博士論文

量子アニーリングを用いたブラックボックス最適化

(Black-Box Optimization with Quantum Annealing)

北井　孝紀

量子アニーリングを用いたブラックボックス最適化

(Black-Box Optimization with Quantum Annealing)

北井　孝紀

Department of Computational Biology and Medical Sciences,
Graduate School of Frontier Sciences

The University of Tokyo

A thesis submitted for the degree of

Doctor of Philosophy

2021

Acknowledgements

The large part of my work is done as a collaboration work with Jiang Guo, Shenghong Ju, Shu Tanaka, Koji Tsuda, Junichiro Shiomi, and Ryo Tamura. I would not have published my work without their efforts. I thank to all of them from the bottom of my heart.

I would also like to show my deepest gratitude to Koji Tsuda, professor at The University of Tokyo and my supervisor, Shu Tanaka, associate professor at Keio University, and Ryo Tamura at the National Institute for Materials Science. They gave me a chance to work with the quantum annealing machine and taught me the intuition and important ideas for my work through all the discussions. I have also learned a lot about the methodology of research from them.

I sincerely thank to the rest of my thesis committee members, Prof. Kiyoshi Asai, Prof. Hisanori Kiryu. Their valuable comments made me notice what was to be cleared in my research and gave me a confidence about my work.

I want to say thanks to my lab members, family, friends, all who supported me to keep going on my research despite all the confusions and difficulties, especially caused by the COVID-19 pandemic.

Abstract

In this thesis a new black-box optimization method, factorization machine for quantum annealing (FMQA), is proposed. The method targets at any combinatorial optimization problems. It is composed of a regression model called factorization machine (FM), and a heuristic minimization solver called quantum annealing (QA). A proof-of-principle demonstration of the FMQA's efficiency on a metamaterial designing problem is made, and it found better materials structure than the ones found with some classical algorithms.

To improve the performance for further optimization problems, the FMQA is modified with a local modeling technique. It is tested on several benchmarking problems and on an application for feature selection problem. The method is revealed to be more robust than the original FMQA for the early steps of optimizations.

Contents

1

Introduction

In science, industry, and engineering, we often have to do experiments or simulations to get objective values that are to be optimized. The process of acquiring the value is considered as an evaluation of *objective function*. The input variables are the settings of the experiments or simulations. Unlike analytical functions, evaluations of those objective functions could be much laborious and costly. Therefore, we should avoid too much number of trials and errors in that case. Additionally, we cannot expect the availability of gradient information, which is essential for gradient-descent optimization methods. Such an optimization task is called black-box optimization (BBO) because there is no information available about the inside of the function. All we can know is only the pairs of input and output values obtained by the evaluations (Figure 1.1). In general, the volume of the possible solutions' space grows exponentially to the number of input variables, and it poses a significant difficulty for finding the optimum. The success of BBO depends on how we utilize the known data to guide through the "ocean" of the possible solutions.

In the thesis, we only consider combinatorial optimization problems under the BBO setting. There are several optimization algorithms for solving it, such as greedy editing, simulated annealing, and tabu search for instance. They are classified as local search methods, in which a method has a current state as a guess for the solution, and iteratively updates it to achieve better objective values. These methods are useful for finding approximate solutions for various combinatorial optimization problems like the traveling salesman's problem (TSP). However, the fact that they require an evaluation

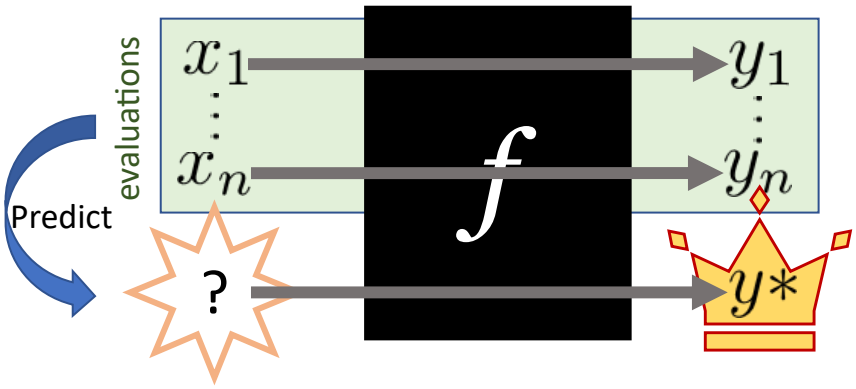


Figure 1.1: A schematic of black-box optimization (BBO). The pairs of input and output values $\{(x_i, y_i)\}_{i=1}^n$, obtained by evaluations of the black-box function f, are used for predicting the input to gain the best output value y^*.

of the target function for updating even one bit restricts their applications for expensive objective functions.

Surrogate-based methods construct regression models and use them as alternatives for the original objective functions (1). The regression models can be quickly evaluated and be optimized with the methods like the afore-mentioned greedy search or simulated annealing. The optimal point of the regression model is considered to be a good approximation to the optimum of the objective function. We evaluate the objective function on that point and append the result to the dataset to train a better regression model. That the optimization process is always taken on the regression model is the reason why it is also termed as a *surrogate model*. Selecting the point from the search space is also called *sampling*. The selection by the optimization over the surrogate model and evaluation of the objective function are continued until some convergence criterion is met. Because of the use of surrogate model, the number of required evaluations of the objective function is significantly reduced. Therefore, the surrogate-based methods are adopted for solving many real-world BBO problems.

In the thesis, a surrogate-based BBO method called factorization machine for quantum annealing (FMQA), driven by a regression model called factorization machine (FM) and an efficient search heuristic called quantum annealing (QA), will be proposed. The use of QA is expected to accelerate the selection step. We also show some applications of the method.

In the remainder of this chapter, existing methods for BBO and the theoretical background of quantum annealing is discussed. In chapter 2, the factorization machine and how to combine it with the quantum annealing into the FMQA is explained. Its proof-of-principle application for materials designing is demonstrated in chapter 3. Chapter 4 extends the FMQA with a local regression technique. We show the application of the FMQA for a drug-discovery process in chapter 5, followed by a conclusion in chapter 6. Appendix A shows the detailed result of the tests validating our algorithm described in chapter 4, and the appendix B is devoted to the description of how to use the FMQA in Python language.

1.0.1 Notations

The notation for the black-box objective function is f, and its input space is D. The dimension of D is represented by d. The dataset S of size N is composed of the pairs of input and output values $(\mathbf{x}, y)$, hence $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$.

Through the thesis, we construct regression models for the objective function, and we represent them by $\hat{f}$.

1.1 Black-box optimization

The target of BBO is an any real-valued function $f(\mathbf{x})$:$D \to \mathbb{R}$, and its minimization problem is written as

$$\min_{\mathbf{x}} \{f(\mathbf{x}) | \mathbf{x} \in D\}. \tag{1.1}$$

The formulation is so general that many optimization problems from wide range of fields fall into BBO. The evaluation of f might be doing an experiment or simulation, which could be time-consuming or cost-intensive. Example cases are planning efficient mining locations (2), designing the wing of airplanes (3), designing problems in materials informatics (4), and protein engineering (5), to name a few. There are several methods available for those problems. The choice of an appropriate method should be made depending on the objective functions' properties (e.g. evaluation cost or smoothness).

In addition to the continuous variables, recent studies shed light on BBOs over discrete or continuous-discrete mixed variables including graph structures, sequences, trees, etc. (6, 7, 8, 9, 10). We focus on BBO problems with binary input space in this thesis, and call each variable x_i of the input vector $\mathbf{x}$ especially as a *bit*.

1.1.1 Local search algorithms

Algorithm 1 Simulated Annealing (SA)

Require: $\mathbf{x}_{\text{init}} \in \{0,1\}^d$ ▷ initial solution
Require: $0 < T_{\min} < T_{\max}$ ▷ temperature range
Require: $0 < \alpha < 1$ ▷ cooling factor

procedure SIMULATEDANNEALING($\mathbf{x}_{\text{init}}, T_{\min}, T_{\max}, \alpha$)
  $\mathbf{x} \leftarrow \mathbf{x}_{\text{init}}$
  $y \leftarrow f(\mathbf{x})$
  $T \leftarrow T_{\max}$
  while $T_{\min} < T$ **do**
    $y_{\text{prev}} \leftarrow y$
    select $k \in \{1, \cdots, d\}$ randomly
    $x_k \leftarrow 1 - x_k$
    $y \leftarrow f(\mathbf{x})$
    sample r from uniform distribution over [0,1]
    if $r < \exp\left(\frac{y_{\text{prev}} - y}{T}\right)$ **then**
      $y_{\text{prev}} \leftarrow y$ ▷ accept the change
    else
      $x_k \leftarrow 1 - x_k$ ▷ rollback the change
    end if
    $T \leftarrow T \times \alpha$
  end while
  return $\mathbf{x}, f(\mathbf{x})$
end procedure

Greedy editing is the most naive method for searching optimal solutions for BBOs with binary variables. In the method, we apply single-bit flipping repeatedly as long as that improves the objective value. In a d-dimensional problem, there are d possible solutions (called *neighbors*) for the next step. All of them, or possibly a fraction of them, will be evaluated and the best one will be chosen. Although the method is very easy to implement, the search process can be very costly because of the need for evaluating all neighbors, and often results in finding a local optimum.

To avoid being caught by a local optimum, we have to sometimes allow the bit flipping yielding a degraded objective value. Simulated annealing (SA) (11, 12), origi-

Algorithm 2 Tabu Search (TS)

Require: $\mathbf{x}_{\text{init}} \in \{0,1\}^d$ ▷ initial solution
Require: $N_{\max} \in \mathbb{N}$ ▷ max evaluations
Require: $0 < l$ ▷ tabu tenure

procedure TabuSearch($\mathbf{x}_{\text{init}}, N_{\max}, l$)
  $\mathbf{x} \leftarrow \mathbf{x}_{\text{init}}$
  $L \leftarrow \underbrace{\{0, \cdots, 0\}}_{d}$ ▷ tabu list
  $n_{\text{eval}} \leftarrow 0$
  while $n_{\text{eval}} < N_{\max}$ **do**
    $k_{\text{best}} \leftarrow -1$ ▷ as a sentinel
    $y_{\text{best}} \leftarrow \infty$
    for $k \leftarrow 1, d$ **do**
      if $L_k == 0$ **then**
        $x_k \leftarrow 1 - x_k$
        $y \leftarrow f(\mathbf{x})$
        $n_{\text{eval}} \leftarrow n_{\text{eval}} + 1$
        if $y < y_{\text{best}}$ **then**
          $k_{\text{best}} \leftarrow k$
          $y_{\text{best}} \leftarrow y$
        end if
        $x_k \leftarrow 1 - x_k$ ▷ rollback
      else
        $L_k \leftarrow L_k - 1$
      end if
    end for
    $x_{k_{\text{best}}} \leftarrow 1 - x_{k_{\text{best}}}$
    $L_{k_{\text{best}}} \leftarrow l$
  end while
  return $\mathbf{x}, f(\mathbf{x})$
end procedure

nating from the statistical modelling in physics, allows an update for the worse solution at some rate governed by a *virtual annealing temperature*, which is termed as T. The temperature usually starts from a large value and monotonically decreases as the update goes on. The higher the temperature is, the more likely an update for the worse solution is allowed. This is an analogy of a development of physical system under the natural/artificial environmental noise and fluctuations. Another important difference from the greedy search is that SA does not check all the neighbors, but only one selected at random. A pseudo-code for the simulated annealing is shown in Algorithm 1.

Another method for escaping from the local minima is a memory-based updating method called tabu search (TS) (13). In the method, we evaluate all the neighbors, the 1-bit flipped states from the current solution, except ones which flip the recently-changed bits. The prohibited moves are memorized in *tabu list*. Thanks to the rule, the newly flipped bit would be fixed at least for the next certain amount of steps. The point is that we always force an update to the best neighbor, even if it degrades the objective value. The strategy is effective for solving problems like TSPs, but hardly applicable for expensive objective functions due to the evaluations of all the neighbors. The pseudo-code of tabu search is shown in Algorithm 2.

Evolutionary algorithms, including evolutionary strategy and genetic algorithm, are also common as BBO methods (14, 15, 16, 17, 18, 19). They search for good solutions not by enumerating the neighbors, but by bleeding a population of solutions. The total number of function's evaluations is only governed by the predefined population size and the number of generations. That means that they can be applied for problems regardless of their dimensions. Since a population can contain points scattered across multiple local optimums, the methods also show the robust performance against multi-modal objective functions. Their simple concept and implementations also made them appealing as instant solvers for wide range of optimization tasks. Here shows the typical implementation of a genetic algorithm in Algorithm 3.

1.1.2 Surrogate-based methods

Surrogate-based methods are the most successful algorithms for BBOs with expensive objective functions (1). We construct a regression model (or *surrogate model*) in a surrogate-based method that tries to regenerate objective values already evaluated. The model is then used for selecting the next point to evaluate. If the BBO is defined

Algorithm 3 Genetic Algorithm (GA)

Require: $\lambda > 0$ ▷ population size
Require: $N_{\text{max\_gen}} > 0$ ▷ number of reproductions
procedure GeneticAlgorithm(λ)
  initialize population $P \subset \{0,1\}^d$ with size $|P| = \lambda$
  $Y_P \leftarrow \{f(\mathbf{x}) | \mathbf{x} \in P\}$ ▷ evaluate all
  for $n_{\text{cycle}} \leftarrow 1, N_{\text{max\_gen}}$ **do**
    $P' \leftarrow$ **RouletteWheelSelection**(P, Y_P) ▷ parent pairs $P' \subset P \otimes P$
    $P'' \leftarrow \{$**Crossover**$(a,b) | (a,b) \in P'\}$
    $P \leftarrow \{$**Mutation**$(\mathbf{x}) | \mathbf{x} \in P''\}$
    $Y_P \leftarrow \{f(\mathbf{x}) | \mathbf{x} \in P\}$
  end for
  $\mathbf{x} \leftarrow \arg\min_{\mathbf{x} \in P} f(\mathbf{x})$
  return $\mathbf{x}, f(\mathbf{x})$
end procedure

on continuous variables, we can use differentiable regression models and optimize them by Newton's method, quasi-Newton method, and other gradient-based optimization methods. For discrete/binary variables, we use the afore-mentioned methods like SA, TS, and GA, to optimize the models. The optimum in the model is expected to perform well on the objective function, too. We evaluate the objective function on the expected minimum and add the result to the dataset to construct a more precise regression model. Then we select the next point to evaluate and run the objective function repeatedly until some convergence criteria are met or one's budget is used up. The loop of regression, selection, and evaluation is sketched in Figure 1.2.

The use of the surrogate model greatly reduces the number of evaluations of the objective function. However, it is only the case when the models are accurate and the optimizations are correct. The difficulty to predict the shape of the objective function based on a limited number of data points is called the *statistical barrier*, and the difficulty for selecting the best promising input from the large search space is called the *computational barrier*. The proposed FMQA method is a surrogate-based method, too, and addresses the computational barrier problem by using quantum annealing.

