Evolutionary Diversity Optimization with Clustering-based Selection for Reinforcement Learning

Yutong Wang*, Ke Xue*, Chao Qian
(*Equal contribution)
Email: {wangyt, xuek, qianc}@lamda.nju.edu.cn
LAMDA Group, Nanjing University, China

ICLR 2022

Background and Motivation

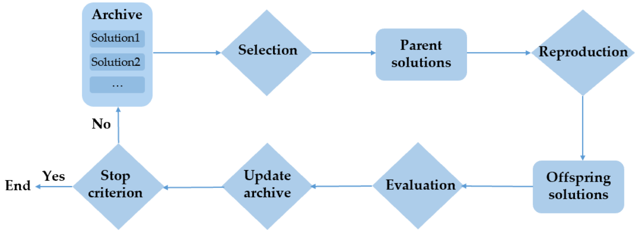
Reinforcement Learning (RL)

- General RL methods obtain a single policy
- Some complex scenarios need a set of diverse policies
  - better exploration
  - faster few-shot adaption
  - greater robustness

How to efficiently obtain a set of high-quality policies with diverse behaviors is a challenging problem in RL

Quality-Diversity (QD) algorithms

- a specific type of Evolutionary Algorithms (EAs)
- aims to return a set of high-quality solutions with diverse behaviors

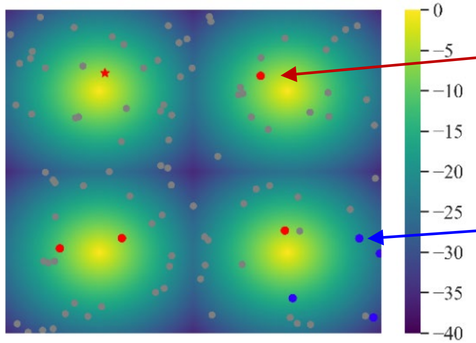


- the inefficient selection results in the poor performance

EDO-CS Method

Clustering-based selection mechanism

- clusters the policies in the archive based on their behaviors
- selects a high-quality policy from each cluster



EDO-CS Method

Self-adjusting reproduction mechanism

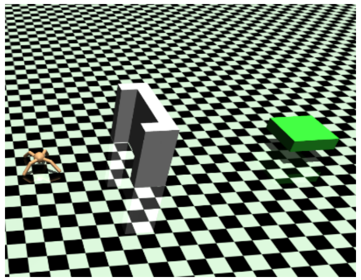
- the objective function to be maximized

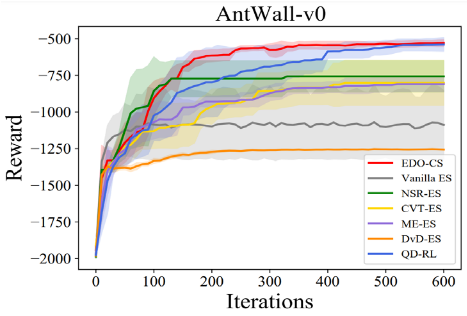
$$J(\theta) = (1-\lambda)E[R(\tau)] + \lambda Div(\theta)$$

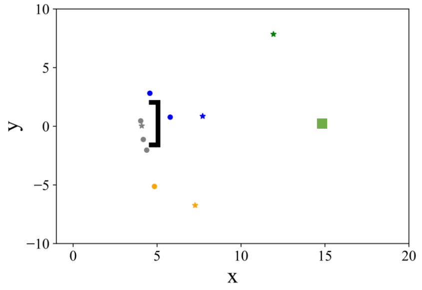
The weight λ controls the trade-off between exploitation and exploration, we use multi-armed bandit to self-adjust it

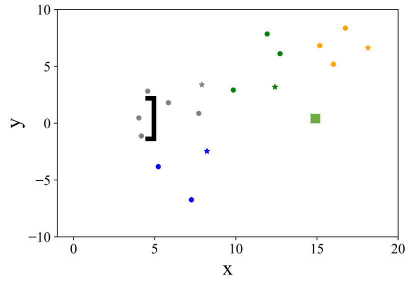
Method	Selection	Reproduction
Vanilla ES	The only parent solution	Quality
NSR-ES	Probabilistic selection	Quality and diversity
CVT-ES	Uniform selection	Quality and diversity
ME-ES	Biased selection	Quality or diversity
DvD-ES	All parent solutions	Quality and diversity
QD-RL	Pareto-based selection	Quality or diversity
EDO-CS	Clustering-based selection	Quality and diversity

Experiment









Environment	EDO-CS	QD-RL	ME-ES	DvD-ES	CVT-ES	NSR-ES	Vanilla ES
HalfCheetahFwd	**4284**	2930	2700	-3419	3219	1346	-5543
HalfCheetahBwd	**6548**	6013	5953	6353	4672	5366	3911
AntFwd	**4617**	4291	4316	4507	3856	1737	1911
AntBwd	**4697**	4164	4123	3498	2958	3961	-851
Performance Ranking	**1**	3	3.5	3.75	4.75	5.25	6.75

EDO-CS shows superior performance on various control tasks