GMAC: A Distributional Perspective to Actor-Critic Framework

Daniel Wontae Nam, Younghoon Kim, Chan Y. Park

ML2

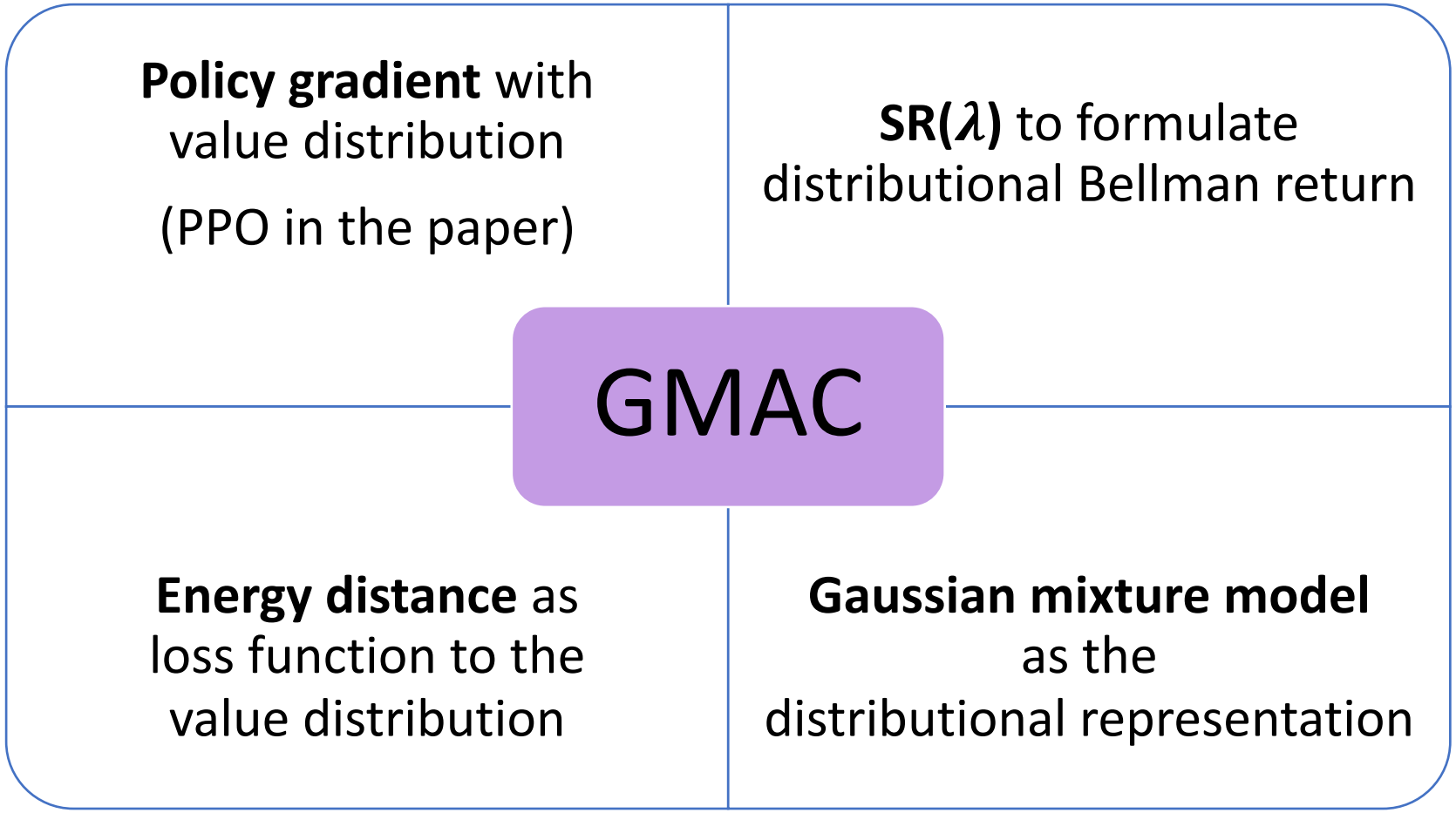
KC Machine Learning Lab

ICML 2021

ML2

GMAC – Gaussian Mixture Actor-Critic

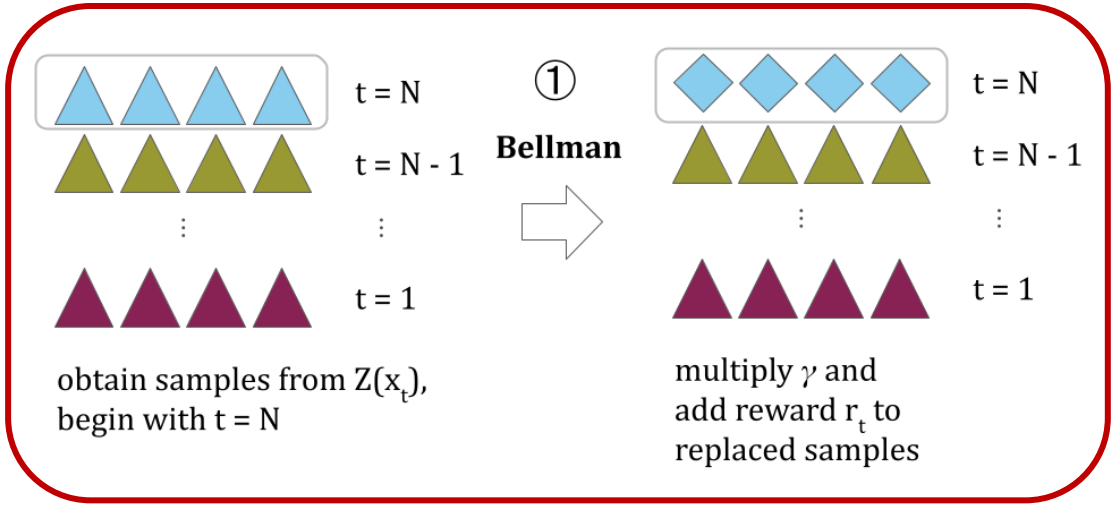
- Distributional RL: Comprehensive distributional actor-critic framework called Gaussian Mixture Actor-Critic (GMAC) which uses