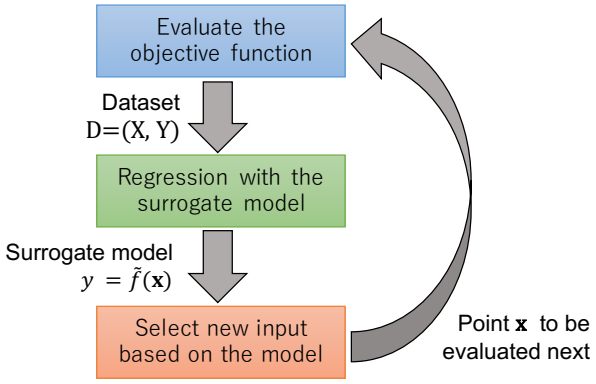


Figure 1.2: Schematic picture of surrogate-based method.

1.1.3 Bayesian optimization

Bayesian optimization (BO), one of the most popular surrogate-based method, should be mentioned here. In BO, a *Gaussian process* (GP) is used as a surrogate model. The GP is not "a" regression model, but a statistical distribution of them calculated as a *posterior distribution* given the dataset. The GP can manage model uncertainties, enabling us to balance the *exploration* and *exploitation* during the search process. Exploration means searching around the highly-uncertain regions where the evaluations are not much made before. Exploitation is searching for solutions extensively around the known good solutions. The former is important for finding the global optima, and the latter for finding the closest local optima quickly. Balancing the dilemma is essential for obtaining good solutions in practical cases. We can change how much weight to put on each of them by modifying a kernel function and the expected noise scale, along with other hyperparameters. However, the drawback of using GP is that it is easily suffered from the *curse of dimensionality*. As the *acquisition function*, which is calculated by the GP model and used for selecting the next evaluation point, can be multi-modal and flat on the most part of the search space, finding the most promising point gets difficult even with quasi-Newton method (20). Some works applying the GP to high-dimensional BBO by dimensionality reduction or designing proper kernel functions exist (21, 22).

1.2 Quantum annealing

The emerging quantum computing researches are sparkled by the recent physical implementations of large scale quantum bits (*qubits*) and their entanglements. Potentially, quantum computers can solve specific computational tasks like database search and prime number factorization faster than the classical computers of the von-Neumann architecture. The supremacy of quantum computing to the classical computers is actually observed on carefully crafted problems (23). However, the current implementations of qubits' interactions are susceptible to environmental noise, and we have to expect so-called noisy intermediate-scale quantum (NISQ) devices in the comming age (24).

In this thesis, we only consider the special type of quantum computing called quantum annealing (QA) (25, 26). QA encodes a problem of interest into the system's Hamiltonian and its ground state, corresponding to the desired solution, is carefully tracked. It should be noted that whether the QA can truly achieve the speedups over the best classical algorithm is still controversial topic (27). However, the current QA machines, which are the executive hardwares to perform QA, are ready to use and already equipped with a few thousands of qubits. Therefore, it is reasonable to assess the scaling performance of quantum devices empirically.

1.2.1 QA capabilities

QA is a kind of heuristic similar to the simulated annealing algorithm, but its variables go through superposed states. The implementations on the current QA machines can solve combinatorial optimization problems represented in the quadratic unconstrained binary optimization (QUBO) format, which is classified as an NP-hard problem (28). QUBO is a minimization problem of the form:

$$\min_{\mathbf{x}\in\{0,1\}^d} \sum_{i\leq j} q_{ij} x_i x_j, \tag{1.2}$$

where d is the number of variables and $q_{ij} \in \mathbb{R}$ is a parameter for configuring problems. Typically, we convert our optimization problems in this QUBO format and solve them with the QA machine.

D-Wave Systems Inc. has developed D-Wave 2000Q$^{\text{TM}}$, a QA machine, to solve QUBO problems with up to around 2000 bits. The size is by far the more than ones

Figure 1.3: Quantum annealing machine D-Wave 2000Q^{TM}(dwavesys.com).

implemented in general-purpose quantum computers so far (29). Diverse applications of QA have been reported in, e.g., biologicl science (30, 31), machine learning (32, 33, 34), and IoT (35, 36).

The device's ability to deal with high-dimensional problems is desirable for BBO to overcome the difficulty of searching in a large space (computational barrier). To use the QA machines for the surrogate-based optimization methods, we adopt a surrogate model which can easily converted to the QUBO format in chapter 2.

1.2.2 Embedding

Despite the large number of qubits implemented, the possible interactions between the qubits are limited on the real devices. For example, the topology of the qubits' connection in the D-Wave 2000Q^{TM} is called *chimera graph* (Figure 1.4 left), where the bipartite graphs composed of 8 qubits (represented by $K_{4,4}$) are arranged in horizontal and vertical directions. The figure means that the coefficient q_{ij} in equation 1.2 can take non-zero value only if an edge connecting i'th and j'th node exists. To deal with the arbitrary QUBO problem with the chimera graph, we have to represent one variable with multiple qubits, instead. This projection is called *embedding* (Figure 1.4 right).

By the embedding process, one variable is represented by several qubits called *chain*. The connections inside a chain are kept strongly ferromagnetic, so that all the members of the chain results in the same state. The accuracy of the QA strongly affected by the length of the chains and strength of their connections. Chains are preferred to be short, and their connections need to be set to minimum required strengths to

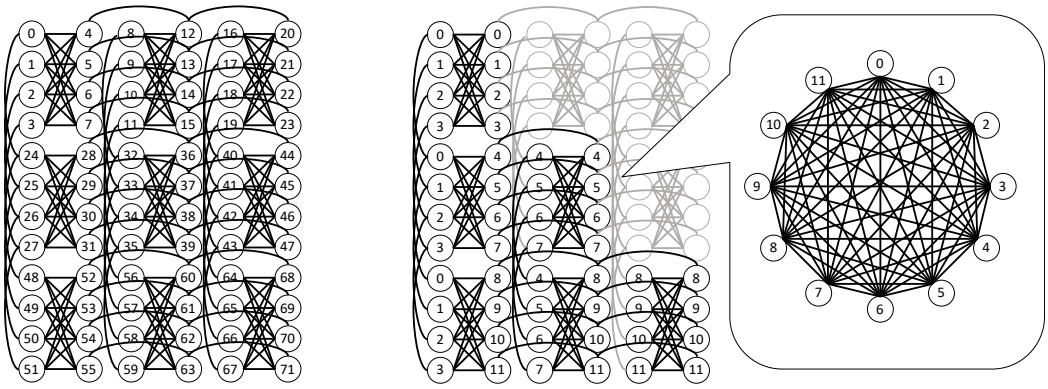


Figure 1.4: (Left) Connections of the qubits in the D-Wave 2000Q^{TM} form chimera graph. (Right) One variable should be embedded into several qubits to deal with arbitrary QUBO.

keep their consistency. Therefore, solving large QUBO problems can be difficult for QA machines. There is a heuristic algorithm to find efficient embedding patterns to amend the difficulty (37). However we do not use the heuristic here, and only use the embedding pattern as shown in Figure 1.4 right because the length of the chains are kept the same and less inhomogeneities between variables are induced (38).

1.3 Conclusion

We reviewed the BBO and several algorithms for solving it in this chapter. Surrogate-based methods are good at grabbing the underlying structures of objective functions. A proper choice of surrogate model and selection schema is important for their performance. Along with the continuous variables considered in BBOs, discrete or binary variables are also gathering more attention. We will challenge these types of problems in this thesis.

Quantum annealing, a special type of quantum computing, has just got ready for business applications. The QA's performance for selecting the best solution from the QUBO problem's solution space is appealing for the selection step of the surrogate-based methods. As an appropriate surrogate model, a statistical model called factorization machine is introduced in the next chapter. We combine those two techniques into a new BBO method and show its practical application.

2

FMQA

2.1 Introduction

A new surrogate-based method, factorization machine for quantum annealing (FMQA), is introduced in this chapter. It is intended for dealing with the large number of candidates during an optimization process. The method is made of two components:

- Factorization Machine (FM) as a surrogate model
- Quantum Annealing (QA) as a selection algorithm

The basics of these components are described here after, and followed by the combining the two. Unfortunately, there are no mathematical guarantees for the convergence, so we would see the empirical performance on real problems to evaluate the method in the next chapter.

2.2 Methods

2.2.1 Regression by FM

Looking at the QUBO defined in the equation 1.2, it is straight-forward to use its objective with a bias term w_0 as a surrogate model:

$$w_0 + \sum_{i \leq j} q_{ij} x_i x_j, \tag{2.1}$$

which is to be optimized by the QA machines. However, we use a factorization machine (FM), a statistical regression model, instead (39). The model has a form:

$$\hat{f}(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=i+1}^{d} \sum_{k=1}^{K} v_{ik} v_{jk} x_i x_j, \tag{2.2}$$

where w_0 is the bias term, $w_i \in \mathbb{R}$ $(i = 1, \cdots, d)$ is the linear coefficient, and the quadratic coefficient for the product $x_i x_j$ is calculated by $\sum_{k=1}^{K} v_{ik} v_{jk} = \mathbf{v}_i \cdot \mathbf{v}_j$. The $\mathbf{v}_i \in \mathbb{R}^K$ $(i = 1, \ldots, d)$ is a K-dimensional vector, where the $K \in \mathbb{N}$ is a hyperparameter called *factorization size*. The set of the model parameters $\{w_i \in \mathbb{R}\ (i = 0, \cdots, d), \mathbf{v}_i \in \mathbb{R}^K\ (i = 1, \cdots, d)\}$ has a dimension of $(K + 1)d + 1$ in total. The important fact is that the number of fitting parameters is reduced from $\frac{d(d+1)}{2} + 1$ in equation 2.1 to $(K + 1)d + 1$ in equation 2.2. The former is quadratically increasing and the latter is linearly increasing to the dimension d. In statistical modeling, models with the large number of parameters are prone to *overfitting*, and require many data points to train generalizable models. Therefore, the models with less parameters are more preferable for the BBO problems with high-dimensional and expensive objective functions. That justifies the use of the FM, rather than the QUBO objective, for the surrogate model.

On the other hand, too a simple model can fail to fit to the dataset as known as the *underfitting* problem. When using the FM as the surrogate-model, the hyperparameter K should be enough large to make the model enough complex. It is known that the factorization size $K = 8$ works empirically well for many cases (39, 40, 41). We mainly use the same K value for our experiments.

In the training step, the model parameters $\{w_0, w_i, \mathbf{v}_i\ (i = 1, \cdots, d)\}$ are tuned so that we obtain the minimum value of a loss function L, which is accounting for the predictive errors on the dataset S:

$$L(w_0, w_i, \mathbf{v}_i\ (i = 1, \cdots, d)) = \sum_{(\mathbf{x}, y) \in S} (y - \hat{f}(\mathbf{x}))^2. \tag{2.3}$$

The minimization of this function is done by the Adam (adaptive moment estimation) algorithm (42), in this thesis.

To select the best configuration in terms of the constructed regression model $\hat{f}(\mathbf{x})$ (equation 2.2) with QA machines, we can convert the regression model to the QUBO

objective (equation 2.1) by

$$w_0 = w_0, \tag{2.4}$$

$$q_{ij} = \begin{cases} w_i & (i = j) \\ \sum_{k=1}^{K} v_{ik} v_{jk} & (i < j) \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

2.2.2 Selection by the D-Wave quantum annealer

A D-Wave 2000Q$^{\text{TM}}$ quantum annealer (29) is utilized to select the next candidate. The working D-Wave 2000Q$^{\text{TM}}$ in Vancouver, Canada has 2038 qubits on the chimera graph, and it can create a virtual fully-connected graph with 63 nodes by regarding some qubits as one variable. The D-Wave Systems Inc.'s implementation of find_clique_embedding method (38) is used for this *embedding* process.

If `num_reads`, which is a parameter for the D-Wave 2000Q$^{\text{TM}}$, is set to 50, the candidate solutions of the QUBO problem is *sampled* for 50 times within 16 ms of *QPU time*. Then the state with the lowest energy out of these 50 states is used as the prediction of the optimal solution. When the suggested solution is already evaluated and contained in the dataset, a randomly generated solution is used for that step instead.

2.2.3 Combining FM and QA

After the optimization by the QA, its actual value is evaluated by the objective function. The result is appended to the training dataset and more accurate FM model is trained. Then, the model is passed to the QA to make a next guess for the optimum. By repeating this procedure, the configuration with a good objective value would be obtained with a small number of evaluations. For machine learning, Ising machines, including QA machines, are experimentally utilized in the training processes of statistical models so far (32, 33). Our use of QA for selecting next solutions in BBO is a new application in this field. The entire algorithm of FMQA is listed in Algorithm 4.

2.3 Conclusion

A new BBO method called FMQA, combining the factorization machine (FM) and quantum annealing (QA), is defined. Overcoming the computational barrier by QA

Algorithm 4 Factorization machine for quantum annealing (FMQA)

Require: $N_{\text{init}} > 0$ ▷ initial data points
Require: $N_{\text{max}} > N_{\text{init}}$ ▷ total evaluations
procedure FMQA($N_{\text{init}}, N_{\text{max}}$)
  select the initial data points $X \subset \{0,1\}^d$ of size $|X| = N_{\text{init}}$ at random
  $S \leftarrow \{(\mathbf{x}, f(\mathbf{x}))|\mathbf{x} \in X\}$ ▷ initial dataset
  train an FM model $\hat{f}$ on S
  for $n_{\text{eval}} \leftarrow N_{\text{init}} + 1, N_{\text{max}}$ **do**
    convert the model $\hat{f}$ into a QUBO parameter $\{q_{ij}\}$
    sample x_{new} from the QUBO $\{q_{ij}\}$ by QA ▷ sampling
    if x_{new} is already in S **then**
      apply random bit flipping on x_{new} until it is not in S
    end if
    $y_{\text{new}} \leftarrow f(\mathbf{x}_{\text{new}})$ ▷ evaluation
    append $(\mathbf{x}_{\text{new}}, y_{\text{new}})$ to S
    train the model parameter of $\hat{f}$ on S again ▷ training
  end for
  return the best data point in S
end procedure

is first in mind and FM is used for constructing effective regression models. There is no mathematical foundation for the effectivity of the method. We will test it on an practical example problem in the next chapter.

3

Application for designing metamaterials

The new BBO method, factorization machine for quantum annealing (FMQA) is introduced in the previous chapter. However, it is difficult to give mathematical guarantees for its performance. Therefore, we show a proof-of-principle application to demonstrate the practical effectivity of the method in this chapter. The objective is a simulated performance of a radiative cooling material which will be discussed in detail later. The result might lead to further applications of the method for designing other materials.

