TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

AlphaStar

AlphaStar

Grandmaster level in StarCraft II using multi-agent reinforcement learning, Nature Oct. 2019, Vinyals et al.

StarCraft:

- Players control hundreds of units.
- Individual actions are selected from 10^{26} possibilities (an action is a kind of procedure call with arguments).
- Cyclic non-transitive strategies (rock-paper-scissors).
- Imperfect information — the state is not fully observable.

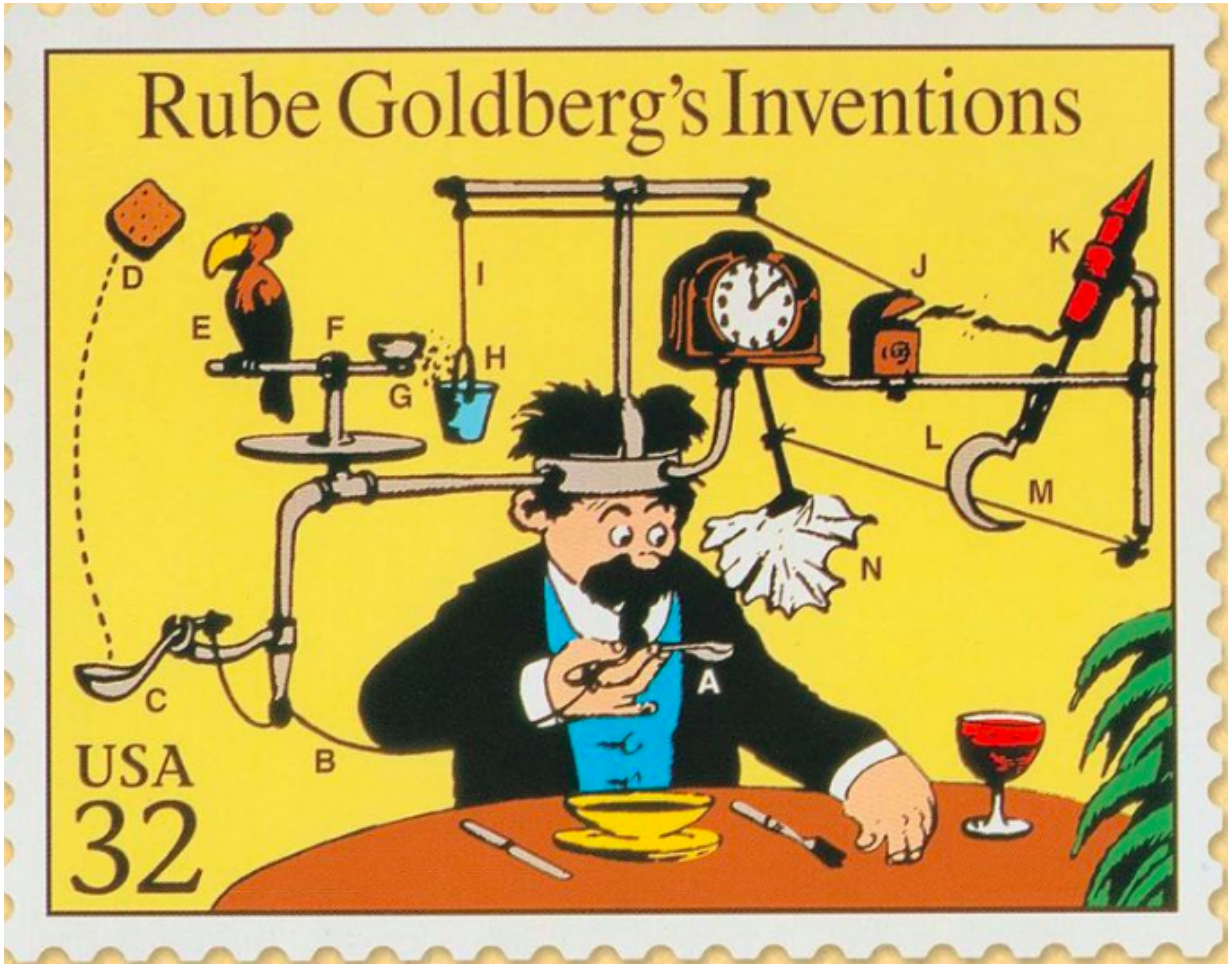
The Paper is Vague

It basically says the following ideas are used:
A policy gradient algorithm, auto-regressive policies, self-attention over the observation history, LSTMs, pointer-networks, scatter connections, replay buffers, asynchronous advantage actor-critic algorithms, TD(λ) (gradients on value function Bellman error), clipped importance sampling (V-trace), a new undefined method they call UPGO that "moves policies toward trajectories with better than average reward", a value function that can see the opponents observation (training only), a "z statistic" stating a high level strategy, supervised learning from human play, a "league" of players (next slide).

The League

The league has three classes of agents: main (M), main exploiters (E), and league exploiters (L). M and L play against everybody. E plays only against M.

A Rube Goldberg Contraption?



Video

https://www.youtube.com/watch?v=UuhECwm31dM

END