Expert Systems With Applications 96 (2018) 25–48

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Empirical studies of Gaussian process based Bayesian optimization using evolutionary computation for materials informatics

Hiroshi Ohno

Toyota Central R&D Labs., Inc., 41-1 Yokomichi, Nagakute, Aichi 480-1192, Japan

ARTICLE INFO

Article history:
Received 10 July 2017
Revised 9 October 2017
Accepted 10 November 2017
Available online 15 November 2017

Keywords:
Bayesian optimization
Markov random search
Gaussian process
Ensemble
Evolutionary computation
Materials informatics

ABSTRACT

Using evolutionary computation, we empirically investigate convergence properties of Gaussian process based Bayesian optimization (BO). We use evolutionary computation for the learning of the prediction model and optimization of the acquisition function (auxiliary search) of BO. For practical use for materials informatics, we address three issues in BO: (1) the stopping conditions, (2) the initial data size, and (3) the unknown smoothness of the target function. Then, we introduce a goal-directed acquisition function in which a target value as a desired property of a compound is incorporated. In addition, we present an ensemble method of BO, in which each BO in the ensemble has a random property and a kernel function with a different smoothness. Experimental results for the materials data sets on melting points of binary compounds and hydrogen weight percentages of hydrogen storage materials with two to four constituent elements show the effectiveness of the ensemble method of BO. Additionally, using an ensemble of BOs presents that the obtained results (increase of the number of samples acquired) are not simply a result of additional BOs. The goal-directed acquisition function and the ensemble of BOs which we propose should be techniques that can be used in the realization of a new materials recommendation system with a self-learning algorithm. Due to the self-learning algorithm realized by BO, the property prediction performance of the algorithm would be increasingly improved.

1. Introduction

Recently, Bayesian optimization (BO) has been a promising tool in the search for materials (Frazier & Wang, 2016). Within the limited search space, BO is successful at finding the best materials (e.g., Seko, Maekawa, Tsuda, & Tanaka, 2014; Seko, Hayashi, Nakayama, Takahashi, & Tanaka, 2017). In Seko et al. (2014), BO was highly effective at finding, within a limited search space, the material with the best melting point as compared with a random search. Furthermore, for a larger limited search space, Ju et al. (2017) demonstrated the efficiency of BO in the optimization of nanostructures consisting of either Si or Ge, related to thermal conductance. It is however not clear how to determine the initial data size of BO in the simulation experiments in these studies. There has also been no discussion relating to the influence of the initial data size on the convergence of BO.

For a materials search, it is not possible to know the suitable size of the initial data set at the start of the search in BO. In a new materials search, in general, there are a little data from a pilot study and domain knowledge. In addition, in the literature of materials science, the data are often obtained from well-performed (successful) experiments. Therefore, the data sets from the literature may have sampling bias. In this study, we address the issue of the initial data size of BO.

In the machine learning and optimization literature, we can see an enormous number of studies of BO which are related to applications and theory (see Shahriari, Swersky, Wang, Adams, and de Freitas, 2016, for a detailed overview). In the present paper, we address a Gaussian process (GP) based BO, which is a popular method in BO. BO has mainly two ingredients: a prediction model and an acquisition function. The prediction model is typically realized by GP regression (Rasmussen & Williams, 2006). In addition, the kernel function of the prediction model determines the smoothness of the model pertaining to the target function of BO which is unknown. The smoothness affects the convergence of the regret of BO (Schulz, Speekenbrink, Hernández-Lobato, Ghahramani, & Gershman, 2016; Srinivas, Krause, Seeger, & Kakade, 2010). The prediction model also has hyper-parameters such as the length scale of the squared exponential kernel function, which is learned from data. During a search process which consists of exploration and exploitation processes by BO, as the number of samples acquired by the search process increases, learning of the model is needed in order to predict the output of the model precisely in the exploitation phase. Therefore, it is desirable that the convergence of the train-

E-mail address: oono-h@mosk.tytlabs.co.jp

https://doi.org/10.1016/j.eswa.2017.11.026
0957-4174/

ing algorithm of the model should be stable under an increasing number of samples acquired and that its computational demand be low.

In this study, we adopt evolutionary computation such as CMA-ES (Hansen & Ostermeier, 2001) for the hyper-parameter tuning. Evolutionary computation has characteristics that are easy to use for practitioners in materials science: a derivative-free and an algorithm with either easy-to-adjust control parameters or no control parameters (e.g., CMA-ES and (1+1)-ES; Auger and Hansen, 2011, chap. 10).

While the study of BO has been steadily growing and expanding, some issues remain with BO (see a summary in Appendix A). For practical use for materials informatics, we mainly address the smoothness of the prediction model pertaining to the target function, the stopping conditions for the search, and the auxiliary search cost. For the stopping conditions for the search, we introduced a goal-directed acquisition function in BO. The goal-directed acquisition function contains the desired output whose value is set by practitioners. In particular, for materials science, this could be useful in the case of the search for alternative materials.