3.1 Introduction

Further evolution in materials that control energy carrier, such as photon, electrons, and phonons, is a condition to realize sustainable industry and society. The key is to manipulate properties at the scale of characteristic length of transport of carriers. Over the last decades, the advance in top-down fabrication and bottom-up synthesis, together with atomistic and spectroscopic characterizations, have given us access to nearly-free exploration of the structures with better energy transport characteristics, and this has resulted in a number of breakthrough in energy materials in photovoltaics (43), thermal radiators (44, 45), batteries (46, 47) thermoelectrics (48), and others. Here, metamaterials are a representative case, where the artificial structures produced inside a material give rise to extraordinary properties.

Now, the above success opens up a new problem that there are too many degrees

of freedom in the structures to explore. Exploration among single crystals is already exhausting when considering compounds but the number of candidates becomes truly massive when extending it to composites in a broad sense with nanoscale inhomogenity in composition or simply "nanostructures". Yet, some materials with such composition inhomogenity have been found to exhibit superior properties over the counterpart ordered structures, particularly in case when structures are smaller than the coherence lengths of energy carriers as shown for photons (49, 50), phonons (51, 52), electrons (53, 54), and magnons (55). Therefore, for further evolution of materials in energy technology, overcoming the challenge of massive candidates via automated materials discovery is crucial.

Automated materials discovery based on BBO is an iterative process of selecting one candidate from massive candidates (i.e., design space) and recommending it for the next experimentation (56, 57, 58). From existing materials properties data, machine learning predicts properties of unobserved candidates and defines an acquisition function in the design space. A global optimization problem in the design space is solved with respect to the acquisition function, and the candidate with the highest one is selected as the next candidate material for experiment. Using the observed properties for the selected candidate, the machine learning model is updated and defines a different acquisition function for the next iteration. By repeating this procedure, we would get materials with desired properties in a small number of experiments. However, the difficulty of BBO is as stated in chapter 1. When a fast simulator, which can calculate properties of target materials, is used to replace the experiment, the statistical barrier is often not as overwhelming as the computational one, as exemplified in several recent studies (56, 58). To circumvent the difficulty of global optimization, heuristic methods such as local search, tree search (59) and genetic algorithms (60, 61) are often used. Instead, to overcome this computational barrier with help of quantum effect, we use a quantum-classical hybrid algorithm, FMQA defined in the previous chapter, employing a D-Wave quantum annealing machine.

3.2 Methods and Target metamaterials

The schematic picture of our algorithm is shown in Figure 3.1. Our algorithm, FMQA, is capable of solving a BBO problem over binary variables representing material's struc-

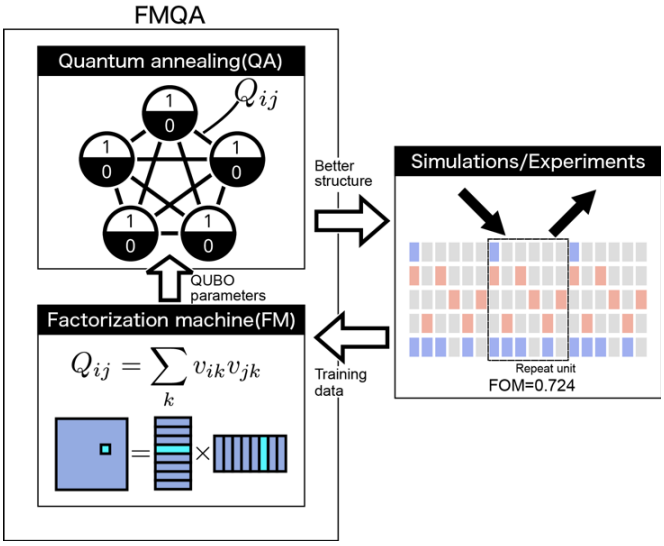


Figure 3.1: Procedure of our automated materials discovery using a factorization machine (FM) for learning and a quantum annealing (QA) for selection. Target property is the figure-of-merit (FOM) for the radiative sky cooling, which is evaluated by the rigorous coupled wave analysis (RCWA). This simulation part can be replaced by other simulation methods or experiments, depending on the target properties.

ture. The property of the new candidate is obtained by an atomistic simulation. In the next step, the new data point is added to the training data, and the FM is retrained. The QA suggests the next structure to be tried based on the FM. These steps are repeated again and again.

3.2.1 Target metamaterials

As we have stated in this chapter, we apply our method to design a metamaterial with tailored thermal radiation spectrum as a proof-of-principle study. Tailoring thermal radiation is fundamentally important as every material emits and absorbs thermal radiation. In engineering viewpoint, wavelength-selective thermal radiation (photons) is needed in a broad range of applications. For instance, narrowband thermal-emission leads to high-efficiency thermophotovoltaics(62, 63), incandescent light source (64), biosensing (65, 66), microbolometers (67, 68), imaging (69), and drying furnace (70).

Another application that has recently attracted much attention, in response to concerns of global warming and energy crises, is radiative sky cooling that utilizes the untapped 3 K cold space as a heat sink. Previous designs on radiative cooling (45, 71, 72,

73, 74, 75, 76, 77, 78, 78, 79, 80) have focused on simultaneously blocking solar energy (0.4-4 μm) while maximizing thermal radiation loss ($> 4\mu$m) to the surroundings. These designs have been experimentally demonstrated to be successful in dry and clear weather. However, this design strategy is not as effective in hot and humid areas because the atmospheric window (8-13μm), which allows thermal radiation to directly transmit to the outer space, becomes less transparent and a large part of the downward radiation that is beyond the window is absorbed by the radiator. Therefore, a radiator that only emits thermal radiation to outer space through the transparency window is desirable. Such a radiator can help maximize the outgoing radiative cooling power while minimizing the ambient radiative energy absorption. In this study, our proposed optimization method is used to design such a radiator with a wavelength selectivity higher than previously designed ones.

A metamaterial, which only emits or absorbs the thermal radiation within the transparency window of the atmosphere (8-13μm), is preferable for radiative sky cooling (75, 81). Various material structures have been proposed to match the spectrum of the atmospheric window, including planar multilayer structures (45, 74), patterned metasurface structures (73, 76, 77, 79), and polymers doped with nanoparticles (71, 78, 80). Most of these structures do not have a high emittance over the whole span of the atmospheric window. Additionally, their cooling capability is insufficient. On the other hand, our design employs SiO_2 and SiC to achieve this stringent spectral selectivity. The dielectric functions of SiO_2 and SiC indicate that they have phonon-polariton resonances positioned at 9.7 and 12.5 μm, respectively. Moreover, they both have very small extinction coefficients in the solar energy wavelength band, implying that the absorption of solar energy should be suppressed.

Inspired by previous research (72), the target metamaterial structure is comprised of SiO_2 and SiC wires placed in poly(methyl methacrylate) (PMMA), which has a negligible absorption in the visible to far-infrared range. Each wire is arranged along the y axis, and light is incident from the top layer (Figure 3.2). The periodic boundary condition is applied in the x directions (i.e., the structure repeats along the x axis). Because the RCWA calculation solves a two-dimensional ($x - z$) problem, the wire is assumed to be infinitely long without any variations in the cross section in the y direction. For structural optimization, the $x - z$ plane is uniformly discretized into

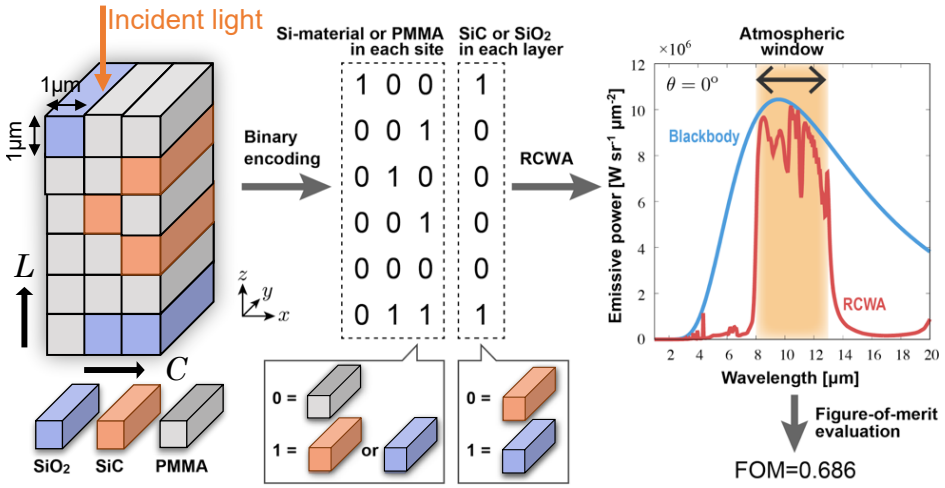


Figure 3.2: Example of the target metamaterial structure for $L = 6$ and $C = 3$, the binary variables expressing it, and the emissive powers of target metamaterial (RCWA) compared with the theoretical upper bound (Blackbody).

square units with 1 μm side lengths, which are either SiO_2, SiC, or PMMA. The numbers of meshes along the z and x directions are defined as L and C, respectively.

One constraint adopted in this study is that a layer can contain only SiO_2 or SiC exclusively. The constraint and the unit size, which correspond to the minimum wire size, are determined considering future fabrication via photolithography. Pretrial simulations confirm that a unit size of 1 μm provides sufficient resolution for the optimization problem. For these structures, the emissivity properties are calculated based on RCWA, and the FOM for radiative cooling is evaluated. In our optimization, the target is the emittance property under p-polarized incidence (TM wave) for a polar angle $\theta = 0°$.

The metamaterial structure is configured by binary variables as shown in Figure 3.2. Initially, the configuration of the wired materials (SiO_2 or SiC) and PMMA is determined by $L \times C$ bits. Then additional bits, which express the type of wired materials in each layer, are prepared. Consequently, the structure of a metamaterial is well-defined using $L \times (C + 1)$ bits. Note that our algorithm can perform a structural optimization when L and C are fixed.

3.2.2 Simulation by rigorous coupled-wave analysis

To calculate the thermal emissivity properties of the target metamaterials, *rigorous coupled-wave analysis* (RCWA) is employed (Figure 3.2 right). RCWA is a semi-analytical method to solve Maxwell's equation and provides a high numerical accuracy (82). The spatial distribution of the dielectric constant and the involved electromagnetic field are decomposed in the x and z directions. By imposing a periodic boundary condition in the x direction and a continuous boundary condition in the z direction, the governing Maxwell's equation can be solved quickly and accurately.

Although PMMA is not universal in identical materials due to the influence of the fabrication process, our calculations assume that the PMMA is pure and the refractive index is fixed as 1.48 for simplicity (83). The dielectric functions of SiO_2 and SiC are obtained from the tabulated data from Palik (84) with interpolation. Comparing the experimental results validates (85) that the RCWA method is a credible approach to design metamaterials for thermal radiators. Figure 3.2 shows a calculated example of the emissive power.

For a good thermal radiator for radiative cooling, the emittance spectra should fall within the wavelength region between 8 and 13 μm. To evaluate the likelihood that the designed metamaterial is the ideal case, the FOM is defined as in (44):

$$\text{FOM} = \frac{\int_{\lambda_i}^{\lambda_f} \epsilon_\lambda E_{b\lambda} d\lambda}{\int_{\lambda_i}^{\lambda_f} E_{b\lambda} d\lambda} - \frac{\int_{\lambda_{\min}}^{\lambda_i} \epsilon_\lambda E_{b\lambda} d\lambda}{\int_{\lambda_{\min}}^{\lambda_i} E_{b\lambda} d\lambda} - \frac{\int_{\lambda_f}^{\lambda_{\max}} \epsilon_\lambda E_{b\lambda} d\lambda}{\int_{\lambda_f}^{\lambda_{\max}} E_{b\lambda} d\lambda} \tag{3.1}$$

where $\lambda_i = 8\mu$m, $\lambda_f = 13\mu$m, $\lambda_{\min} = 1\mu$m, and $\lambda_{\max} = 20\mu$m. Here, ϵ_λ and $E_{b\lambda}$ are the spectral emittance and the spectral blackbody emissive power calculated by RCWA, respectively.

In equation 3.1, $\int_{\lambda_i}^{\lambda_f} E_{b\lambda} d\lambda$ is the energy flux radiated out by the blackbody at the same temperature in the atmospheric-window wavelength range. This is the maximum energy flux that an object at that temperature can emit. Thus, the first term in equation 3.1 is the ratio that quantitatively describes the similarity of the designed structure to the ideal emissivity with respect to the energy flux. For the remaining wavelength range (i.e., $\lambda < 8\mu$m and 13μm $< \lambda$), the optimization strives to make the emissivity as low as possible. Therefore, in the definition of FOM, the coefficients are negative. Since the ideal FOM value is 1 and the worst FOM value is -2, where

the emissivity functions take exactly the opposite shape, our optimization task is to maximize the FOM.

It should be noted that according to Kirchhoff's law, the spectral emittance is equal to the spectral absorptance in the thermal equilibrium state. In the following calculations, the spectral absorptance property is obtained directly from the spectral emittance.

3.3 Metamaterials design by FMQA

The key feature of our method is the use of FM and QA in combination. To check only the validity of using FM as a regression model at first, we compared it against another common regression model. However, comparing the performance of QA against other corresponding methods is beyond our scope. Therefore, the second step of the validation is made on the optimization performance of the entire FMQA algorithm.

3.3.1 Performance of FM

First, to clarify the usefulness of an FM as a regression model for our metamaterial design, the performances using an FM and a Gaussian process (GP), which is mentioned at the chapter 1.1.3, to find the best structure for the $L = 4$ and $C = 3$ case are compared. It takes 16 bits to encode the material's structure in this setting, and an exhaustive search of the acquisition function defined by the negative FOM is even possible by classical computers to select the next candidate material. For the sake of a comparison against the most "thoughtless" method, a search where the next candidate material is randomly generated from all candidates in the selection part is also tried. Since the number of candidate material structures is only $2^{16} = 65536$ in this case, all the FOMs by RCWA are evaluated and the best structure has already been identified.