ML2

SR(λ) – Sample Replacement

Finding the distributional λ-return in linear time

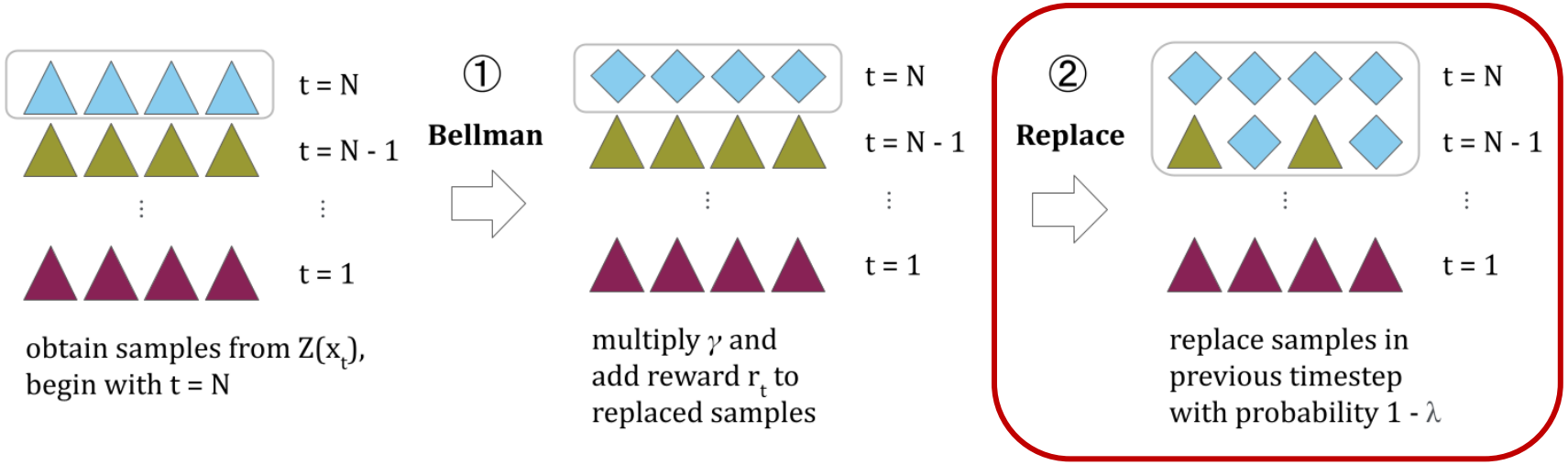


Apply Bellman operation to the samples given at time t=N

ML2

SR(λ) – Sample Replacement

Finding the distributional λ-return in linear time

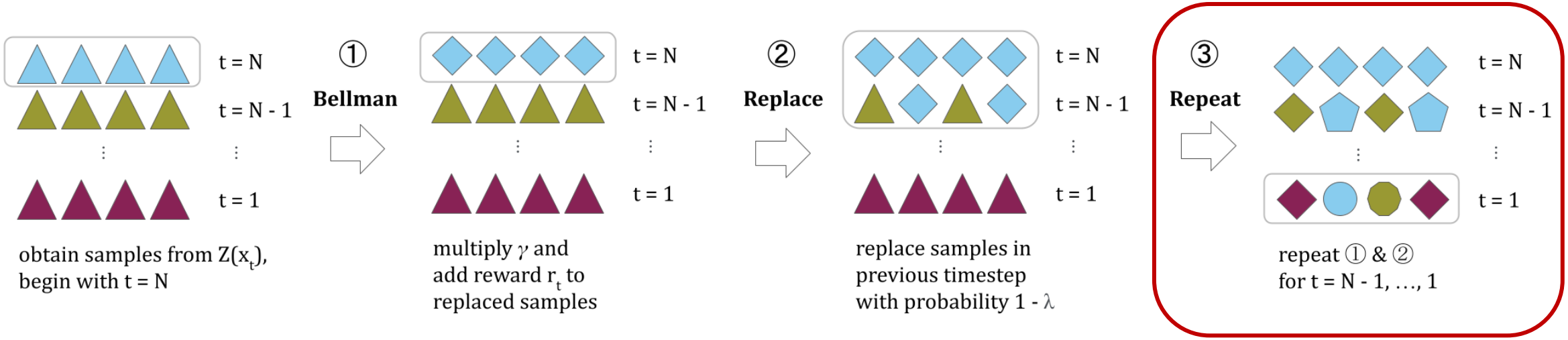


Mix the resulting samples to the samples in t=N-1 with the probability $1 - \lambda$

ML2

SR(λ) – Sample Replacement

Finding the distributional λ-return in linear time



Repeat until the beginning of the trajectory and $Z_t^{(\lambda)}$ is obtained for all t

