Parallel exploration via negatively correlated search

Peng YANG, Qi YANG, Ke TANG, Xin YAO

Frontiers of Computer Science, DOI: 10.1007/s11704-020-0431-0

Problems & Ideas

- **Problems:** How to get high quality solutions for a problem with local optima and the objective function is non-differentiable?
- **Ideas:** Adaptively seeking multiple probabilistic models that both lead to solutions of high quality and are distant from previous obtained models.

(1) Defining the objective of Parallel Exploration

The fitness objective for exploitation

The diversity objective for exploration

$$
\begin{aligned}
\mathcal{J} &= \mathcal{F} + \mathcal{D} \\
&= \sum_{i=1}^{\lambda} f(\boldsymbol{\theta}_i) + \sum_{i=1}^{\lambda} d\left(p(\boldsymbol{\theta}_i)\right) \\
&= \sum_{i=1}^{\lambda} \underbrace{\int f(\boldsymbol{x}) p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)\,\mathrm{d}\boldsymbol{x}} + \sum_{i=1}^{\lambda} \sum_{j=1, i \neq j}^{\lambda} \underbrace{\left(-C\left(p(\boldsymbol{\theta}_i), p\left(\boldsymbol{\theta}_j\right)\right)\right)}
\end{aligned}
$$

The expectation of solution qualities from each distribution

Negative Correlations between pairwise distributions

(2) Maximizing the objective by evolving each distribution in parallel to cover different promising areas.

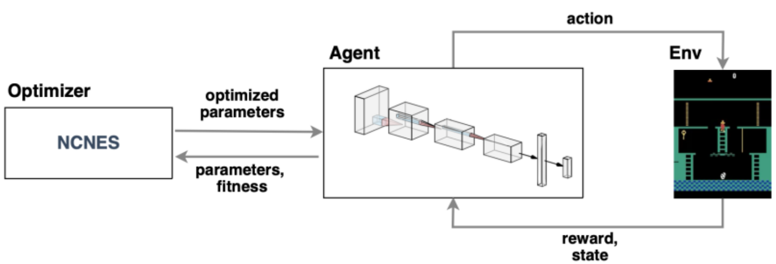
$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}_i} \mathcal{J} &= \nabla_{\boldsymbol{\theta}_i} \mathcal{F} + \nabla_{\boldsymbol{\theta}_i} \mathcal{D} \\
&= \nabla_{\boldsymbol{\theta}_i} f(\boldsymbol{\theta}_i) + \nabla_{\boldsymbol{\theta}_i} d\left(p(\boldsymbol{\theta}_i)\right), \quad i = 1, \ldots, \lambda
\end{aligned}
$$

$$
\boldsymbol{\theta}_i = \boldsymbol{\theta}_i + \eta \cdot \nabla_{\boldsymbol{\theta}_i} \mathcal{J}
$$

Main Contributions

- A parallelizable exploration strategy which is equivalent to an intuitive sequential exploration process, namely NCNES.
- An Evolutionary Reinforcement Learning method based on the proposed NCNES.

Evolutionary Reinforcement Learning Flowchart



Key Challenge:
Training a neural network involving 1.7 millions weights, with uncertain and delayed rewards.

Scoring higher
than the well-known Deep Reinforcement Learning methods

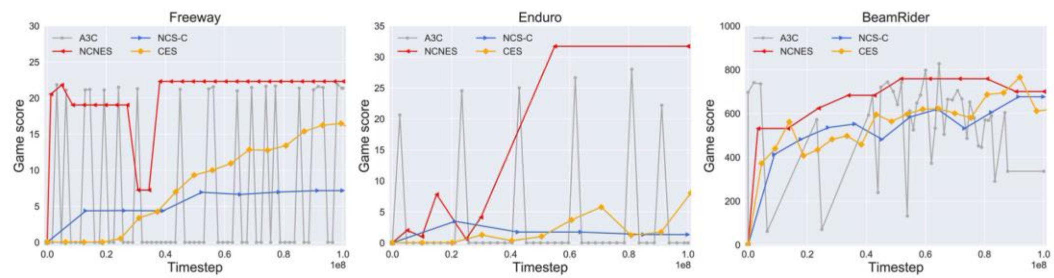
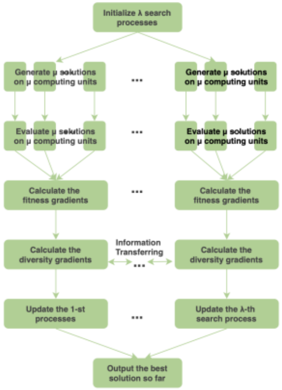


Fig. 3 The score curves of four algorithms on three games, respectively.

Table 5 The runtime of three implementations of NCNES

Computing model	Game	Serial model	Island-model	Hybrid-model
Computing Units(i.e. m)		1	5	75
Runtime(hours)	Freeway	122.6 ± 0.5	31.2 ± 0.2	2.28 ± 0.0
	BeamRider	116.0 ± 18.8	58.8 ± 22.2	19.48 ± 4.3
	Enduro	119.6 ± 0.7	30.4 ± 0.1	2.16 ± 0.0
Speedup Ratio $\frac{t_{\text{serial}}}{t_{\text{parallel}\cdot m}}$	Freeway	-	0.78 ± 0.01	0.72 ± 0.00
	BeamRider	-	0.43 ± 0.20	0.09 ± 0.03
	Enduro	-	0.79 ± 0.01	0.74 ± 0.02



Speedup faster
accelerating 55 times faster with 75 parallel computing units.