

Fuzzy Rule BayesOpt AutoRL

Bayesian Optimization

$$z^* = \operatorname{argmax}_z f(z)$$

f : Team 1's End-Game Winrate/Reward/Advantage

$z = [1.00, 1.00, 4.00, 6.00, 0.42, \dots]$

Fuzzy Logic Decoder

Fuzzy Logic Encoder

IF ... THEN ...

Fuzzy Rules



Opponent
RL Learner 3

Opponent
RL Learner 2

POV (Point of View) of Learner 1

Generic statistics & feedback

- $\{P\}$ Reward Winrate
- $\{P\}$ Accumulated Winrate
- $\{P\}$...
- $\{R\}$ Agent Life Length
- $\{R\}$ Number of Alive Agent
- $\{R\}$...

agent trajectories
(observation & reward)

agent policies (actions)

Fuzzifier

Fuzzy Logic Processor

Mamdani-type Fuzzy Controller (MFC)

Defuzzifier

Fuzzy Feedback Adjustment

Model Feedback Loop

Centralized
Policy Optimizer

Decentralized
Actor

Centralized
Critic

Improved RL Learner for Team 1

RL Learner
Dynamic
Adjustment

- RL Hyper-Parameters
- Intrinsic Rewards
- Explorative-Exploitative Balance
- ...