Figure 3.3(a) shows the best FOM as a function of the number of calculated structures (iterations of the cycle depicted in Figure 3.1) by each method. For each method, 16 optimization runs with different initial choices are performed and the FOMs are averaged out. All three methods use the same first 50 randomly selected structures as the initial data. The regression results are used from step 51 for the FM and the GP cases. The results in order from worst to best are the random search, the GP, and the FM. Because the FM finds the best structure within 300 iterations, it can be stated

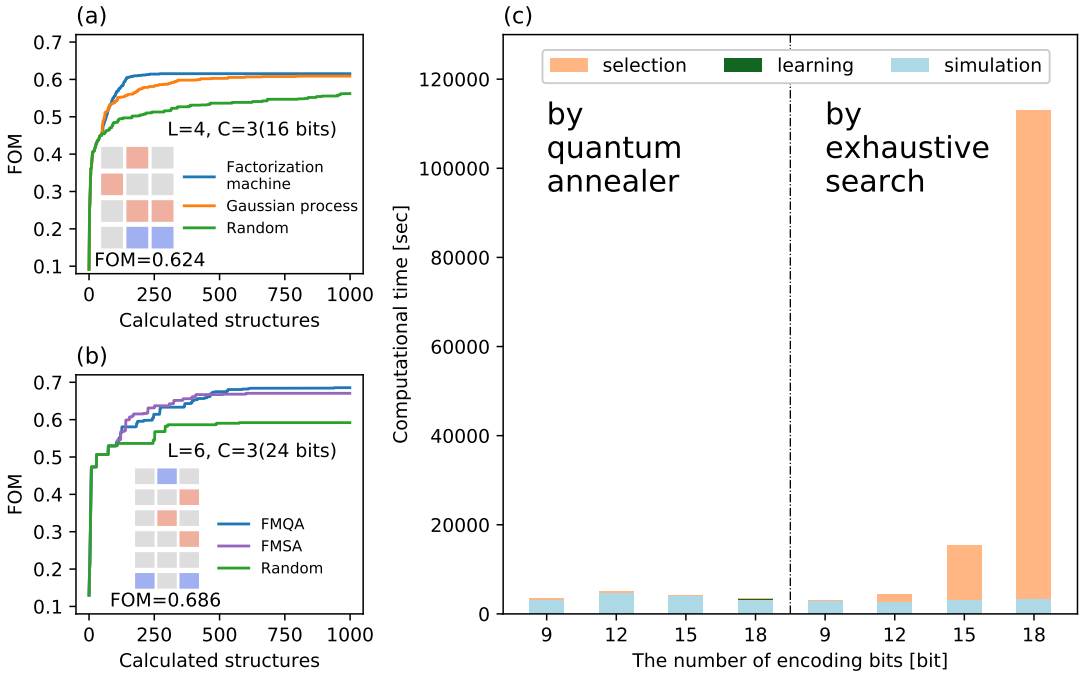


Figure 3.3: (a) Dependence of the best FOM on the number of calculated structures (iterations) by automated materials discovery using an FM, a Gaussian process, and a random search for the $L = 4$ and $C = 3$ case. Inset is the optimum structure and its FOM. Blue, red, and gray squares denote SiO_2, SiC, and PMMA, respectively. (b) Best FOM by FMQA using a quantum annealer, FMSA using a simulated annealing, and the random search for the $L = 6$ and $C = 3$ case. Inset is the structure with the best FOM identified by FMQA. (c) Computing time to perform 500 iterations of automated materials discovery using a quantum annealer and an exhaustive search on a classical computer with an Intel Xeon E5-2690 v3 @ 2.6GHz for the selection part, respectively. Learnings and simulations are performed by the same classical computer.

to be suitable for the regression model of our target. This result says that, at least, the learning by the FM is more useful than a random search to discover metamaterial structures with a high FOM within a small number of simulations. This result is promising for our metamaterial design.

Next, the problem with $L = 6$ and $C = 3$ is considered with the use of QA. Since the candidate number is $2^{24} = 16777216$, an evaluation of all FOMs predicted by regression models such as an FM and a GP to select the next candidate is a time-consuming task. A more realistic option is to use a QA or similar classical algorithms like simulated annealing (SA) for the selection part rather than an exhaustive search. Therefore, we

Table 3.1: Empirical computing time to perform 500 iterations in our automated materials discovery depending on the number of encoding bits.

	by quantum annealer			by exhaustive search		
Bits	Selection	Learning	Simulation	Selection	Learning	Simulation
9	361 [sec]	54.9 [sec]	2920 [sec]	236 [sec]	50.0 [sec]	2830 [sec]
12	385 [sec]	53.0 [sec]	3620 [sec]	1730 [sec]	50.4 [sec]	2630 [sec]
15	423 [sec]	51.1 [sec]	3690 [sec]	12500 [sec]	45.0 [sec]	2920 [sec]
18	374 [sec]	52.2 [sec]	3140 [sec]	110000 [sec]	50.3 [sec]	3170 [sec]

compared the best FOM by FMQA, FMSA (a method using FM as a surrogate model and SA as a selection method) and a random search as functions of the iteration number. The result is shown in Figure 3.3(b). Each line represents the average values from 16 independent runs with 50 different initial structures. The FMQA could reduce the number of simulations to find a better metamaterial structure, with a high probability to reach the higher FOM than the FMSA. Hence, it is a useful tool to design new metamaterials.

Because the computing time to perform FMQA is important, the computing time of automated materials discovery based on an FM when the selection part is conducted by a quantum annealer is compared to an exhaustive search by a classical computer with an Intel Xeon E5-2690 v3 @ 2.6GHz. Note that to this point, the BBO algorithms accompanied by acquisition functions (e.g., Bayesian optimization) have employed an exhaustive search for the selection part. Here, an exhaustive search is compared to the computing time between our algorithm and a conventional algorithm. Figure 3.3(c) plots the computing time to perform 500 iterations as a function of the problem size (number of encoding bits). The selection time, learning time by an FM, and simulation time by RCWA are separately illustrated. The target structure is the $L = 3$ case with various $C = 2, 3, 4$, or 5, which are encoded using 9, 12, 15, and 18 bits, respectively. Their empirical computing times are also summarized in Table 3.1.

Since the learning and simulation are conducted on the same classical computer, these parts require about the same time in both cases. The use of a quantum annealer reduces the selection time. Both methods provide the structure with the highest FOM within only 500 iterations, and the number of simulations can be reduced to find the best one. Note that if the FOMs of all structures are evaluated by the RCWA simulations for

the 18 bits case to identify the best structure, the required time is more than ten times longer than the algorithm where the exhaustive search method is used for the selection part. Although RCWA is a relatively high-speed simulation method (one calculation takes one minute at most), the most time-consuming part becomes the simulation time when FMQA is performed. Consequently, a quantum annealer can be used to solve the hard computational barrier in the automated materials discovery.

3.3.2 Optimum metamaterial structure search by FMQA

We search the optimum structure of the metamaterial for radiative cooling. Varying the number of layers and columns of the target structure should help elucidate a trend to achieve a high FOM. Starting from the previous setting ($L = 6$ and $C = 3$), the number of layers L is changed. The range of L is varied from 3 to 9, and a single run with the first 50 randomly generated initial structures and 2000 iterations is conducted for each optimization trial. Figure 3.4(a) plots the best FOMs as a function of the number of calculated structures for various numbers of layers. The structure with five layers ($L = 5$) exhibits the highest FOM. Figure 3.4(b) shows the found structure with the best FOM for each L. Interestingly, when $L \geq 6$, some layers only contain PMMA but removing the PMMA-only layers decreases the FOM. Consequently, the existence of the PMMA layer plays an important role in improving the FOM for thick metamaterials.

Next, the number of columns C in the target structures is changed while the number of layers is fixed to five ($L = 5$). Figure 3.4(c) shows the best FOM as a function of the iteration number. Larger FOMs appear for the $C = 4$ and 6 cases. Increasing the number of columns to more than seven ($C \geq 7$) does not yield a higher FOM. Since larger cases are multiples of smaller ones with commensurate periods, a metamaterial structure for $C = 8$ should have a similar FOM as the $C = 4$ case. Thus 2000 samplings are too small to find the optimum structure with a higher FOM due to the massive number of candidates for $C = 8$. In fact, the change in the FOMs with the number of calculated structures gradually increases for larger systems [Figure 3.4(c)]. Consequently, searches should be continued to find the best structure for larger systems.

Figure 3.4(b) summarizes the metamaterial structures with a high FOM found by FMQA for various L and C values. The structure for the $L = 5$ and $C = 6$ case

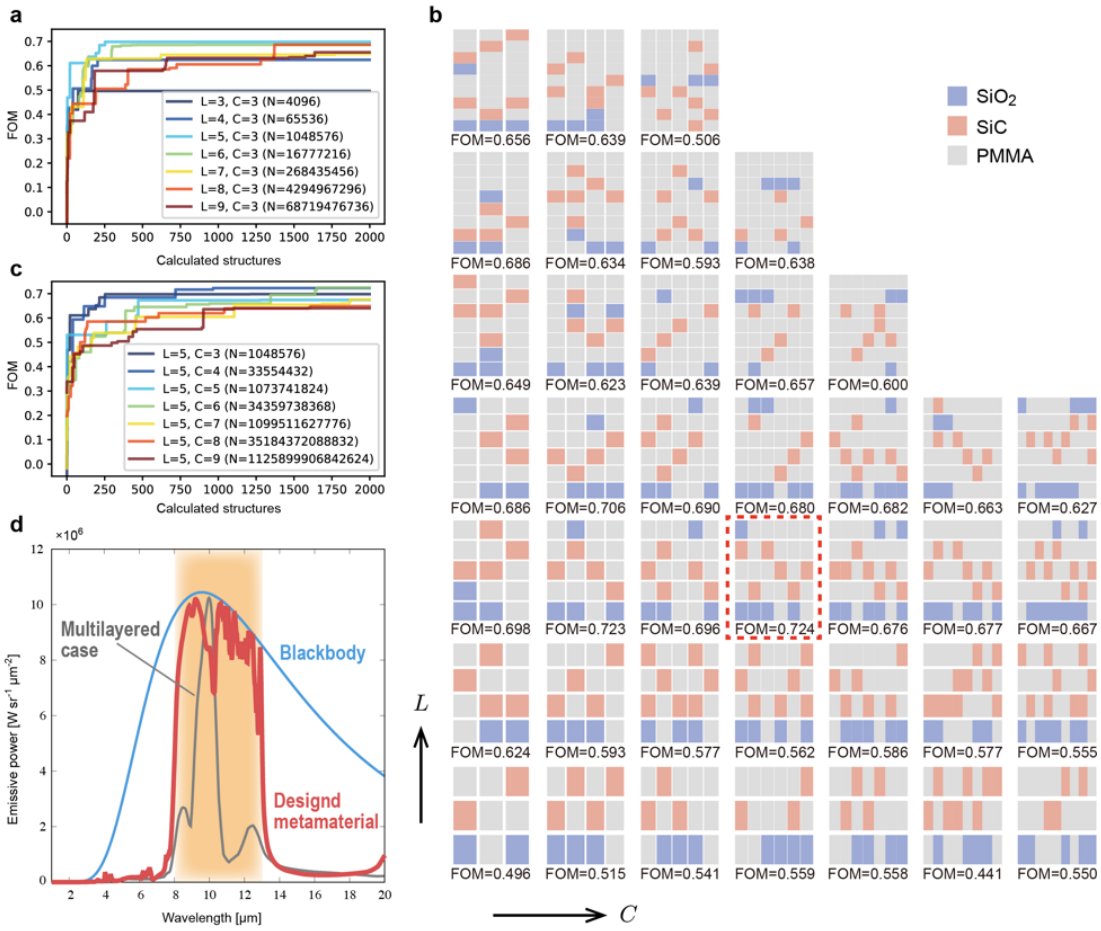


Figure 3.4: (a) Best FOM as a function of the iteration number for $C = 3$ with various L. Plotted results are for a single run with the first 50 randomly generated initial structures. Parentheses denote the number of candidates. (b) Structures with a high FOM designed by FMQA depending on L and C. Red dotted line denotes the structure with the highest FOM in our search. (c) Change in the best FOM for $L = 5$ with various C obtained by a single run with the first 50 randomly generated initial structures. (d) Emissive power calculated by RCWA (red curve) of the designed structure with the highest FOM for the $L = 5$ and $C = 6$ case, which is surrounded by the red dotted line in (b). For comparison, the blackbody emissive power (blue curve) and the multilayer optimum case with five layers (grey curve) are shown (see Supplemental note D in (86)).

has the highest FOM, and its value is 0.724 for radiative cooling. Figure 3.4(d) shows the emissive power of the structure. The large emittances fall into the atmospheric window. Note that the maximum FOM of a multilayer structure constructed by SiO_2, SiC, and PMMA is 0.250 for five layers. In the multilayer structure, each planar layer is constructed by the same material similar to the $C = 1$ case. Hence, the material arrangement in the z direction is optimized. (For the detail optimized structure, see Supplemental note D in (86).) This means that our designed metamaterial structure is essential to obtain a high FOM for radiative cooling.

3.3.3 Mechanism of high emittance in designed metamaterial

From the list of optimized structures (Figure 3.4(b)), structures with SiO_2 located separately at the top and bottom where SiC mediates the middle part always show a higher FOM. To understand which part of the structure absorbs the wave energy, the electric power dissipation density w_e of each part is evaluated. This value is calculated as (87).

$$w_e = \frac{1}{2}\epsilon_0 \epsilon_{\mathrm{Im}} \omega |\mathbf{E}|^2, \tag{3.2}$$

where ϵ_0 is the permittivity in a vacuum, ϵ_{Im} is the imaginary part of the dielectric function, ω is the angular frequency, and $\mathbf{E}$ is the complex electric field calculated by RCWA.

Figure 3.5(a) shows w_e for several typical wavelengths of the optimum structure designed above [i.e., the metamaterial structure surrounded by the red dotted line in Figure 3.4(b)]. The top and bottom SiO_2 absorb most of the wave energy within 8–11μm, whereas the middle parts of the SiC layers dominate the absorption between 11–13μm. Furthermore, around 11.8μm, SiO_2 also facilitates absorption.

Next, the mechanism of the high emittance of the designed metamaterial is examined. The emittance contour plot of the p-wave dispersion relation indicates that a high emittance is almost insensitive against the incident angle (see Supplemental note E in (86)). Hence, the resonance is not due to the surface phonon polariton, which is highly sensitive to the incident angle. On the other hand, in terms of the magnetic polariton (88), the diamagnetic response between the external field and centralized magnetic field inside the structure is usually excited. Consequently, the magnetic polariton is almost insensitive of the incident angle.

