Reinforcement Learning, MOEA/D and the Sample Complexity

Chen Jingjing, 55766475

July 12, 2020

Abstract

This article includes the introduction and methodologies of Reinforcement Learning(RL) and MOEA/D, as well as the analysis of sample complexity.

Keywords: reinforcement learning, MOEA/D, sample complexity

1 Multi-armed Bandit Problems

1.1 Techniques of Exploration versus Exploitation

I have written the most of them in the IPad.

1.1.1 ϵ -greedy

1.1.2 Optimistic Initial value

The Greedy actions are those that look best at present, but some of the other actions may actually be better. ϵ -greedy action selection forces the non-greedy actions to be tried, but indiscriminately, with no preference for those that are nearly greedy or particularly uncertain. It would be better to choose among the non-greedy actions according to their potential for actually being optimal.

But it's only a trick, which can not be used in general problems.

1.1.3 Upper-Confidence-Bound Action Selection

1.1.4 Gradient Bandit algorithms

Methods above are all in common which considered calculating the estimate action values then selecting actions according to these estimated value.

1.2 Contextual Bandits

Contextual Bandits aim to find the policy: a mapping from situations(states) to the actions that are best in those situations.

Contextual Bandits is a bridge between MAB and general RL for the only one situation in MAB and the states transitions in general RL.

1.3 Articles related to MAB

1.3.1 PAPER 1: Bayesian optimization for modular black-box systems with switching costs [10]

Main idea

The cost function of Bayesian Optimization in previous research depends on current inputs, while in this article, cost depends on the changes between iterations. The optimization problem is cast into a selection problem where optimal variables i (arms) are selected from a set $\mathcal{K}$ to minimize the loss function $l : \mathcal{K} \to R$. At each iteration, the loss is $l(i^t)$ of arm i^t. GP-UCB as an acquisition function is not a new idea used in Bayesian Optimization. Under certain conditions, the iterative application of this acquisition function will converge to the true global minimum of f.

Slowly moving Bandit(SMB) algorithm took the switching cost into account, $c(i^t, i^{t-1})$, a cost when switching between arms from $t-1$ to t. SMB aims to minimize the linear combination of the cost and the switching cost.

Comments

Authors extended SMB to the setting of black-box optimization with modular structure. At each iteration, SMB choose an arm based on the probability distribution p_t. It's s conditional sampling to encourage slow switching, to make the arm selection close to the previous choice.

2 The Research about Sample Complexity in Reinforcement Learning

At first, we want to extend the bound in sample complexity in Azar's work to the non-markovian environment. So I searched articles and listened to the conference related to this area. Meanwhile, the theoretical guarantee that huiling proposed is about sample complesity and UCB algorithm. So sample complexity in RL is a line for me to consider.

2.1 Paper 1: The Optimal Sample Complexity of PAC Learning [1]

waiting to update

Main idea

This is a basic article related to the sample complexity in supervised learning.

Comments

2.2 Paper 2: When to Trust Your Model: Model-Based Policy Optimization [5]

Abstract

Designing effective model-based reinforcement learning algorithms is difficult because the ease of data generation must be weighed against the bias of model- generated data. In this paper, we study the role of model usage in policy optimization both theoretically and empirically. We first formulate and analyze a model-based reinforcement learning algorithm with a guarantee of monotonic improvement at each step. In practice, this analysis is overly pessimistic and suggests that real off-policy data is always preferable to model-generated on-policy data, but we show that an empirical estimate of model generalization can be incorporated into such analysis to justify model usage. Motivated by this analysis, we then demonstrate that a simple procedure of using short model-generated rollouts branched from real data has the benefits of more complicated model-based algorithms without the usual pitfalls. In particular, this approach surpasses the sample efficiency of prior model-based methods, matches the asymptotic performance of the best model-free algorithms, and scales to horizons that cause other model-based methods to fail entirely.

Comments

Zhang weinan and Yu yang's work all pay attention to the model-based RL and the algorithms they proposed motivated by this article, then they all analyzed the sample complexity of the new algorithm in their article.

2.3 Paper 3: The Sample-Complexity of General Reinforcement Learning

Main idea Waiting to update. **Comments**

3 Multi-objective Evolutionary Algorithm based on Decomposition

3.1 MOEA/D [2]

There are two main lines in the MOEAD related works [4].

We will concentrate on the selection of weighted vectors next week.

3.2 MOEA/D-M2M [11]

MOEA/D-M2M decomposes the multi-objective problems(MOP) into several small MOPs. The decision space is divided by k unit direction vectors into k regions. Compared to MOEA/D, it can promote the diversity of the population.