$$F_{Z_t^{(\lambda)}} = (1-\lambda)\sum_{n=1}^{\infty} \lambda^{n-1} F_{Z_t^{(n)}}$$

ML2

Energy Distance from the Cramér distance

Minimize the energy distance between obtained Bellman target distribution and the prediction

Cramer Distance

$$l_p(P,Q) = \left(\int_{-\infty}^{\infty} \left| F_P(x) - F_Q(x) \right|^p dx \right)^{1/p}$$

Energy Distance

$$2l_2^2\,(P,Q) = \mathcal{E}(P,Q)$$

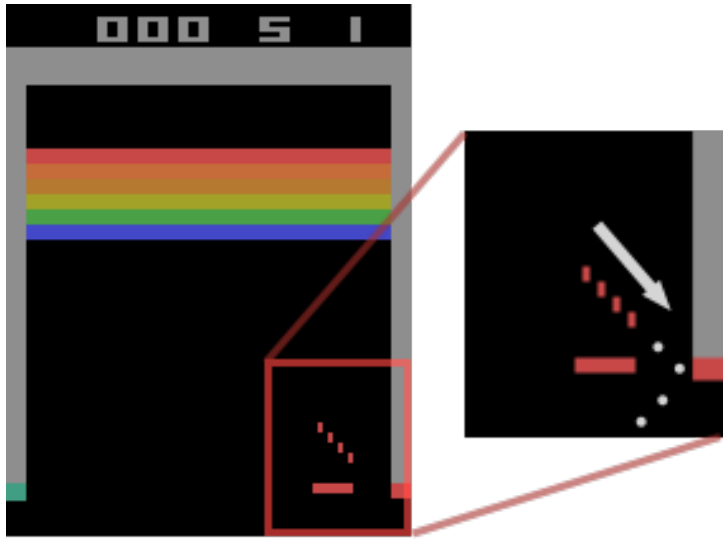
$$= 2\mathbb{E}\|U - V\|_2 - \mathbb{E}\|U - U'\|_2 - \mathbb{E}\|V - V'\|_2,$$

