Contextual Stochastic Search

Abbas Abdolmaleki
IEETA, University of Aveiro
DSI, University of Minho
LIACC, University of Porto
Portugal
abbas.a@ua.pt

Nuno Lau
DETI, IEETA, University of Aveiro, Portugal
nunolau@ua.pt

Luis Paulo Reis
DSI, University of Minho
LIACC, University of Porto
Portugal
lpreis@dsi.uminho.pt

Gerhard Neumann
CLAS, TU Darmstadt,
Darmstadt, Germany
neumann@ias.tu-darmstadt.de

ABSTRACT

Many stochastic search algorithms require relearning if the task changes slightly to adapt the solution to the new situation or the new context. Therefore in this research, we investigate the contextual stochastic search algorithms that can learn from multiple tasks simultaneously. Here, we want to find good parameter vectors for multiple related tasks, where each task is described by a continuous context vector. Hence, the objective function might change slightly for each parameter vector evaluation.

1. INTRODUCTION

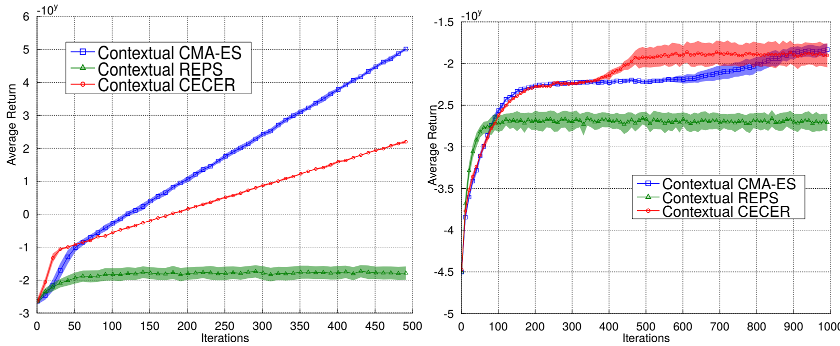
Stochastic search algorithms are gradient-free black-box optimizers of some performance function dependent on a high dimensional parameter vector. Here, we directly evaluate the execution of a parameter vector by using the return of an episode. Stochastic search algorithms typically maintain a search distribution over the parameters that we want to optimise. This search distribution is used to create samples of the parameter vector. Subsequently, the performance of the sampled parameters is evaluated. Using the samples and their evaluations, a new search distribution is computed. However, many of stochastic search algorithms can not be applied for multi-task learning. Therefore if the task setup changes slightly, relearning is needed to adapt the solution to the new situation or the new context. For example, consider optimising the parameters of a robot arm controller to pick up an object. Once the characteristics of the object, such as weight or material, changes, relearning is needed. In order to generalize a solution for a context to the other contexts, for example, picking up an object with different weights, the parameters can be optimized for several target contexts independently. Subsequently, regression methods can be used to generalize the optimized contexts to a new, unseen context. However such approaches are time consuming and inefficient in terms of the number of needed training samples as optimizing for different contexts and the generalization between optimized parameters for different contexts, are two independent processes. Hence, we cannot reuse data-points obtained from optimizing a task with context $\boldsymbol{s}$ to improve and accelerate the optimization of a task with another context $\boldsymbol{s}'$. Hence, it is desirable to learn the selection of the parameter for multiple tasks at once without restarting the learning process once we see a new task. This problem setup is also known as contextual policy search [1]. Recently, such multi-task learning capability was established for information-theoretic policy search algorithms [2], such as the episodic Relative Entropy Policy Search (REPS) algorithm [1]. However, as many other stochastic search algorithms, which update the search distribution, the search distribution used in contextual REPS might collapse prematurely to a point-estimate, resulting in premature convergence. In order to alleviate this problem, we use a recently proposed regularization technique for updating covariance matrices called CECER [3] for updating the covariance matrix of contextual REPS. CECER has been shown to be highly competitive to other stochastic optimizers. We call the resulting algorithm contextual CECER. On the other hand, competing stochastic search algorithms, such as CMA-ES [4] and NES, and commonly used policy search methods are lacking this important contextual feature. Therefore, we also extend covariance matrix adaptation algorithm (CMA-ES) which is the state of art algorithm in stochastic search to be applicable for contextual setup which we refer to contextual CMA-ES. In order to do that, we maintain a context dependent search distribution, where the mean of this distribution now depends on the context. We use a linear parametrization for the mean, i.e., the mean linearly depends on a given context feature vector. In each iteration we update the weights of this linear function as well as covariance matrix of the search distribution. We compare these two algorithms along with standard contextual REPS on standard functions.