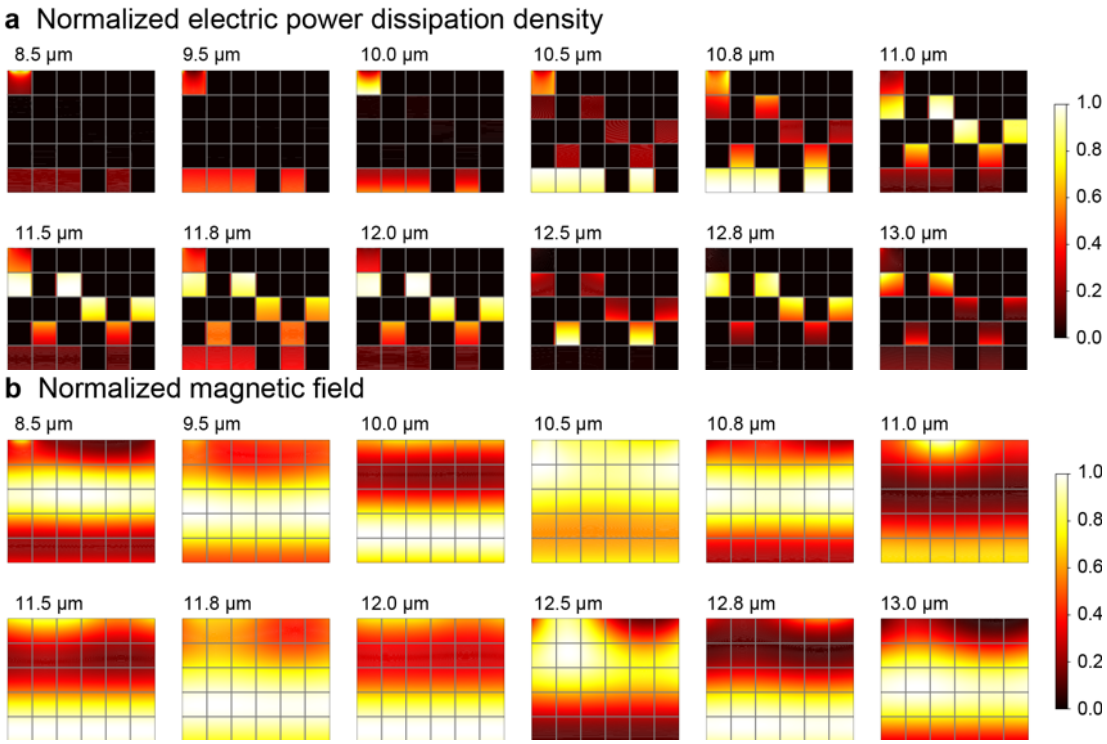


Figure 3.5: Contour plots of (a) the normalized electric power dissipation density at select wavelengths and (b) the normalized magnetic field for the designed optimum structure with $L = 5$ and $C = 6$ which is shown in Figure 3.4 (b).

To further elucidate the magnetic polariton resonance, Figure 3.5(b) plots the magnetic field normalized by the maximum value at several typical wavelengths. Comparing Figure 3.4(d) and 3.5(b) reveals that the magnetic fields always show a strong confinement at the part with a high emittance, which is where the polariton resonance is excited. Furthermore, when the emittance curve falls, the confined magnetic field becomes flatter and less centralized. These results suggest that the high emittance of the designed structure originates from the magnetic polariton resonance. Since a multilayer structure with five layers only shows a narrow absorption band between 8.5 and 10.5μm (Figure 3.4(d)), the structure design of materials is critical to confine the magnetic field. To promote the understanding of our designed thermal radiator, the angle dependence and theoretical cooling power analysis are discussed in Supplemental notes E and F in (86).

3.4 Discussion and summary

In summary, we have proposed a new optimization technique called FMQA, which uses a quantum annealer for automated materials discovery. In our algorithm, the next candidate material, which is selected with respect to the acquisition function, is represented as a solution to a combinatorial optimization by using an FM, and this optimization problem is solved by a quantum annealer. By performing FMQA using the D-Wave 2000Q$^{\text{TM}}$ quantum annealer, we have demonstrated that a metamaterial can be designed for radiative cooling within a small number of RCWA simulations. In the target metamaterial, which has SiO_2 and SiC wires placed in PMMA, a high FOM of 0.724 is achieved for radiative sky cooling. Compared to previously designed structures, the targeted single polarization FOM is far greater than the best reported structures with a comparable polarization- and angle-averaged cooling power (see Supplemental note F in (86)).

Although fabrication is beyond the scope of this thesis, a stratified structure similar to the designed structures has been prepared using current technology (89). Hence, it should be possible to fabricate the designed optimal structure. The results will be discussed elsewhere. Here, we demonstrate that our method is applicable to the case where the metamaterial design is expressed by binary bits. However, if combined with methods for encoding integer variables (90, 91), our algorithm is applicable to discrete

structural and compositional optimizations of any property as long as the property calculation is relatively fast with respect to the optimization process. Therefore our algorithm should help manage the transportation of various carriers (e.g., phonons, electrons, and magnons) and contribute to creation of new energy materials by combining with a straight-forward extension such as tailoring spectral and angular-dependent radiative heat transfer.

Ising machines are conventionally used to optimize explicitly defined functions. Herein quantum annealer as an Ising machine is used for the black-box optimization algorithm. In the future, the application domain of our algorithm will expand to even larger problems as next-generation quantum annealers or other Ising machines equipped with many bits (51, 92, 93, 94, 95, 96, 97) become available. As demonstrated by our experiments, the hard computational barrier (e.g., candidate selection) in automated materials discovery can be partly resolved with the help of an Ising machine. Accordingly, the most time-consuming part will be the simulation of material properties even if a fast simulation on classical computers is used. Therefore it is essential to speed up the materials simulations. As there are increasing number of studies on materials simulations by quantum computing to overcome this problem (66, 98, 99, 100, 101, 102, 103, 104, 105, 106), we expect in the future that our algorithm will further accelerate materials discovery.

4

Local modeling

The FMQA showed a good performance for the metamaterial designing problem. However, a difficulty which is mentioned as the statistical barrier still remains in the construction of a regression model on high-dimensional space with a small dataset. In this chapter, we modify the FMQA so that the FM considers only the data points on a local region of the input space for ensuring the accuracy inside it. The restricted search space is still large, though, and the annealing solvers should be used for selecting the next solution from that region. After the definition of the method, the demonstration for solving black-box optimization on randomly-generated objective function is made.

4.1 Introduction

In surrogate-based optimization methods, the predictive accuracy of the model is essential for the effective search. There are two scenarios where the required accuracy is not satisfied. The first case is when the regression model does not have enough complexity to capture the shape of the objective function. The second is when not enough data points are available for the regression. There are no ways to make accurate predictions in both cases. As long as we stick to use the QA or other Ising machines, we cannot replace the FM for the regression model. To amend the first difficulty, as a consequence, we need to regulate the area of the target region, and inside of which, train a local regression model. This should be easier than modeling the whole input space. Once the search space is localized, it also becomes easier to collect enough data points for the regression, and the second difficulty is also amended. Localized regression models are

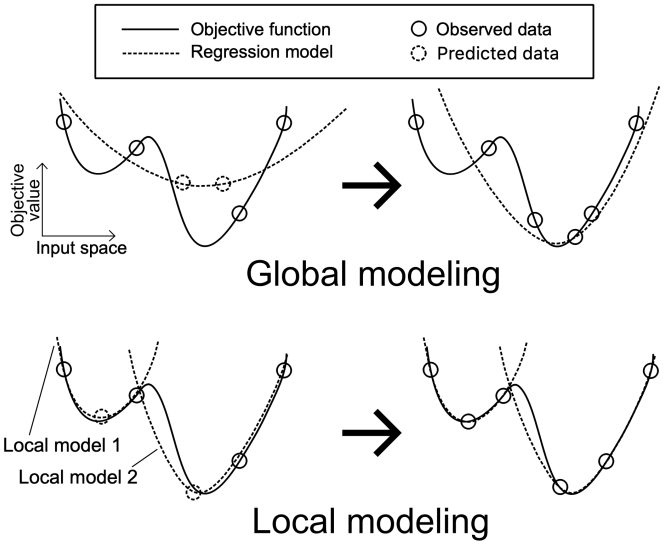


Figure 4.1: Schematic image of BBO with global/local modeling.

actually reported to make more accurate predictions in machine learning tasks (107). Schematic images of fitting objective function by a single global model or multiple local models are shown in Figure 4.1. A model trained with the whole dataset (global modeling) is expected to gradually fit to local optimum region as the data points accumulate around the optimum solution. On the other hand, multiple models meant for fitting local parts of the dataset (local modeling) would converge faster to local optimum solutions of their respective regions. Some of them converge to the global optimum. Since it is difficult to tell which local region is promising beforehand, we have to leave a flexibility so that a model can shift its domain and escape from a local optimum.

Based on the idea of local modeling, we will try to use the FM as a local surrogate model to guide through the search space. We will call this method *local FMQA* against the original FMQA in chapter 2. For the simplicity, no regularization for the model complexities would be introduced here. Note that it can be also important to avoid over-fitting or to keep the interpretability of the model (108).

4.2 Training local models

4.2.1 Loss functions

The local modeling has been considered in the context of evolutionary optimization algorithms (109, 110), to deal with inhomogeneity of the objective function across the search space. According to the methods, a local model in this chapter is trained by a local loss function, which try to fit the model parameter on only a local fraction of the dataset around a queried point $\mathbf{q}$. The squared predictive error of each data point $(\mathbf{x}_i, y_i)$ is multiplied by a certain weight determined by the distance of $\mathbf{x}_i$ from $\mathbf{q}$. The loss function for training the FM model (equation 2.2) is

$$L_{\mathbf{q}}(w_0, w_i, \mathbf{v}_i \ (i = 1 \cdots, d)) = \sum_{i=1}^{N} (\hat{f}(\mathbf{x}_i \oplus \mathbf{q}) - y_i)^2 K\Big(\frac{d(\mathbf{x}_i, \mathbf{q})}{\sigma}\Big), \tag{4.1}$$

where $\oplus$ is the XOR operation (applied bitwise for vectors), which makes 1 if the two bits disagree and 0 if not. $d(\cdot, \cdot)$ is the distance function between two points, and σ is the diameter of the local region which needs to be adjusted for each problem. $K(\cdot)$ is a *kernel function*, which convert the magnitude of the distance to a weight with which the error is accounted for. A most simple distance metric for the binary bits is *Hamming distance*:

$$d(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{d} a_i \oplus b_i. \tag{4.2}$$

However, we would use a different distance metric:

$$d_\mu(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{d} \{1 + 4p_i(\mu)(p_i(\mu) - 1)\}(a_i \oplus b_i), \tag{4.3}$$

where μ is a inverse temperature parameter (≥ 0) and $p_i(\mu)$ is an estimated probability that the i'th bit of the optimum solution is 1. The definition of $p_i(\mu)$ is given later. The intuition behind equation 4.3 is that the flipping of the almost-confirmed ($p_i(\mu) \approx 0$ or 1) bits matters more than unconfirmed ($p_i(\mu) \approx 0.5$) ones. The kernel adopted in equation 4.1 is:

$$K(\zeta) = \begin{cases} (1 - \zeta^2)^2 & \text{if } \ \zeta < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{4.4}$$

This kernel puts more emphasis on the closer data points and ignores ones outside of the $1 \leq \zeta$ region (hence, outside of the σ-width region in input space is dismissed).

As a related work, an approximated covariance matrix between variables, C, is used as a metric in an evolutionary search method called *local meta-model CMA-ES* (111):

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(\mathbf{p} - \mathbf{q})^\top C^{-1} (\mathbf{p} - \mathbf{q})}. \tag{4.5}$$

The C is consecutively updated based on each population. The use of a metric defined by the empirical dataset is useful for ill-conditioned objective functions (16).

4.2.2 Balancing exploration and exploitations

The bitwise probability $p_i(\mu)$, which is introduced in *global equilibrium search* (GES) method (112, 113), is an approximation to a Boltzmann distribution governed by the objective value as energy. Its definition is based on the observed dataset:

$$p_j(\mu) = \frac{\exp[-\mu(y^{\min}_{x_j=1} - y^{\min})]}{\exp[-\mu(y^{\min}_{x_j=0} - y^{\min})] + \exp[-\mu(y^{\min}_{x_j=1} - y^{\min})]} \quad (j = 1, \cdots, d), \tag{4.6}$$

where μ is an inverse temperature and $y^{\min}$ is the best objective value found so far in the dataset S. $y^{\min}_{x_i=0}$ and $y^{\min}_{x_i=1}$ are the best found objective values with its j'th bit being 0 or 1:

$$y^{\min} = \min_y \{(\mathrm{x}, y) \in S\}, \tag{4.7}$$

$$y^{\min}_{x_j=0} = \min_y \{(\mathrm{x}, y) \in S \mid x_j = 0\}, \tag{4.8}$$

$$y^{\min}_{x_j=1} = \min_y \{(\mathrm{x}, y) \in S \mid x_j = 1\}. \tag{4.9}$$

When the inverse temperature $\mu = 0$, the probability makes no prediction and all the data points are treated equally in the loss function:

$$p_i(\mu = 0) = \frac{1}{2}, \tag{4.10}$$

$$d(\mathbf{x}, \mathbf{q})|_{\mu=0} = 0, \tag{4.11}$$

$$L_{\mathbf{q}}(w_0, w_i, \mathbf{v}_i \ (i = 1 \cdots, d))|_{\mu=0} = \sum_{i=1}^{N} (\hat{f}(\mathbf{x}_i \oplus \mathbf{q}) - y_i)^2. \tag{4.12}$$

When μ is large, the probability converges to 1 only on the best observed configuration,

and the locality is strictly defined by the Hamming distance:

$$p_i(\mu \to \infty) = \begin{cases} 1 & (y^{\min}_{x_j=1} = y^{\min}) \\ 0 & (y^{\min}_{x_j=0} = y^{\min}) \end{cases}, \tag{4.13}$$

$$d(\mathbf{x}, \mathbf{q})|_{\mu\to\infty} = \sum_{i=1}^{d} x_i \oplus q_i, \tag{4.14}$$

$$L_{\mathbf{q}}(w_0, w_i, \mathbf{v}_i \ (i = 1 \cdots, d))|_{\mu\to\infty} = \sum_{i=1}^{N} (\hat{f}(\mathbf{x}_i \oplus \mathbf{q}) - y_i)^2 K\Big(\frac{\sum_{j=1}^{d} \{\mathbf{x}_i\}_j \oplus q_j}{\sigma}\Big). \tag{4.15}$$

By the use of inverse temperature parameter μ, we can manage the metric, and change to what extent the model is localized. Enforcing no localization will encourage an active *exploration* of the search space, while strong localization will encourage a heavy *exploitation* from around the existing good solutions. In many optimization tasks, managing the balance between the exploitation and exploration is important to find good solutions within a limited number of evaluations. However, in this thesis, the μ is fixed throughout the optimization to keep the experimental settings simple and tractable.

4.3 Sampling from local models

By the sampling from the local model around the point $\mathbf{q}$, a relative coordinate δ^* is returned for the next evaluation of a point $\mathbf{q} \oplus \delta^*$. We need to limit the norm of δ^* so that $\mathbf{q} \oplus \delta^*$ is within the scope of the local model. We therefore minimize the model with a penalty term to suppress the norm of δ^* from exceeding the predefined width σ:

$$\delta^* = \arg\min_{\delta} \big[(1-\lambda)\hat{f}(\delta) + \lambda \cdot \max(0, \sum_{i=1}^{d} p_i(\mu)\delta_i - \sigma)\big]. \tag{4.16}$$

Here, λ is the penalty strength (> 0) which is set manually in this experiment. The penalty term can be encoded into the QUBO format with additional auxiliary bits. We then evaluate the objective function on $\mathbf{q} \oplus \delta^*$ and append the result to the dataset. If $f(\mathbf{q} \oplus \delta^*) < f(\mathbf{q})$, $\mathbf{q}$ is replaced with $\mathbf{q} \oplus \delta^*$. Repeating the modelings, samplings, and evaluations, the $\mathbf{q}$ is expected to become better.