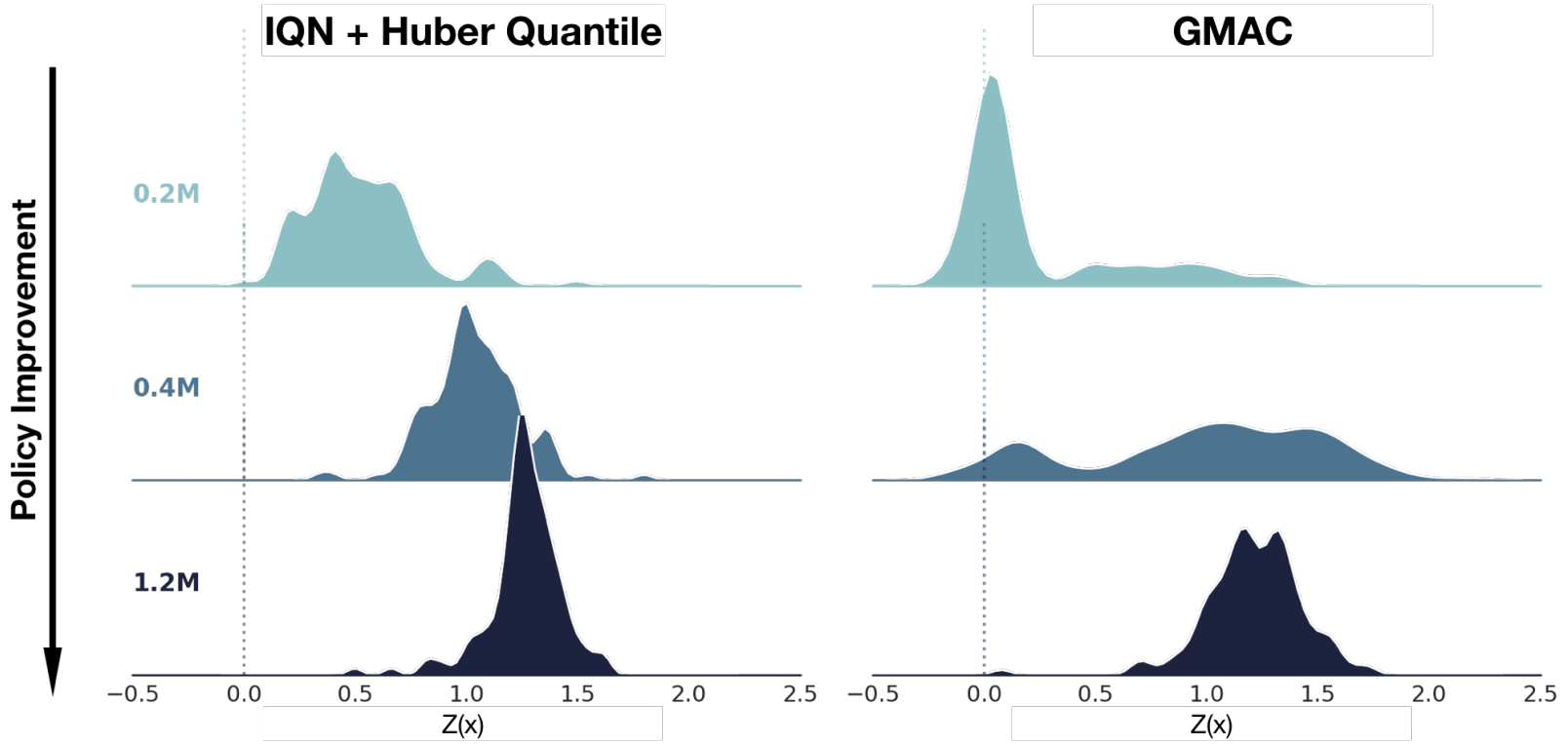
$$U, U' \sim P, \qquad V, V' \sim Q$$

ML2

Learning the right value distribution

- Value distribution learned from actual game of Breakout from ALE through multi-step Bellman return distribution