1.1 Problem Statement

Given a context vector $\boldsymbol{s}$ which defines a task, we want to find a function $\boldsymbol{m}(\boldsymbol{s}) : \mathbb{R}^m \to \mathbb{R}^n$ that outputs a parameter vector $\boldsymbol{\theta}$ with n dimensions such that it maximizes an objective function $R(\boldsymbol{\theta}, \boldsymbol{s}) : \{\mathbb{R}^n, \mathbb{R}^m\} \to \mathbb{R}$. The only accessible information on $R(\boldsymbol{\theta}, \boldsymbol{s})$ are evaluations $\{R^{[k]}\}_{k=1\ldots N}$ of samples $\{\boldsymbol{s}^{[k]}, \boldsymbol{\theta}^{[k]}\}_{k=1\ldots N}$, where k is the index of the sample and N is number of samples. We maintain a search distribution $\pi(\boldsymbol{\theta}|\boldsymbol{s})$ over the parameter space $\boldsymbol{\theta}$ of the objective function $R(\boldsymbol{\theta}, \boldsymbol{s})$. The search distribution $\pi(\boldsymbol{\theta}|\boldsymbol{s})$

GECCO '16 July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07.
DOI: http://dx.doi.org/10.1145/2908961.2909012



(a) Sphere (b) Rastrigin

Figure 1: The performance comparison of stochastic search methods for optimising contextual version of standard functions (a)Sphere and (b)Rastrigin, The results show that while both contextual CECER and contextual CMA-ES perform well, Contextual REPS suffers from premature convergence.

is modeled as linear Gaussian policy, i.e.,

$$\pi(\boldsymbol{\theta}|\boldsymbol{s}) = \mathcal{N}\left(\boldsymbol{\theta}|\boldsymbol{m}_\pi(\boldsymbol{s}) = \boldsymbol{A}_\pi^T\varphi(\boldsymbol{s}), \Sigma_\pi\right),$$

where $\varphi(\boldsymbol{s})$ is an arbitrary feature function of context $\boldsymbol{s}$. In each iteration, a new coefficient matrix $\boldsymbol{A}_\pi$ and a new covariance matrix Σ_π is obtained. Typically $\boldsymbol{\varphi}(\boldsymbol{s}) = [1 \quad \boldsymbol{s}]$, which results in linear generalization over contexts. However we could use non-linear feature functions such as radial basis functions(RBF) for non linear generalization over contexts [5]. In each iteration, given context samples $\boldsymbol{s}^{[k]}$[1], the current search distribution $q(\boldsymbol{\theta}|\boldsymbol{s})$ is used to create samples $\boldsymbol{\theta}^{[k]}$ of the parameter vector $\boldsymbol{\theta}$. Subsequently, the evaluation $R^{[k]}$ of $\{\boldsymbol{s}^{[k]}, \boldsymbol{\theta}^{[k]}\}$ is obtained by querying the objective function $R(\boldsymbol{\theta}, \boldsymbol{s})$. Now the samples $\{\boldsymbol{s}^{[k]}, \boldsymbol{\theta}^{[k]}, R^{[k]}\}_{k=1\ldots N}$ are used to compute a weight $d^{[i]}$ for each sample k. Subsequently, using $\{\boldsymbol{s}^{[k]}, \boldsymbol{\theta}^{[k]}, d^{[k]}\}_{k=1\ldots N}$, a new Gaussian search distribution $\pi(\boldsymbol{\theta}|\boldsymbol{s})$ is estimated. This process will run iteratively until the algorithm converges to a solution.