In both the plain FMQA and local FMQA, the model can be trapped by a local optimum and the increasing data points around it makes it further difficult for the algorithms to escape from the local region. In local FMQA, we can change the inverse

temperature μ, and/or reset the center point of the local model by sampling from $p_i(\mu)$. The way of resetting the local region according to the equation 4.6 is also coming from the GES method. Although we won't try these methods here, the tuning of μ and resetting the local region could help the optimization process.

4.3.1 Updating model parameters

When we change the center point of the local model from $\mathbf{q}$ to $\mathbf{q} \oplus \delta^*$, we convert the parameter $\{w_0, w_i, \mathbf{v}_i \ (i = 1, \cdots, d)\}$ of the original model $\hat{f}(\delta)$ to the newer $\{w_0', w_i', \mathbf{v}_i' \ (i = 1, \cdots, d)\}$ of $\hat{f}'(\delta)$ by shifting variable:

$$\forall \delta \quad \hat{f}'(\delta) = \hat{f}(\delta \oplus \delta^*) \tag{4.17}$$

$$\Leftrightarrow \begin{cases} \mathbf{v}_i' & = (1 - 2\delta_i^*)\mathbf{v}_i, \\ w_i' & = (1 - 2\delta_i^*)\big(w_i + \sum_{j \neq i}(\mathbf{v}_i \cdot \mathbf{v}_j)\delta_j^*\big), \\ w_0' & = \sum_{i<j}(\mathbf{v}_i \cdot \mathbf{v}_j)\delta_i^*\delta_j^* + \sum_i w_i\delta_i^* + w_0 = \hat{f}(\delta^*). \end{cases} \tag{4.18}$$

By inheriting the parameters from the previous model, the training cost of the current model over the new local region is reduced.

4.4 Benchmark on CUBO

Our interest is in that the use of local modeling encourages the better regression accuracy for the black-box objective function, in local regions. The FM model is inherently a quadratic function and, for instance, would fail to fit to cubic functions if trained over the entire input space. To test the robustness of the local FMQA in such a case, we generate *cubic unconstrained binary optimization* (CUBO) problems at random and solve them in the BBO setting to compare with other methods such as the plain FMQA and greedy search.

A CUBO problem is formulated as follows:

$$\min_{\mathbf{x} \in \{0,1\}^d} \sum_{i \leq j \leq k} q_{ijk} x_i x_j x_k, \tag{4.19}$$

where the problem is formulated by the coefficients q_{ijk}. According to the similar rules for generating QUBO (not CUBO) problems for benchmarking in (114), several CUBO problems are randomly generated so as that: (i) a problem has non-zero coefficients for 1% of possible $\frac{n(n-1)(n-2)}{3!}$ cubic terms, $\frac{n(n-1)}{2}$ quadratic terms, and n linear terms. (ii)

All non-zero coefficients are sampled from integer-valued uniform distribution between $[-100, 100]$. The problem size d is set to 64, and 5 problems are generated at total.

4.5 Results

We compared the local FMQA method against the plain FMQA discussed in Chapter 2, greedy search algorithm, and random search. In the local FMQA method, we used the inverse temperature $\mu = 1.0$, local region width $\sigma = 8$, and regularization weight for the step size $\lambda = 0.05$. Simulated annealing is used for the sampling, as a general 64 bit problem is still difficult to solve with the D-Wave machine. This does not matter here because our focus is on the use of local FM model, rather than on the use of QA.

The greedy search is the most naive local search method. Starting from a random point, it try to flip one bit at a time. In the experiment, we do not evaluate all the configurations next to the current solution to select which bit to flip. Instead, we confirm a bit flipping as soon as the evaluated score is revealed to be better than current one. This accelerates the optimization process in practice. We can think of this method as the most exploitation-oriented one.

The result of an experiment on a CUBO problem is shown in Figure 4.2. Each curve is a minimization trajectory averaged over 10 times of trials. The colored area shows the standard deviation. At the beginning of the optimization (right after the evaluations of 100 random structures), the improvement of the solution by local FMQA and the greedy search was faster than FMQA. While the greedy search tended to stuck at higher values, the local FMQA reached as good value as the plain FMQA achieved finaly, with fewer evaluations. Though the FMQA method found good configurations after 3000 times of evaluations, it showed a plateau-like curve in the early stage of the optimizations. This was not always the case in the all generated problems, but FMQA tended to be slower than greedy method at the beginning of the optimization. The results on all of the problems are shown in the Appendix A of the thesis.

4.6 Discussion

The local FMQA was as fast as greedy search at the beginning, and found as good solutions as plain FMQA. It seems that we could incorporate the early converging

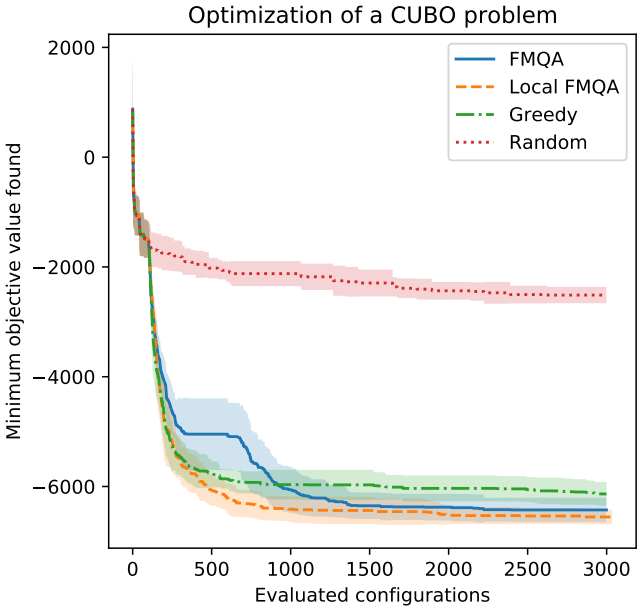


Figure 4.2: Result of optimization for a generated 64bit CUBO problem.

speed of the local search to the FMQA without causing the higher probability of being caught by local minima. We neither reset the center point of local model, nor changed the inverse temperature for this experiment, but the method was still resistant to the local optimums. It can be assumed that the method might work robustly without the delicate tuning of the hyper parameters. We would like to empirically investigate their effects and best configurations in the future research.

5

An application for analyzing quantitative structure-activity relationship

The local FMQA method made the optimization process fast especially in the earlier steps. The improvement would be more beneficial when dealing with higher dimensional optimization problems because learning global FM models in the plain FMQA with only dozens of data points gets more and more difficult as the dimension increases. An important example of that sort of problems is feature selection where we have to choose a subset of available variables for constructing better machine learning models. The optimization of models for predicting drug-like molecules' activity against a specific protein is tried in this chapter.

5.1 Introduction

The structure of organic molecules plays an essential role in chemical reactions in biological systems such as selectively triggering the function of a targeted enzyme or propagating information between neural cells. Many drugs enhances or suppresses the process to modify the chemical dynamics in the body toward the desired one. Computer-aded drug design/discovery (CADD) has been an indispensable tools for pharmaceutical researches to tame the structural diversity and complex interactions of molecules. Although molecular mechanics (MM), molecular dynamics (MD), and

other simulation methods considering even quantum effects have been evolved and are available for replacing the part of initial screening steps, they cost a lot of computational time/resources and it is still unrealistic to examine every single candidate molecule. Therefore, many machine-learning methods to reduce the number of candidate molecules or to suggest a good lead compound to developers have been proposed.

The attempt to construct predictive models of organic compounds' activities/properties is called the quantitative structure-activity relationship (QSAR) analysis(115, 116). Scientists have noticed that the structual information of a group of compounds can describe their some important properties by simple regression models (117, 118). The very basic type of QSAR models is linear regression for some essential variables like electronegativity and hydrophobicity based on the molecule's residues. Some recent QSAR models utilize the neural networks to capture non-linear effects. Their input space can be high-dimensional feature vectors which are generated by fingerprinting methods (discussed later).

One important goals of statistical modeling is to gain insight behind the target system. Expanding the feature vectors to achieve better predictive accuracy is the half of the work. The other half is to investigate the resulting model. Before looking at, say, thousands of input features, we need to extract the only features that are relevant to the objective value or performance.

5.2 Feature selection

Feature selection is the problem of choosing only the important features out of many unneeded or redundant features. All feature selection methods are classified into three groups: *filter method*, *wrapper method*, and *embeedded method*.

5.2.1 Filter method

Reducing the features based on the data itself and then passing the remained features to the training of the predictive model is called *filter method*. The correlation between features and labels, or information gain are often used. This method is relatively handier, faster and can be applied independent of the successive model to be trained.

5.2.2 Wrapper method

A feature reduction method, where the predictive model is repeatedly trained on the subset of features until finding the good subset, is classified as *wrapper*. Since we directly check the model's performance during the process, this method can find the best models while the computational cost is also the highest. This is the same for the black-box optimization, which is the repeating of observing the objective value and changing the input as many times as allowed. This means that the FMQA method is also classified as this wrapper method, when applied for the feature selection problem.

5.2.3 Embedded method

Some machine-learning models automatically dismiss irrelevant features for their predictions. We say that the feature selection is *embedded* in the model for this case. Once the model training is finished, we can safely omit the unused variables from the input.

5.3 Application

In (119), 1746 molecules from a dataset by Tox21/ToxCast programs (120, 121) of U.S. Environmental Protection Agency (U.S. EPA) were extracted for training QSAR models of androgen receptor (AR) activities. The molecules' 1024-bits extended connectivity binary fingerprints (ECFPs) were prepared in (122), which is made available on the UCI online repository (123). The ECFP is a binary-valued feature vector of a molecule, and each bit of which represents whether or not one or more specific substructures are contained in the molecule. The QSAR models based on the molecular fingerprints are getting more popular thanks to the increasing processing power of computers and statistical models which can express nonlinear effects between input variables. However, the drawback of using them is the excess amount of variables which are prone to the overfitting problem. The models are expected to show better generalizability with the dimension reduction by feature selection techniques. Here, the local FMQA is adopted for selecting the effective set out of the 1024 features to construct a classifier for predicting the AR activity.

We need an index to indicate the "goodness" of a subset of variables for the feature selection. The N-Nearest Neighbor (N3) method (124), a variation of nearest-neighbor classifiers, was reported to show a better accuracy along with other methods like naive

Bayes and random forest classifier on the dataset (122). Therefore, we use the cross-validated (CV) F1 score of a similar model, k-Nearest Neighbors (k-NN) models with $k = 3$ and Jaccard similarity, in this chapter. Technically speaking, the optimization is treated as a minimization of the negated F1 score in our black-box optimization experiments.

The performance of local FMQA was compared with several algorithms: random search, greedy search, forward feature selection, genetic algorithm (GA), and FMQA. The forward feature selection is a common method where the considering feature subset in the modeling is initialized with null set and gradually incremented. At each step of the increment, a feature variable which gives the best improvement on the score is joined into the modeling. We here avoid checking all the effect of adding each unused variable exhaustively at each step, and instead, check only subsampled 10 candidate variables from which the best one is joined. This makes it possible to choose 100 variables after 1000 times of score evaluations. In greedy search, we do the same subsampling for candidates' evaluations to update the initial configuration 100 times within the 1000 times of evaluations.

GA is also used for reducing the input variables of QSAR models (125, 126). In GA, a group of variable subsets is generated at first. The group experience alternating selection and reproduction (*cross-over* and *mutation*) phases to make new generations so that the only well-performing subsets survive in the end. In the experiment, we used the group size of 50 and made it pass through 20 generations, which entail 1000 times of score function's evaluations. The 8 parent subsets were selected for one cross-over, mixing in all the configuration, with roulette wheel selection, and the 10 percent of the bits in the resulting configuration are mutated at random. For the experiment, an implementation of GA in Python language, PyGAD (127), is used.

For the local FMQA, we used the diameter of the local region $\sigma = 16$, inverse temperature $\mu = 3.0$, and regularization weight $\lambda = 0.1$. SA solver, instead of QA, is used for sampling due to the large problem size.

5.4 Result

The optimization curves are shown in figure 5.1. The local FMQA method showed the best optimization performance of all the methods. However, the result of the normal

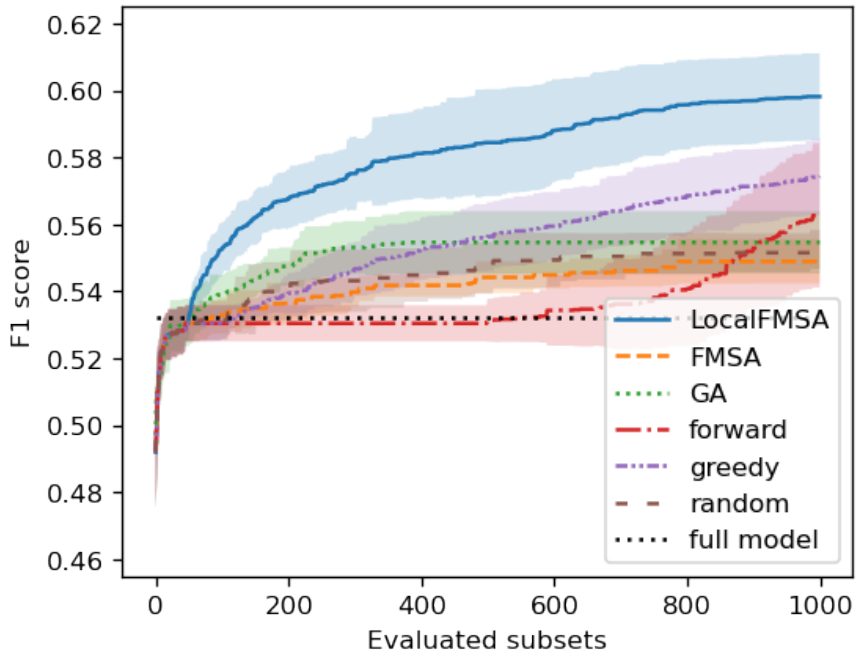


Figure 5.1: Optimization of the feature subset for constructing QSAR model of AR activity.

FMQA tended to be worse than random search.

The forward selection began to improve at around 600 evaluations, which tells that the models with more than 60 variables included would be better. The initial improvement rate in GA was relatively faster, but it soon got stuck. Its resulting configuration was slightly better than random search. The greedy search method showed the steady improvement along the optimization steps.

