SCIENCE CHINA
Information Sciences

• LETTER •

December 2018, Vol. 61 129205:1–129205:3
https://doi.org/10.1007/s11432-018-9521-8

Incremental data-driven optimization of complex systems in nonstationary environments

Cuie YANG[1], Jinliang DING[1*], Yaochu JIN[1,2*] & Tianyou CHAI[1]

[1]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang* 110004*, China;*
[2]*Department of Computer Science, University of Surrey, Guildford* GU2 7XH*, UK*

Received 16 June 2018/Accepted 29 June 2018/Published online 20 November 2018

Citation Yang C E, Ding J L, Jin Y C, et al. Incremental data-driven optimization of complex systems in nonstationary environments. Sci China Inf Sci, 2018, 61(12): 129205, https://doi.org/10.1007/s11432-018-9521-8

Dear editor,

In a complex system, many real-world optimization problems do not have explicit optimization functions or constraint functions, whereas only data from production processes are available because of the complexity of the system [1]. Thus, optimization in this scenario can generally rely on collected historical datasets, and these problems are also known as offline data-driven optimization problems [2]. In the last decade, studies on offline data-driven optimization have been conducted by first constructing a surrogate model with the collected data and then considering the best solution from the surrogate model as the optimal decision [1]; few studies, however, have considered the errors existing in surrogate models. In addition, production processes are often subject to changes, leading to the attainment of data that are non-independent identically distributed. In this study, we aim to deal with the offline data-driven optimization problem for data generated in nonstationary systems in an incremental manner; specifically, we assume the data come in a chunk as shown in Figure A1 in Appendix A. This type of data-driven optimization poses new challenges to the current data-driven optimization algorithms [2]. We focus on the following three issues. The first is how to build a high-quality surrogate model for each environment. The second issue concerns optimization, i.e., how to quickly exploit the optimal solution of each new environment. The last issue is the final solution creation, which is necessary because errors exist between surrogate models and corresponding real fitness functions (the unknown formulations of real systems), resulting in less reliable solutions from the surrogate models. To alleviate the above difficulties, this study suggests a general method described as follows.

While new data chunk D_t is income from complex systems at the t-th environment, do Steps 1–4:

Step 1. Update the surrogate model based on knowledge transfer technique to adapt the t-th environment;

Step 2. Initialize population based on historical knowledge;

Step 3. Optimize the surrogate model by using differential evolution (DE) algorithm;

Step 4. Produce the final solution for complex systems.

Approach. A general framework of the proposed approach is described as above. Upon the observation of each new data chunk from complex systems, the surrogate model is first updated to adapt to the new environment. A DE algorithm [3], a type of population-based global optimization approach, is adopted as the optimizer. Thus, the

* Corresponding author (email: jlding@mail.neu.edu.cn, yaochu.jin@surrey.ac.uk)

next step is population initialization for the DE algorithm. Then, the DE algorithm with the rand/1 strategy is applied to optimize the surrogate model to obtain an optimal landscape of the real fitness problems. Lastly, the final solution is generated from the obtained optimal landscape. Details of the surrogate model update, population initialization, and final solution production are presented as follows.

• Knowledge transfer-based surrogate model adaptation. Ensemble approaches are a popular means of handling incremental learning, which uses models or training instances of historical environments to improve model quality in the current environment [4]. Nevertheless, most work on incremental learning has focused on classification tasks [4]. In the optimization problem outlined in this study, we introduce an ensemble approach for regression tasks to formulate the surrogate model. Specifically, a base regression learner is first trained via the new data chunk D_t, equivalently $(\boldsymbol{x}_t, \boldsymbol{y}_t)$, and denoted by h_t. The radial basis neural network (RBFN) is applied as a base learner in this study because of its universal approximation ability [5]. Then, a set of base learners trained by using the data chunks of each past environment is also constructed separately. To improve the adaptability of past training instances, we first map the historical data chunks $D_1, D_2, \ldots, D_{t-1}$ to the current data chunk space D_t, thereby facilitating knowledge transfer between the historical data set and the current data set. We then use a combination of each of the transferred historical data chunk D_{ni} and the current data chunk D_t to build each historical base surrogate model h_i. Note that we are interested in the dynamics of the system, which would result in a change in the function of the value $\boldsymbol{y}_t$. Thus, we transform $\boldsymbol{y}_i$, $i = 1, 2, \ldots, t-1$ in each D_i to the current objective space $\boldsymbol{y}_t$ by using (1). We define the transferred D_i, which contains $(\boldsymbol{x}_i, \boldsymbol{y}_{ni})$, as D_{ni}, $i = 1, 2, \ldots, t-1$.

$$\boldsymbol{y}_{ni} = \frac{\boldsymbol{y}_i - y_i^{\min}}{y_i^{\max}} - y_i^{\min} \times (y_t^{\max} - y_t^{\min}) + y_t^{\min}, \quad (1)$$

where $y_i^{\max}$ and $y_i^{\min}$ are the maximum and minimum values of $\boldsymbol{y}_i$, respectively, $y_t^{\max}$ and $y_t^{\min}$ are the maximum and minimum values of $\boldsymbol{y}_t$, respectively.

In the next step, all the base surrogate models h_i, $i = 1, 2, \ldots, t$ are integrated into the final perfect surrogate f using

$$f = \frac{\sum_{i=1}^{t} w_i h_i}{\sum_{i=1}^{t} w_i}, \quad (2)$$

where $w_i = \frac{1}{\mathrm{RMSE}_i + \mathrm{RMSE}_t}$, $i = 1, 2, \ldots, t-1$, $w_t = \frac{1}{\mathrm{RMSE}_t}$ and $\mathrm{RMSE}_i = \sqrt{\frac{1}{|D_i|} \Sigma_{y \in \boldsymbol{y}_i} (\hat{y} - y)^2}$, $i = 1, 2, \ldots, t$.

