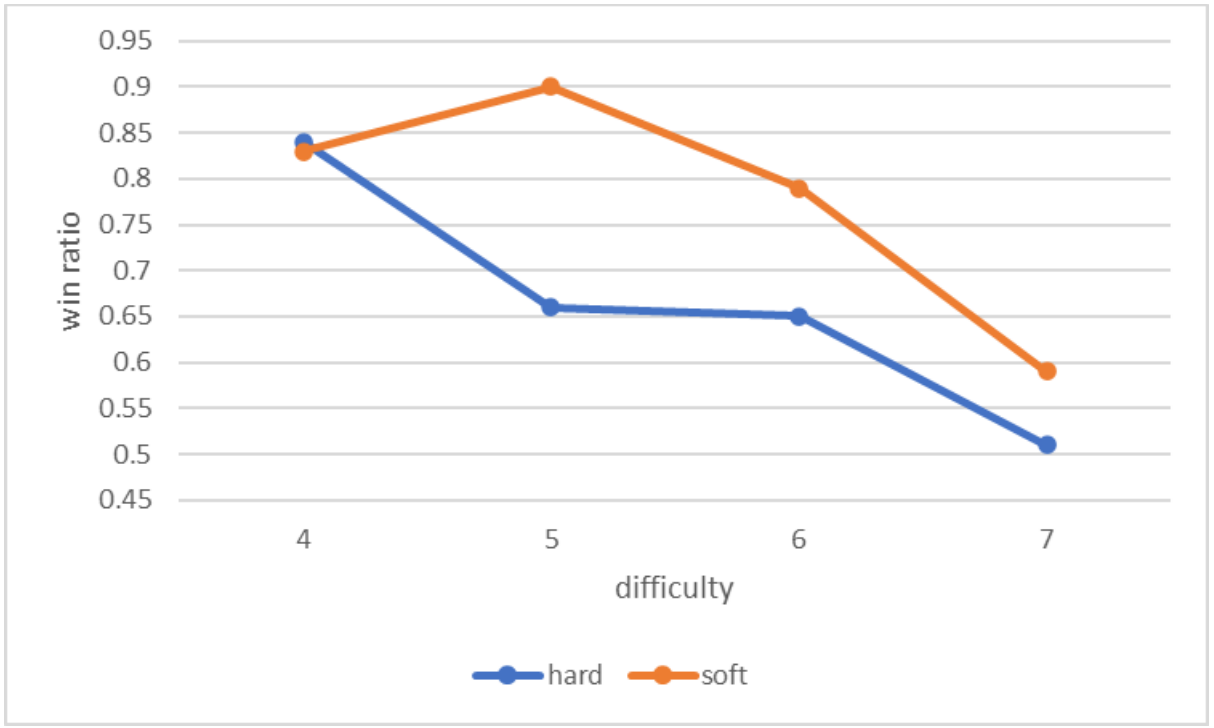
Results

- Abnormal behavior in the StarCraft 2 AI bot at difficulties above 8 (With maphack)
- The bots try to gather the agent's base resources as the bot's resources.

Figure 5. Abnormal behavior of StarCraft 2 AI Bot at difficulties above 8

Results

- Soft restricted model(purposed) had better win ratio than hard restricted model.



t-검정: 등분산 가정 두 집단

	hard6	soft6
평균	0.65	0.79
분산	0.229798	0.167576
관측수	100	100
P(T<=t) 단측 검정	0.013746	

Figure 6. Win Ratio: Soft Agent vs Hard Agent (Difficulty Levels 4-7) for SC2 AI Bot

Results

- Action ratio of Restrict trained model in removed restriction
- Frequent actions with removed restriction
- Models can act abnormally in restriction-removed environments.

Agent vs. Diff. 7	Restriction	Restriction removed
action ratio(hard)	0.48	0.73
action ratio(soft)	0.49	0.75

Table 1. Action Ratio Comparisons under Imposed and Removed Restrictions

Discussion

- Training with gradual imposing restrictions performs better than training with direct imposing.
- This methodology can be implied in environment adaptation training. Robot AI adaptation from the Earth to Mars.

Figure 7. NASA Spirit rover

Reference

- Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A. S., Yeo, M., ... & Tsing, R. (2017). Starcraft ii: A new challenge for reinforcement learning. arXiv preprint arXiv:1708.04782.
- https://github.com/deepmind/pysc2
- Lee, D., Tang, H., Zhang, J., Xu, H., Darrell, T., & Abbeel, P. (2018). Modular Architecture for StarCraft II with Deep Reinforcement Learning. Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 14(1), 187-193. https://doi.org/10.1609/aiide.v14i1.13033
- Liu, Y., Wang, W., Hu, Y., Hao, J., Chen, X., & Gao, Y. (2020, April). Multi-agent game abstraction via graph attention neural network. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 7211-7218).
- Sun, P., Sun, X., Han, L., Xiong, J., Wang, Q., Li, B., ... & Zhang, T. (2018). Tstarbots: Defeating the cheating level builtin ai in starcraft ii in the full game. arXiv preprint arXiv:1809.07193. https://doi.org/10.48550/arXiv.1809.07193
- https://github.com/Tencent/PySC2TencentExtension
- https://github.com/Tencent/TStarBot1
- https://commons.wikimedia.org/wiki/File:NASA_Mars_Rover.jpg

Thank you for listening to our presentation.

QnA