2. EXPERIMENTS

We compare contextual CECER, contextual CMA-ES and the standard contextual REPS algorithms. we use standard optimization test functions, such as the Sphere and the Rastrigin (multi modal) function. We extend these functions to be applicable for contextual setting. The task is to find the optimum 15 dimensional parameter vector $\boldsymbol{\theta}$ for a given 1 dimensional context $\boldsymbol{s}$. We show the average as well as two times the standard deviation of the results over 10 trials for each experiment. Note that the y-axis of all plots is in a logarithmic scale.

2.1 Standard Optimization Test Functions

We chose two standard optimization functions which are the Sphere function $f(\boldsymbol{s}, \theta) = \sum_{i=1}^{p} \boldsymbol{x}_i^2$ and a multi-modal function which is known as the Rastgirin function $f(\boldsymbol{s}, \theta) = 10p + \sum_{i=1}^{p}[\boldsymbol{x}_i^2 - 10\cos(2\pi\boldsymbol{x}_i)]$. Where $\boldsymbol{x} = \boldsymbol{\theta} + \boldsymbol{A}\boldsymbol{s}$. The matrix A is a constant matrix that was chosen randomly. In our case, because the context $\boldsymbol{s}$ is 1 dimensional, A is a $n \times 1$ dimensional vector where n is the dimension of parameter space $\boldsymbol{\theta}$. Now, the optimum $\boldsymbol{\theta}$ for these functions is linearly dependent on the given context $\boldsymbol{s}$. The initial search area of $\boldsymbol{\theta}$ for all experiments is restricted to the hypercube $-5 \leq \boldsymbol{\theta}_i \leq 5, i = 1, \ldots, n$ and contexts are samples uniformly from interval $0 \leq \boldsymbol{s}_i \leq 3, i = 1, \ldots, m$ where m is dimension of the context space $\boldsymbol{s}$. In our experiments, the mean of the initial distributions have been chosen randomly in the defined search area.

Algorithmic Comparison.

We compared contextual CECER, contextual CMA-ES and standard contextual REPS. In each iteration, we generated 50 new samples. The results in figure 1 shows that both contextual CMA-ES and contextual CECER could successfully learn the contextual tasks while standard contextual REPS suffers from premature convergence.

3. CONCLUSION

Stochastic search methods such as evolutionary strategies, e.g. CMA-ES, have been employed extensively for black box optimization. However, these algorithms fail to generalize the optimized parameters to related tasks. Therefore, in this research, we investigated contextual stochastic search methods for multi task learning.

4. REFERENCES

[1] A. Kupcsik, M. P. Deisenroth, J. Peters, and G. Neumann. Data-Efficient Contextual Policy Search for Robot Movement Skills. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2013.

[2] J. Peters, K. Mülling, and Y. Altun. Relative Entropy Policy Search. In *Proceedings of the 24th National Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2010.

[3] A. Abdolmaleki, N. Lua, L.P. Reis, and G. Neumann. Regularized covariance estimation for weighted maximum likelihood policy search methods. In *Proceedings of the International Conference on Humanoid Robots (HUMANOIDS)*, 2015.

[4] N. Hansen, S.D. Muller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, 2003.

[5] A. Abdolmaleki, N. Lua, L.P. Reis, J. Peters, and G. Neumann. Contextual Policy Search for Generalizing a Parameterized Biped Walking Controller. In *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2015.

[1]Please note that the way we sample contexts $\boldsymbol{s}^{[k]}$ depends on the task. We consider scenarios where the context vector changes for each sample. Typically we use a uniform distribution to sample contexts $\boldsymbol{s}^{[k]}$.