3.3 MOEA/D-VLP [8]

Motivated by metameric genetic algorithm(MGA) [13], Test problems proposed by Lihui were all composed by the shape function $\alpha_i(x_I)$ plus distance function $\beta_i(x_{II})$. The challenge of these kind of test problems I found is that MOEA/D, MOEA/D-M2M and NSGA-II in approximating the higher dimension.

MOEA/D-VLP combined the idea of MOEA/D and MOEA/D-M2M, but not the same. It decompose the original MOP into small MOPs according to the dimension.

Meanwhile, I read Zhenkun's article [9] about the construction of the shape function $h_k(x_I)$ and distance function $g_k(x_{II})$.

3.4 MOEAD-DE and test problems with complicated pareto sets [14]

This article was motivated by the influence of PS shapes over the performance of MOEAs. Li and Zhang proposed multiobjective test problems with prescribed pareto set.

A multiobjective optimization problem (MOP) can be stated as follows:

$$\begin{aligned} & minimize \quad F(x) = (f_1(x), \ldots, f_m(x)) \\ & s.t. \quad x \in \Omega \end{aligned} \tag{3.1}$$

The decision space:
$\Omega = \prod_{i=1}^{n} [a_i, b_i] \subset R^n$, where $-\infty < a_i < b_i < +\infty$ for all $i = 1, \ldots, n$.
Its m objectives to be minimized take the following form:

$$\begin{aligned} f_1(x) =& \alpha_1(x_I) + \beta_1(x_{II} - g(x_I)) \\ & \vdots \\ f_m(x) =& \alpha_m(x_I) + \beta_m(x_{II} - g(x_I)) \end{aligned} \tag{3.2}$$

where

- $x = (x_1, \ldots, x_n) \in \Omega, x_I = (x_1, \ldots, x_{m-1})$ and $x_{II} = (x_m, \ldots, x_n)$ are two subvectors of x;
- $\alpha_i (i = 1, \ldots, m)$ are functions from $\prod_{i=1}^{m-1} [a_i, b_i]$ to R ;
- $\beta_i (i = 1, \ldots, m)$ are functions from R^{n-m+1} to R^+;
- g is a function from $\prod_{i=1}^{m-1} [a_i, b_i]$ to $\prod_{i=m+1}^{n} [a_i, b_i]$.

Based on this article, UF1-UF10 were proposed in [15]. Apart from common PF shape like linear, convex or concave in bi-objective test problems and semi-sphere in tri-objective test problems, the PF of UF5, UF6 and UF9 are more complicated.

3.5 MOEA/D-GRA [16]

This article introduced a generalized resource allocation strategy to assign computational resource to subproblems in MOEA/D using a probability of improvement vector. Details in this article I will study with the code in this week.

3.6 the research work in the next two months

3.6.1 Background

In a sensor converge problem, sensors are placed in a bounded domain A and the objective of this problem is to maximize the coverage and minimizing the cost. The sensing ranges of different sensors are various. The larger the sensing radius r_i, the higher the cost. So the number of the sensors placed in the bounded area is variable. M.L. Ryerkerk et al. [13] combined these two objectives together by a single weighted sum objective function and solved it with metameric genetic algorithm(MGA) based on the variable-length genome. Ting et al. [18] present a multiobjective variable-length genetic algorithm to solve this problem.

This kind of variable-length optimization with varying number of parameters strikes us to consider: how will we deal with two or more kinds of designing problems with different number of parameters under the identical criteria? In other words, how will we choose a suitable plan or co-worked scheme among various projects.

This is the reason why we concentrated on this problem. (I will continue to think a more specific example.)

3.6.2 Test Problems

More details will be added in this week.

$$\begin{aligned} MOP_1 \quad & minimize \quad F_1(x) = \left(f_1^1(x), \ldots, f_m^1(x)\right) \\ & s.t. \quad x \in \Omega_1 \\ & \vdots \\ MOP_K \quad & minimize \quad F_K(x) = \left(f_1^K(x), \ldots, f_m^K(x)\right) \\ & s.t. \quad x \in \Omega_K \end{aligned} \tag{3.3}$$

The decision space of the $k-$th MOP:

$\Omega_k = \prod_{i=1}^{n_k} [a_i, b_i] \subset R^{n_k}$, where $-\infty < a_i < b_i < +\infty$ for all $i = 1, \ldots, n_k$, for all k=1, . . . , K.