In this ensemble, we assign a larger w_t than w_i, $i = 1, 2, \ldots, t$ to ensure a higher weight for the base model in the current environment. This is because the data set of the current environment is more reliable, and thus it should be fully used.

• A priori knowledge-based population initialization. After the surrogate model of the current environment is obtained, an initial population should be created before surrogate model optimization is started. The initial population is often randomly generated in the decision space in traditional DE algorithms. In reality, the surrogate models of different environments are not isolated because their training instances belong to the same system. Therefore, using the historical knowledge of the past environments in the population initialization would benefit convergence of the current environment. For simplicity, the candidates of the latest environment are applied as the initial population in this study. Note that the initial population is randomly generated for the first environment because there is no historical information at the beginning.

• Top best solution averaging-based final solution production. As mentioned above, solutions obtained during optimization are not allowed to be evaluated by true complex problems; instead, they are only evaluated by surrogate models. In this case, it is interesting to create a high-quality final solution for real fitness functions because no surrogate can be updated using the real fitness function, and the fitness value of a solution evaluated by surrogates may contain large errors compared to that evaluated by real fitness problems. This study proposes a top best solution averaging technique to generate the final solution for a real fitness function instead of directly using the best solution of the obtained candidates. Specifically, in the final population of each environment, the average of the top 10% best individuals is considered as the final solution. In this manner, the errors of the final solution induced by surrogates can be smoothed by consulting a number of candidates.

Experimental results. The six dynamic optimization benchmark problems [6] are applied to examine the transferred surrogate model construction, population initialization, and final solution production strategies. The number of decision variables D of each problem is set to 10. The total number of environments in each test problem is set to 50. In each environment, $3D$ points generated by Laplace sampling and evaluated by the real fitness function are taken as the historical dataset, where D is the number of decision variables. The experiment is conducted on different approaches of incremental data-driven optimization in nonstationary environments to verify each of the proposed techniques: SS (single dataset-based surrogate model construction technique), KTS (knowledge transfer-based surrogate model construction), KTSPI (the version of KTS by inducing a prior knowledge-based population Initialization), and KTTLSA-TBA (the KTSPI algorithm with top best solution averaging-based final solution production technique).

Table 1 Average results of 50 environments over 20 independent runs

Name	SS	KTS	KTSPI	KTSPI-TBA
F1	3.7472	3.4762	3.0220	2.9770
F2	578.2938	546.8831	528.7354	528.9487
F3	1.121E+03	1.094E+03	1.077E+03	1.074E+03
F4	646.0489	610.8318	592.6871	589.7063
F5	2.021E+03	2.016E+03	1.999E+03	1.997E+03
F6	2.341E+03	1.399E+03	1.317E+03	1.314E+03

Table 1 presents the average results of the 50 environments over 20 independent runs of the compared algorithms. We can see from this table that the values obtained by SS, KTS, KTSPI, and KTSPI-TBA decrease in general, indicating the suitability of the knowledge transfer-based surrogate model update technique, historical knowledge population initialization technique, and averaging-based final solution production technique. The results of each environment over 20 independent runs of the compared algorithms on F1 and F5 is presented in Figure A2 in Appendix A. This figure clearly shows that KTSPI and KTSPI-TBA outperform SS and KTS in most of the experiments on F1. With regards to the F5 problem, we find that KTS clearly outperforms SS in almost all experiments; in addition, it can also be found that the KTSPI-TBA algorithm achieves a more robust performance than KTSPI in the experiment.

Conclusion. This study proposes a general method to solve data-driven problems in nonstationary environments, which includes three techniques, i.e., knowledge transfer-based surrogate model adaptation technique, historical knowledge-based population initialization technique, and top best solution averaging-based final solution production technique. We systematically compare each technique by applying the 4 proposed incremental data-driven optimization approaches to 6 benchmark problems. The statistical results revealed that each strategy exhibited good performance in addressing the incremental offline data-driven problem in dynamic environments.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61525302, 61590922), Project of Ministry of Industry and Information Technology of China (Grant No. 20171122-6), Projects of Shenyang (Grant No. Y17-0-004), Fundamental Research Funds for the Central Universities (Grant Nos. N160801001, N161608001), and Outstanding Student Research Innovation Project of Northeastern University (Grant No. N170806003).

Supporting information Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Yang C E, Ding J L. Constrained dynamic multi-objective evolutionary optimization for operational indices of beneficiation process. J Intel Manuf, 2017. doi: 10.1007/s10845-017-1319-1
2 Wang H D, Jin Y C, Jansen J O. Data-driven surrogate-assisted multiobjective evolutionary optimization of a trauma system. IEEE Trans Evol Comput, 2016, 20: 939–952
3 Storn R, Price K. Differential evolutionary simple and efficient heuristic for global optimization over continuous spaces. J Global Optim, 1997, 11: 341–359
4 Ditzler G, Roveri M, Alippi C, et al. Learning in nonstationary environments: a survey. IEEE Comput Intel Mag, 2015, 10: 12–25
5 Park J, Sandberg I W. Universal approximation using radial-basis-function networks. Neural Comput, 1991, 3: 246–257
6 Li C H, Yang S X, Nguyen T T, et al. Benchmark Generator for CEC'2009 Competition on Dynamic Optimization. IEEE Congress on Evolutionary Computation Technical Report, 2008