Furthermore, we adopt an ensemble search of BOs in order to deal with the smoothness of the prediction model and facilitate the learning of the models included in the ensemble. Unlike in other studies (e.g., Hutter, Hoos, & Leyton-Brown, 2012), in our study, the ensemble search consists of many workers (hereafter, searchers) with a shared data set that are not independent of each other. Because the data set is common across the searchers, a Gram matrix (kernel matrix) of GP regression of each searcher is constructed.

A knowledge base and an inference mechanism are the main ingredients in expert and intelligent systems. In materials science, there is insufficient prior work on novel situations (e.g., prediction of chemical reactions and new materials search) to apply the intelligent systems with a predetermined knowledge base and a predefined inference logic (e.g., Kayala & Baldi, 2011). Thus, we attack a new materials search using a self-learning algorithm (see our related work in Miwa & Ohno, 2017). The goal-directed acquisition function and the ensemble of BOs which we propose should be useful techniques that can be used in the realization of a new materials recommendation system with a self-learning algorithm.

Experimental results for learning on synthetic and benchmark data sets reveal that a Markov random search method (Tikhomirov, 2011) for the learning of the model parameters had reasonably good performance compared to other method such as kernel partial least squares (PLS) (Rosipal & Trejo, 2001). The Markov random search method that we use[1] is a simplified variant of CMA-ES, which has few control (hyper-) parameters for the algorithm and would be easy for practitioners in materials science to use.

In the experiments with a small initial data set (even one sample) for the case of materials data sets (melting points of binary compounds and hydrogen weight percentages of hydrogen storage materials), we adopted a random error insertion in the search such as an ϵ-greedy strategy (Sutton and Barto, 1998, Section 2.2) in order to improve the convergence speed. From the experimental results on materials data sets, we confirmed the effectiveness of the random error insertion. In the case of ensemble search of BOs, it was observed that using an ensemble of BOs facilitated an increase of samples acquired, compared with a single BO.

Our contributions on the practical use for materials informatics are outlined as follows:

- We present a goal-directed acquisition function with regard to the stopping conditions, that is, the desired output value which is the target in the optimization. The goal-directed acquisition function is useful for searching for alternative materials in terms of setting a desired property of the materials.
- We show the usefulness of a random search in the cases of a small initial data set (including a size-one data set as the extreme case).
- We confirm the effectiveness of the ensemble search for handling a mismatch between the smoothness of the prediction model and that of the target function, and facilitating the learning of the models included in the ensemble.
- We show the convergence guarantee (regret bound) for the goal-directed acquisition function and the lower bound of the minimum number of search steps based on the desired and initial outputs of the prediction model on the Markov random search.

The paper is organized as follows. In Section 2, we survey the related work regarding evolutionary computation and BO. We then describe the issues of previous studies with respect to searching for materials. In Section 3, the problem setting is provided in order to clarify the issues that we address with our proposed method. In Section 4, we describe the algorithms of learning of model parameters and search for BO. The kernel function in the GP regression is also defined for the purpose of dealing with materials with different numbers of elements in their compounds. The experiments for the learning of model parameters (hyper-parameter tuning) and the search on the materials data sets that we conducted are described in Section 5. Finally, the paper is concluded with a summary in Section 6.

2. Related work

In this section, we survey the related work for hyper-parameter tuning of the kernel function and the issues of BO.

2.1. Optimization and learning of model parameters

Roman, Santana, Mendiburu, and Lozano (2014) used DIRECT (Jones, Perttunen, & Stuckman, 1993) for optimization (maximization) of an acquisition function (auxiliary search). Wang, Zoghi, Hutter, Matheson, and de Freitas (2016) also used DIRECT for the hyper-parameter tuning of the kernel function in BO. Krause, Glasmachers, and Igel (2016) proposed UP-MO-CMA-ES and MO-CMA-ES for multi-objective optimization, based on CMA-ES. CMA-ES has also been used for hyper-parameter tuning in supervised learning (e.g., Friedrichs & Igel, 2005). For example, Mierswa (2006) developed evolutionary support vector machines, which were developed by combining evolution strategies and particle swarm optimization in order to solve a constrained optimization problem.

For a random search, Bergstra and Bengio (2012) described the effectiveness of hyper-parameter tuning by using a random search. Poggio, Mhaskar, Rosasco, Miranda, and Liao (2017) used a random search for the hyper-parameter search of both shallow and 2-layer binary tree networks.

Therefore, derivative-free (black box) optimization algorithms such as CMA-ES and DIRECT have been proved useful for the optimization of hyper-parameters and acquisition functions. Hence, for the present work, we adopted evolutionary computation for the hyper-parameter tuning of GP regression and the optimization of the acquisition function of BO. However, DIRECT is suitable for relatively low-dimensional search spaces (up to five dimensions) (Pošík, 2009). Since CMA-ES has the population of solutions and evaluates their fitness values, the calculation cost becomes somewhat high.

[1] This search method is often called (1+1)-ES in the evolutionary computation literature (Auger and Hansen, 2011, chap. 10).