5.5 Discussion

The better performance of local FMQA than other methods indicated its optimality for the high-dimensional feature selections. The fail of FMQA, on the contraly, suggest the difficulty of training a global FM model over high dimensional space with data points fewer than the problem dimension. The steady improvement in greedy method implies that the method did not reach local optimums within the 1000 evaluations. Finding even a local optimum seems to be an expensive task in high dimensional problems. Therefore, the greedy-like methods would be reasonable options when the construction

of the model is costly. The forward feature selection is also a method which collects features in a greedy manner and its performance has been dramatically risen up after selecting 60 features in this case. It suggest that the better performance might be achieved with the forward selection method if more evaluations are allowed.

5.5.1 Validation of the models

In the experiment, the feature subset with the better CV score was obtained. A good score means that the models trained on the chosen features "generalize well" for test data. However, the model can be still overfitting to the specific patterns of splitting data used in the CVs. With the local FMQA, about 500 features were selected for the modeling (Figure 5.2). It is still too many to understand the reasons why the choice was made, but there are two other ways to validate the selected features. The one way is that if the selected features are preserved across the different trials of optimization, then the features are considered to have some knowledge for the classification. Unfortunately, though, no consistent patterns were found in this experiment. The other way is to re-evaluate the F1 scores of the found feature subsets over different dataset, or different data splitting patterns of CVs. The distribution of the F1 scores of the feature subsets found by local FMQA, greedy, and forward feature selection were measured on different CV trials and shown in Figure 5.3. For the box plots, 50 different random seeds were used for CV calculations. On all the methods, the score with the original random seed showed the best score over different seeds, and there were significant gap between the best and the typical scores. However, the order of the median scores of the methods were the same as their best F1 scores. This indicate that the best feature subset found here has some amount of justification to give better classification models on different datasets on average.

5.6 Conclusion

An application of local FMQA method is demonstrated in this chapter. Note that we used the SA sampler rather than the QA, to deal with the high-dimensionality of the problem. The local FMQA method found the better feature subsets for statistical classification models with fewer actual constructions of the models. This would be a

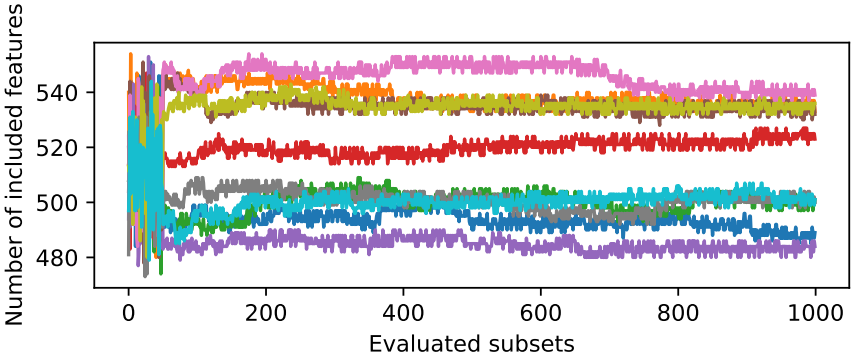


Figure 5.2: Number of features selected out of 1024 for the modeling in the 10 different trials of the local FMQA.

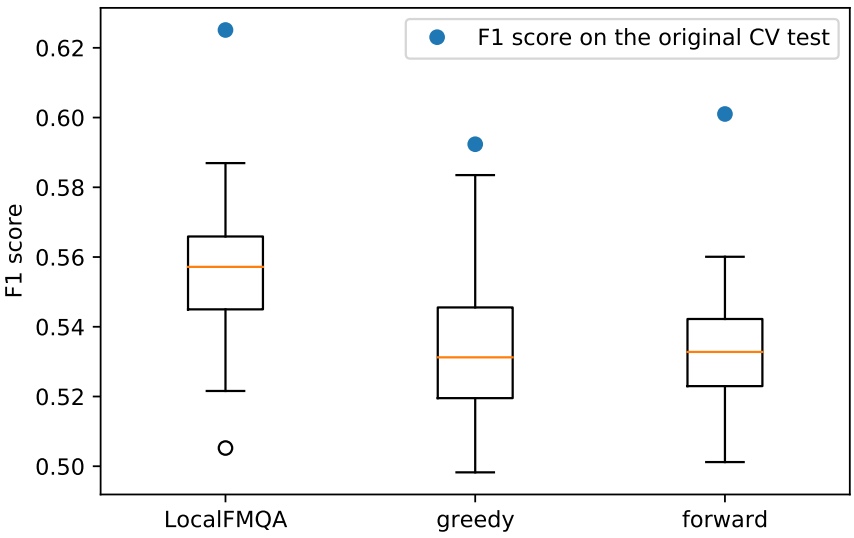


Figure 5.3: F1 scores of best feature subsets found by each method is validated by different random seeds for CV.

help especially when the training of the model is costly, like training neural networks or other complex statistical models.

The resulting feature subset was validated, and it showed a bit of robustness for different cross-validation patterns. It would have contained the important combinations of features for the classification task, but yet to be discovered. To facilitate the understanding of the resulting models, we need to further reduce the number of selected features without loosing the F1 score. Therefore, imposing the penalties for selecting many features in the local FMQA to obtain compact and understandable models will be the future work.

6

Conclusion and perspective

6.1 Conclusion

The basics of black-box optimization is discussed in chapter 1. The versatility of the methods are crucial for real-world optimization tasks. Among the methods, the surrogate-based methods are promising for reducing the number of evaluations of objective functions, but suffers from computational barrier due to a large search space. As a new black-box optimization method for combinatorial optimization problems, the FMQA is proposed in chapter 2. It utilizes a quantum annealing solver to shorten the time for selecting the next configuration to be evaluated, and partly solved the computational barrier. It actually showed a good performance for a materials designing problem in chapter 3. Those results encourages the use of FMQA for many combinatorial optimization problems.

In chapter 4, the factorization machine in the FMQA method is extended with local modeling technique (local FMQA). The experiments showed that the use of local models accelerated the optimization in the beginning of the process. This would be beneficial especially for problems on high dimensional space, where a normal factorization machine requires a lot of data points until it makes predictions with some accuracy. In chapter 5, we demonstrated the performance of local FMQA on a feature selection problem for efficient modeling of drug activities. The successful result of the optimization supported the usefulness of the method, and the possible scope of the applications are widened for even higher dimensional problems.

The crucial point of our methods is that it accesses annealing solvers many times

to solve one optimization problem. This is different from the usual use-cases where the targeted combinatorial optimization problem is formulated into a QUBO and solved at one shot. Our approach is more like solving many subproblems to solve one big and hard problem. The application of annealing solvers as a backend to solve subproblems in classical optimization frameworks is also discussed in (128). As the quantum system and various Ising solvers are not freely scalable, the common approach in the application will be chunking the problems into smaller sizes at which the solvers exceed the performance of classical algorithms.

6.2 Future perspective

Though the empirical benefits of the proposed methods were observed in the experiments, the best practice for using them is yet to be clarified. The most important part is the tuning of hyperparameters, which were manually set in this thesis. The factorization machine requires a fixed factorization size K, whose default value is commonly 8. The local modeling by a factorization machine requires further parameters like the inverse temperature μ, the width σ of a local region, and the weight λ on which the sampling from the proximity is facilitated. A guideline for selecting these values properly are essential for both the theoretical understanding of the methods and the practical use cases. Those knowledge will be pursued by carefully examining each method in the future.

The pros and cons of using a factorization machine as a local model are also not addressed here. Looking at the results, the original FMQA would be better for problems with around few dozens of dimensions, while local FMQA would perform better for larger problems. While we further explore other applications, a criteria for deciding which method to use will be defined, too.

6.3 Different types of annealing solvers

In the experiments of the thesis, only the SA and QA were used for the QUBO minimization part. However, along with the quantum annealing machine, there are numerous annealing algorithms/hardwares available like parallel tempering (129), population annealing (130, 131), Digital Annealer (132), Simulated Bifurcation Machines (133), and

CMOS annealing machines (92). They are very competitive with the QA machines, especially for the problems with dense connections between the bits. As our FMQA models output densely-connected QUBO objectives to be minimized, using those options might improve the performance.

References

[1] Ky Khac Vu, Claudia d'Ambrosio, Youssef Hamadi, and Leo Liberti. **Surrogate-based methods for black-box optimization**. *International Transactions in Operational Research*, **24**(3):393–424, 2017. 2, 6

[2] Nasser Madani, Saffet Yagiz, and Amoussou Coffi Adoko. **Spatial mapping of the rock quality designation using multi-Gaussian Kriging method**. *Minerals*, **8**(11):530, 2018. 3

[3] Masahiro Kanazaki and Naoto Seto. **Efficient Global Optimization Applied to Design and Knowledge Discovery of Supersonic Wing**. *Journal of Computational Science and Technology*, **6**(1):1–15, 2012. 3

[4] Daniel Packwood. *Bayesian Optimization for Materials Science*. Springer, 2017. 3

[5] Yutaka Saito, Misaki Oikawa, Hikaru Nakazawa, Teppei Niide, Tomoshi Kameda, Koji Tsuda, and Mitsuo Umetsu. **Machine-learning-guided mutagenesis for directed evolution of fluorescent proteins**. *ACS synthetic biology*, **7**(9):2014–2022, 2018. 3

[6] Laurens Bliek, Sicco Verwer, and Mathijs de Weerdt. **Black-box combinatorial optimization using models with integer-valued minima**. *Annals of Mathematics and Artificial Intelligence*, pages 1–15, 2020. 3

[7] Carola Doerr. **Complexity Theory for Discrete Black-Box Optimization Heuristics**. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 133–212. Springer International Publishing, Cham, 2020. 3

[8] Ricardo Baptista and Matthias Poloczek. **Bayesian Optimization of Combinatorial Structures**. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, **80** of *Proceedings of Machine Learning Research*, pages 462–471, Stockholmsmssan, Stockholm Sweden, July 2018. PMLR. 3

[9] Akshay Iyer, Yichi Zhang, Aditya Prasad, Siyu Tao, Yixing Wang, Linda Schadler, L Catherine Brinson, and Wei Chen. **Data-centric mixed-variable bayesian optimization for materials design**. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, **59186**, page V02AT03A066. American Society of Mechanical Engineers, 2019. 3

[10] Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. **Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations**. In *International Conference on Artificial Intelligence and Statistics*, pages 3393–3403. PMLR, 2020. 3

[11] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. **Optimization by simulated annealing**. *science*, **220**(4598):671–680, 1983. 4

[12] Peter JM Van Laarhoven and Emile HL Aarts. **Simulated annealing**. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987. 4

[13] Fred Glover and Manuel Laguna. **Tabu search**. In *Handbook of combinatorial optimization*, pages 2093–2229. Springer, 1998. 6

[14] Antonio LaTorre, Santiago Muelas, and Jos-Mara Pea. **A comprehensive comparison of large scale global optimizers**. *Information Sciences*, **316**:517–549, 2015. 6

[15] Daniel Molina, Antonio LaTorre, and Francisco Herrera. **SHADE with iterative local search for large-scale global optimization**. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE, 2018. 6

[16] Nikolaus Hansen, Sibylle D Mller, and Petros Koumoutsakos. **Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)**. *Evolutionary computation*, **11**(1):1–18, 2003. 6, 38

[17] Ilya Loshchilov, Marc Schoenauer, and Michele Sebag. **Kl-based control of the learning schedule for surrogate black-box optimization**. *arXiv preprint arXiv:1308.2655*, 2013. 6

[18] Zbigniew Michalewicz. *Genetic algorithms+ data structures= evolution programs*. Springer Science & Business Media, 2013. 6

[19] Rachid Chelouah and Patrick Siarry. **Genetic and NelderMead algorithms hybridized for a more accurate global optimization of continuous multiminima functions**. *European Journal of Operational Research*, **148**(2):335–348, 2003. 6

[20] Santu Rana, Cheng Li, Sunil Gupta, Vu Nguyen, and Svetha Venkatesh. **High dimensional Bayesian optimization with elastic Gaussian process**. In *International Conference on Machine Learning*, pages 2883–2891, 2017. 8

[21] Kirthevasan Kandasamy, Jeff Schneider, and Barnabs Pczos. **High dimensional Bayesian optimisation and bandits via additive models**. In *International conference on machine learning*, pages 295–304, 2015. 8

[22] Zi Wang, Chengtao Li, Stefanie Jegelka, and Pushmeet Kohli. **Batched high-dimensional Bayesian optimization via structural kernel learning**. In *International Conference on Machine Learning*, pages 3656–3664. PMLR, 2017. 8

REFERENCES

[23] Vasil S. Denchev, Sergio Boixo, Sergei V. Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven. **What is the computational value of finite-range tunneling?** *Physical Review X*, **6**(3):031015, 2016. 9

[24] John Preskill. **Quantum Computing in the NISQ era and beyond**. *Quantum*, **2**:79, 2018. 9

[25] Tadashi Kadowaki and Hidetoshi Nishimori. **Quantum annealing in the transverse Ising model**. *Physical Review E*, **58**(5):5355, 1998. 9

[26] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. **Quantum computation by adiabatic evolution**. *arXiv preprint quant-ph/0001106*, 2000. 9

[27] Salvatore Mandr, Zheng Zhu, Wenlong Wang, Alejandro Perdomo-Ortiz, and Helmut G. Katzgraber. **Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches**. *Physical Review A*, **94**(2):022337, 2016. 9

[28] Katta G Murty and Santosh N Kabadi. **Some NP-complete problems in quadratic and nonlinear programming**. Technical report, 1985. 9

[29] Mark W. Johnson, Mohammad HS Amin, Suzanne Gildert, Trevor Lanting, Firas Hamze, Neil Dickson, R. Harris, Andrew J. Berkley, Jan Johansson, and Paul Bunyk. **Quantum annealing with manufactured spins**. *Nature*, **473**(7346):194, 2011. 10, 15

[30] Alejandro Perdomo-Ortiz, Neil Dickson, Marshall Drew-Brook, Geordie Rose, and Aln Aspuru-Guzik. **Finding low-energy conformations of lattice protein models by quantum annealing**. *Scientific reports*, **2**:571, 2012. 10

[31] Richard Y. Li, Rosa Di Felice, Remo Rohs, and Daniel A. Lidar. **Quantum annealing versus classical machine learning applied to a simplified computational biology problem**. *NPJ quantum information*, **4**(1):14, 2018. 10

[32] Steven H Adachi and Maxwell P Henderson. **Application of quantum annealing to training of deep neural networks**. *arXiv preprint arXiv:1510.06356*, 2015. 10, 15

[33] Marcello Benedetti, John Realpe-Gmez, Rupak Biswas, and Alejandro Perdomo-Ortiz. **Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning**. *Physical Review A*, **94**(2):022308, 2016. 10, 15

[34] Daniel O' Malley, Velimir V. Vesselinov, Boian S. Alexandrov, and Ludmil B. Alexandrov. **Nonnegative/binary matrix factorization with a D-Wave quantum annealer**. *PloS one*, **13**(12):e0206653, 2018. 10