m objectives in MOP_k to be minimized take the following form:

$$\begin{aligned} f_1^k(x) =& \alpha_1^k\left(x_I\right)+\beta_1^k\left(x_I, x_{II}\right) \\ &\vdots \\ f_m^k(x) =& \alpha_m^k\left(x_I\right)+\beta_m^k\left(x_I, x_{II}\right) \end{aligned} \tag{3.4}$$

where

- $x=\left(x_1, \ldots, x_{n_k}\right) \in \Omega_k, x_I=\left(x_1, \ldots, x_{m-1}\right)$ and $x_{II}=\left(x_m, \ldots, x_{n_k}\right)$ are two subvectors of x;
- $\alpha_i(i=1, \ldots, m)$ are functions from $\prod_{i=1}^{m-1}\left[a_i, b_i\right]$ to R ;
- $\beta_i(i=1, \ldots, m)$ are functions from R^{n_k-m+1} to R^{+};
- g is a function from $\prod_{i=1}^{m-1}\left[a_i, b_i\right]$ to $\prod_{i=m+1}^{n_k}\left[a_i, b_i\right]$.

3.6.3 Evaluation Criteria

Waiting to update.

3.6.4 the framework of the proposed algorithm

Waiting to update.

References

[1] Hanneke S. The optimal sample complexity of PAC learning[J]. The Journal of Machine Learning Research, 2016, 17(1): 1319-1333.

[2] Zhang Q, Li H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition[J]. IEEE Transactions on evolutionary computation, 2007, 11(6): 712-731.

[3] Li K, Fialho A, Kwong S, et al. Adaptive operator selection with bandits for a multiobjective evolutionary algorithm based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2013, 18(1): 114-130.

[4] Trivedi A, Srinivasan D, Sanyal K, et al. A survey of multiobjective evolutionary algorithms based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2016, 21(3): 440-462.

[5] Janner M, Fu J, Zhang M, et al. When to trust your model: Model-based policy optimization[C]//Advances in Neural Information Processing Systems. 2019: 12519-12530.

[6] Li K, Malik J. Learning to optimize[J]. arXiv preprint arXiv:1606.01885, 2016.

[7] Lattimore T, Hutter M, Sunehag P. The sample-complexity of general reinforcement learning[C]//Proceedings of the 30th International Conference on Machine Learning. Journal of Machine Learning Research, 2013.

[8] Li H, Deb K, Zhang Q. Variable-length Pareto optimization via decomposition-based evolutionary multiobjective algorithm[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(6): 987-999.

[9] Z. Wang, Y. Ong and H. Ishibuchi, ”On Scalable Multiobjective Test Problems With Hardly Dominated Boundaries,” in IEEE Transactions on Evolutionary Computation, vol. 23, no. 2, pp. 217-231, April 2019, doi: 10.1109/TEVC.2018.2844286.

[10] Chi-Heng Lin, Joseph D. Miano, Eva L. Dyer: “Bayesian optimization for modular black-box systems with switching costs”, 2020; [http://arxiv.org/abs/2006.02624 arXiv:2006.02624].

[11] H. Liu, F. Gu and Q. Zhang, ”Decomposition of a Multiobjective Optimization Problem Into a Number of Simple Multiobjective Subproblems,” in IEEE Transactions on Evolutionary Computation, vol. 18, no. 3, pp. 450-455, June 2014, doi: 10.1109/TEVC.2013.2281533.

[12] Handoko S D, Nguyen D T, Yuan Z, et al. Reinforcement learning for adaptive operator selection in memetic search applied to quadratic assignment problem[C]//Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation. 2014: 193-194.

[13] Ryerkerk, M.L., Averill, R.C., Deb, K. et al. Solving metameric variable-length optimization problems using genetic algorithms. Genet Program Evolvable Mach 18, 247–277 (2017). https://doi.org/10.1007/s10710-016-9282-8.

[14] Li H, Zhang Q. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II[J]. IEEE transactions on evolutionary computation, 2008, 13(2): 284-302.

[15] Zhang Q, Zhou A, Zhao S, et al. Multiobjective optimization test instances for the CEC 2009 special session and competition[J]. 2008.

[16] Zhou A, Zhang Q. Are all the subproblems equally important? Resource allocation in decomposition-based multiobjective evolutionary algorithms[J]. IEEE Transactions on Evolutionary Computation, 2015, 20(1): 52-64.

[17] Deb K, Thiele L, Laumanns M, et al. Scalable test problems for evolutionary multiobjective optimization[M]//Evolutionary multiobjective optimization. Springer, London, 2005: 105-145.

[18] Ting C K, Lee C N, Chang H C, et al. Wireless heterogeneous transmitter placement using multiobjective variable-length genetic algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2009, 39(4): 945-958.