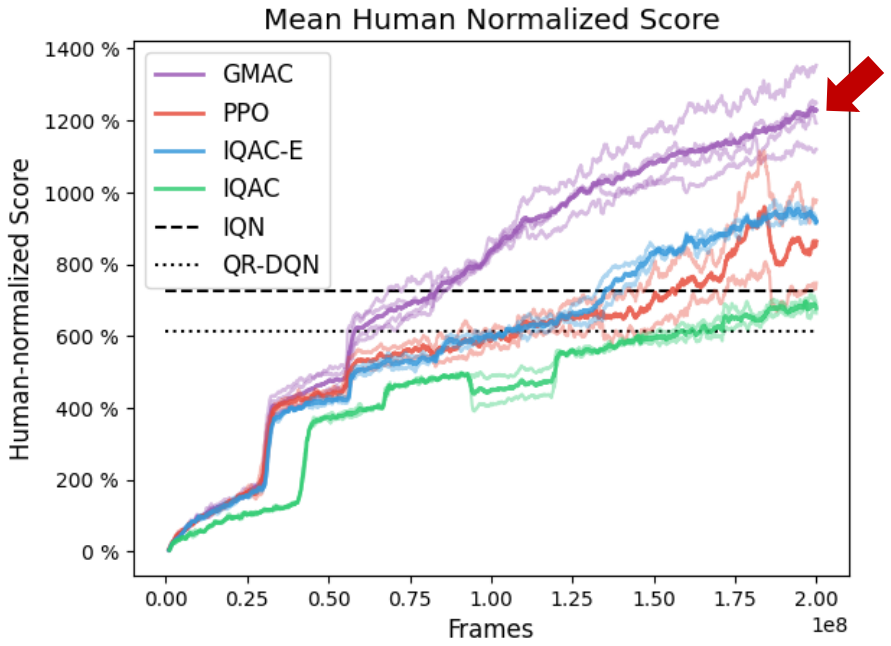


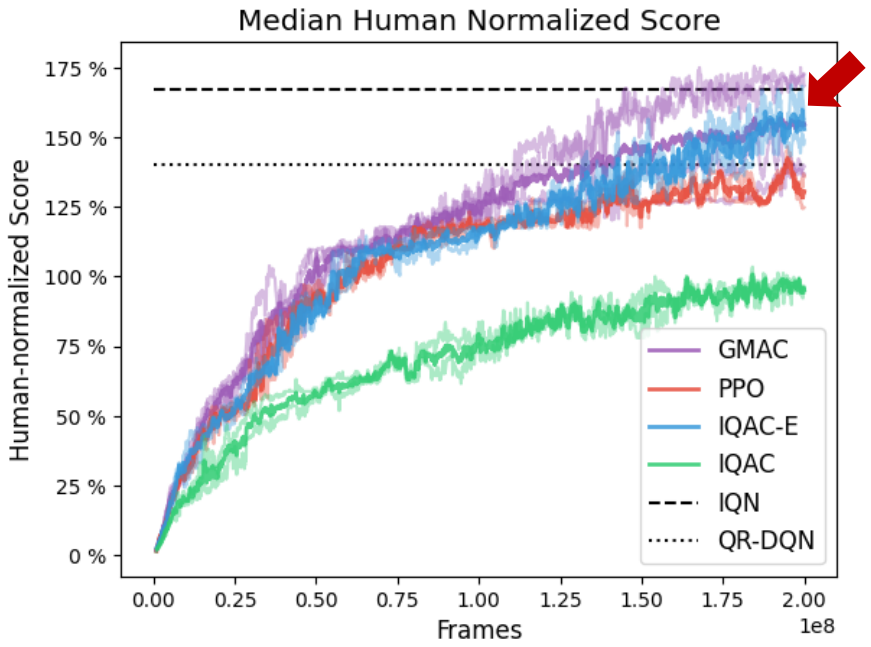


ML2

Improvement in performance

On Atari ALE tasks (57 that have human normalized scores)





FLOPs comparison to IQN architectures

Algorithm	Params (M)	FLOPs (G)	
		Inference	Update
PPO	0.44	1.73	5.19
IQAC	0.52	2.98	12.98
IQAC-E	0.52	2.98	8.98
GMAC	**0.44**	**1.73**	**5.27**

Wall clock time comparison to imputation strategy (using SciPy) on tabular case

ER-naive	ER-imputation	Energy	GM
0.0021	0.0340	0.0011	0.0035

ML2

Conclusion

GMAC

SR(λ) | Energy Distance | Gaussian Mixture Model

- Efficient algorithm to obtain and learn distributional value function for actor-critic methods like PPO that uses multi-step returns, e.g. λ-return

Code: https://github.com/kc-ml2/gmac

Our website: https://www.kc-ml2.com/

Daniel Wontae Nam, Younghoon Kim, Chan Y. Park

ML2