[35] Florian Neukart, Gabriele Compostella, Christian Seidel, David Von Dollen, Sheir Yarkoni, and Bob Parney. **Traffic flow optimization using a quantum annealer**. *Frontiers in ICT*, **4**:29, 2017. 10

[36] Edward Boyda, Saikat Basu, Sangram Ganguly, Andrew Michaelis, Supratik Mukhopadhyay, and Ramakrishna R. Nemani. **Deploying a quantum annealing processor to detect tree cover in aerial imagery of California**. *PloS one*, **12**(2):e0172505, 2017. 10

[37] Jun Cai, William G Macready, and Aidan Roy. **A practical heuristic for finding graph minors**. *arXiv preprint arXiv:1406.2741*, 2014. 11

[38] D.-Wave Sytems Inc. *dwave.embedding.chimera.find_clique_embedding — dwave-system 1.2.1 documentation*. 11, 15

[39] Steffen Rendle. **Factorization machines**. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010. 14

[40] Steffen Rendle. **Factorization machines with libfm**. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **3**(3):57, 2012. 14

[41] Steffen Rendle. *libFM*. 14

[42] Diederik P Kingma and Jimmy Ba. **Adam: A method for stochastic optimization**. *arXiv preprint arXiv:1412.6980*, 2014. 14

[43] Martin A. Green. **Commercial progress and challenges for photovoltaics**. *Nat. Energy*, **1**(1):15015, 2016. 19

[44] Atsushi Sakurai, Kyohei Yada, Tetsushi Simomura, Shenghong Ju, Makoto Kashiwagi, Hideyuki Okada, Tadaaki Nagao, Koji Tsuda, and Junichiro Shiomi. **Ultranarrow-Band Wavelength-Selective Thermal Emission with Aperiodic Multilayered Metamaterials Designed by Bayesian Optimization**. *ACS Central Science*, 2019. 19, 24

[45] Aaswath P. Raman, Marc Abou Anoma, Linxiao Zhu, Eden Rephaeli, and Shanhui Fan. **Passive radiative cooling below ambient air temperature under direct sunlight**. *Nature*, **515**(7528):540, 2014. 19, 21, 22

[46] Quan Pang, Xiao Liang, Chun Yuen Kwok, and Linda F. Nazar. **Advances in lithiumsulfur batteries based on multifunctional cathodes and electrolytes**. *Nature Energy*, **1**(9):16132, 2016. 19

[47] Arumugam Manthiram, Xingwen Yu, and Shaofei Wang. **Lithium battery chemistries enabled by solid-state electrolytes**. *Nature Reviews Materials*, **2**(4):1–16, 2017. 19

[48] Arun Majumdar. **Thermoelectricity in semiconductor nanostructures**. *Science*, **303**(5659):777–778, 2004. 19

[49] Jonathan C. Knight. **Photonic crystal fibres**. *nature*, **424**(6950):847, 2003. 20

[50] John D Joannopoulos, Pierre R Villeneuve, and Shanhui Fan. **Photonic crystals: putting a new twist on light**. *Nature*, **386**(6621):143–149, 1997. 20

[51] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L. McMahon, Takeshi Umeki, and Koji Enbutsu. **A coherent Ising machine for 2000-node optimization problems**. *Science*, **354**(6312):603–606, 2016. 20, 33

[52] Jayakanth Ravichandran, Ajay K. Yadav, Ramez Cheaito, Pim B. Rossen, Arsen Soukiassian, S. J. Suresha, John C. Duda, Brian M. Foley, Che-Hui Lee, and Ye Zhu. **Crossover from incoherent to coherent phonon scattering in epitaxial oxide superlattices**. *Nature materials*, **13**(2):168, 2014. 20

[53] Erwann Bocquillon, Vincent Freulon, J.-M. Berroir, Pascal Degiovanni, Bernard Plaais, A. Cavanna, Yong Jin, and Gwendal Fve. **Coherence and indistinguishability of single electrons emitted by independent sources**. *Science*, page 1232572, 2013. 20

[54] MA Topinka, Brian J LeRoy, RM Westervelt, SEJ Shaw, R Fleischmann, EJ Heller, KD Maranowski, and AC Gossard. **Coherent branched flow in a two-dimensional electron gas**. *Nature*, **410**(6825):183–186, 2001. 20

[55] Jiang Xiao, Gerrit EW Bauer, Ken-chi Uchida, Eiji Saitoh, and Sadamichi Maekawa. **Theory of magnon-driven spin Seebeck effect**. *Physical Review B*, **81**(21):214418, 2010. 20

[56] Shenghong Ju, Takuma Shiga, Lei Feng, Zhufeng Hou, Koji Tsuda, and Junichiro Shiomi. **Designing nanostructures for phonon transport via Bayesian optimization**. *Physical Review X*, **7**(2):021024, 2017. 20

[57] Keith T. Butler, Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. **Machine learning for molecular and materials science**. *Nature*, **559**(7715):547, 2018. 20

[58] Masato Sumita, Xiufeng Yang, Shinsuke Ishihara, Ryo Tamura, and Koji Tsuda. **Hunting for organic molecules with artificial intelligence: molecules optimized for desired excitation energies**. *ACS central science*, **4**(9):1126–1133, 2018. 20

[59] Thaer M. Dieb, Shenghong Ju, Kazuki Yoshizoe, Zhufeng Hou, Junichiro Shiomi, and Koji Tsuda. **MDTS: automatic complex materials design using Monte Carlo tree search**. *Science and technology of advanced materials*, **18**(1):498–503, 2017. 20

[60] Wojciech Paszkowicz. **Genetic algorithms, a nature-inspired tool: Survey of applications in materials science and related fields**. *Materials and Manufacturing Processes*, **24**(2):174–197, 2009. 20

[61] Andriy O Lyakhov, Artem R Oganov, Harold T Stokes, and Qiang Zhu. **New developments in evolutionary structure prediction algorithm USPEX**. *Computer Physics Communications*, **184**(4):1172–1182, 2013. 20

[62] Menaka De Zoysa, Takashi Asano, Keita Mochizuki, Ardavan Oskooi, Takuya Inoue, and Susumu Noda. **Conversion of broadband to narrowband thermal emission through energy recycling**. *Nature Photonics*, **6**(8):535–539, 2012. 21

[63] David M. Bierman, Andrej Lenert, Walker R. Chan, Bikram Bhatia, Ivan Celanovi, Marin Soljai, and Evelyn N. Wang. **Enhanced photovoltaic energy conversion using thermally based spectral shaping**. *Nature Energy*, **1**(6):16068, 2016. 21

[64] Ognjen Ilic, Peter Bermel, Gang Chen, John D. Joannopoulos, Ivan Celanovic, and Marin Soljai. **Tailoring high-temperature radiation and the resurrection of the incandescent source**. *Nature nanotechnology*, **11**(4):320, 2016. 21

[65] Chihhui Wu, Alexander B. Khanikaev, Ronen Adato, Nihal Arju, Ahmet Ali Yanik, Hatice Altug, and Gennady Shvets. **Fano-resonant asymmetric metamaterials for ultrasensitive spectroscopy and identification of molecular monolayers**. *Nature materials*, **11**(1):69, 2012. 21

[66] Shiwen Luo, Jun Zhao, Duluo Zuo, and Xinbing Wang. **Perfect narrow band absorber for sensing applications**. *Optics express*, **24**(9):9288–9294, 2016. 21, 33

[67] X. L. Liu, L. P. Wang, and Z. M. Zhang. **Wideband tunable omnidirectional infrared absorbers based on doped-silicon nanowire arrays**. *Journal of Heat Transfer*, **135**(6):061602, 2013. 21

[68] Kaikai Du, Qiang Li, Weichun Zhang, Yuanqing Yang, and Min Qiu. **Wavelength and thermal distribution selectable microbolometers based on metamaterial absorbers**. *IEEE Photonics Journal*, **7**(3):1–8, 2015. 21

[69] N. I. Landy, C. M. Bingham, T. Tyler, N. Jokerst, D. R. Smith, and W. J. Padilla. **Design, theory, and measurement of a polarization-insensitive absorber for terahertz imaging**. *physical review B*, **79**(12):125104, 2009. 21

[70] Julin Obando, Yonatan Cadavid, and Andrs Amell. **Theoretical, experimental and numerical study of infrared radiation heat transfer in a drying furnace**. *Applied Thermal Engineering*, **90**:395–402, November 2015. 21

[71] Angus R. Gentle and Geoffrey B. Smith. **Radiative heat pumping from the earth using surface phonon resonant nanoparticles**. *Nano letters*, **10**(2):373–379, 2010. 21, 22

[72] Angus R. Gentle and Geoff B. Smith. **A subambient open roof surface under the Mid-Summer sun**. *Advanced Science*, **2**(9):1500119, 2015. 21, 22

[73] Md Muntasir Hossain, Baohua Jia, and Min Gu. **A metamaterial emitter for highly efficient radiative cooling**. *Advanced Optical Materials*, **3**(8):1047–1051, 2015. 22

[74] Zhen Chen, Linxiao Zhu, Aaswath Raman, and Shanhui Fan. **Radiative cooling to deep sub-freezing temperatures through a 24-h daynight cycle**. *Nature communications*, **7**:13729, 2016. 22

[75] Xingshu Sun, Yubo Sun, Zhiguang Zhou, Muhammad Ashraful Alam, and Peter Bermel. **Radiative sky cooling: fundamental physics, materials, structures, and applications**. *Nanophotonics*, **6**(5):997–1015, 2017. 22

[76] Chengjun Zou, Guanghui Ren, Md Muntasir Hossain, Shruti Nirantar, Withawat Withayachumnankul, Taimur Ahmed, Madhu Bhaskaran, Sharath Sriram, Min Gu, and Christophe Fumeaux. **Metal-Loaded Dielectric Resonator Metasurfaces for Radiative Cooling**. *Advanced Optical Materials*, **5**(20):1700460, 2017. 22

REFERENCES

[77] Kai Sun, Christoph A. Riedel, Yudong Wang, Alessandro Urbani, Mirko Simeoni, Sandro Mengali, Maksim Zalkovskij, Brian Bilenberg, C. H. De Groot, and Otto L. Muskens. **Metasurface optical solar reflectors using AZO transparent conducting oxides for radiative cooling of spacecraft**. *ACS Photonics*, **5**(2):495–501, 2017. 22

[78] Yao Zhai, Yaoguang Ma, Sabrina N. David, Dongliang Zhao, Runnan Lou, Gang Tan, Ronggui Yang, and Xiaobo Yin. **Scalable-manufactured randomized glass-polymer hybrid metamaterial for daytime radiative cooling**. *Science*, **355**(6329):1062–1066, 2017. 22

[79] Armande Herv, Jrmie Drvillon, Youns Ezzahri, and Karl Joulain. **Radiative cooling by tailoring surfaces with microstructures: Association of a grating and a multi-layer structure**. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **221**:155–163, 2018. 22

[80] Sarun Atiganyanun, John B. Plumley, Seok Jun Han, Kevin Hsu, Jacob Cytrynbaum, Thomas L. Peng, Sang M. Han, and Sang Eon Han. **Effective Radiative Cooling by Paint-Format Microsphere-Based Photonic Random Media**. *ACS Photonics*, **5**(4):1181–1187, 2018. 22

[81] Mehdi Zeyghami, D. Yogi Goswami, and Elias Stefanakos. **A review of clear sky radiative cooling developments and applications in renewable power systems and passive building cooling**. *Solar Energy Materials and Solar Cells*, **178**:115–128, 2018. 22

[82] MG Moharam and TK Gaylord. **Rigorous coupled-wave analysis of planar-grating diffraction**. *JOSA*, **71**(7):811–818, 1981. 24

[83] Mikhail N. Polyanskiy. *Refractive index database*. 24

[84] Edward D. Palik. *Handbook of Optical Constants of Solids*. Academic Press, Burlington, January 1997. 24

[85] Katherine Han and Chih-Hung Chang. **Numerical modeling of sub-wavelength anti-reflective structures for solar module applications**. *Nanomaterials*, **4**(1):87–128, 2014. 24

[86] Koki Kitai, Jiang Guo, Shenghong Ju, Shu Tanaka, Koji Tsuda, Junichiro Shiomi, and Ryo Tamura. **Designing metamaterials with quantum annealing and factorization machines**. *Phys. Rev. Research*, **2**(1):013319, March 2020. 29, 30, 32

[87] J. M. Zhao and Z. M. Zhang. **Electromagnetic energy storage and power dissipation in nanostructures**. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **151**:49–57, 2015. 30

[88] L. P. Wang and Z. M. Zhang. **Phonon-mediated magnetic polaritons in the infrared region**. *Optics express*, **19**(102):A126–A135, 2011. 30

[89] Diana M. Chambers, Gregory P. Nordin, and Seunghyun Kim. **Fabrication and analysis of a three-layer stratified volume diffractive optical element high-efficiency grating**. *Optics express*, **11**(1):27–38, 2003. 32

[90] Nicholas Chancellor. **Domain wall encoding of discrete variables for quantum annealing and QAOA**. *Quantum Science and Technology*, 2019. 32

[91] Stuart Hadfield, Zhihui Wang, Bryan O'Gorman, Eleanor G. Rieffel, Davide Venturelli, and Rupak Biswas. **From the quantum approximate optimization algorithm to a quantum alternating operator ansatz**. *Algorithms*, **12**(2):34, 2019. 32

[92] Masanao Yamaoka, Chihiro Yoshimura, Masato Hayashi, Takuya Okuyama, Hidetaka Aoki, and Hiroyuki Mizuno. **A 20k-spin Ising chip to solve combinatorial optimization problems with CMOS annealing**. *IEEE Journal of Solid-State Circuits*, **51**(1):303–309, 2015. 33, 53

[93] Gili Rosenberg, Poya Haghnegahdar, Phil Goddard, Peter Carr, Kesheng Wu, and Marcos Lpez De Prado. **Solving the optimal trading trajectory problem using a quantum annealer**. *IEEE Journal of Selected Topics in Signal Processing*, **10**(6):1053–1060, 2016. 33

[94] Peter L. McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, and Kazuyuki Aihara. **A fully programmable 100-spin coherent Ising machine with all-to-all connections**. *Science*, **354**(6312):614–617, 2016. 33

[95] Satoshi Matsubara, Hirotaka Tamura, Motomu Takatsu, Danny Yoo, Behraz Vatankhahghadim, Hironobu Yamasaki, Toshiyuki Miyazawa, Sanroku Tsukamoto, Yasuhiro Watanabe, and Kazuya Takemoto. **Ising-model optimizer with parallel-trial bit-sieve engine**. In *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 432–438. Springer, 2017. 33

[96] Hayato Goto, Kosuke Tatsumura, and Alexander R Dixon. **Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems**. *Science advances*, **5**(4):eaav2372, 2019. 33

[97] Takuya Okuyama, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Masanao Yamaoka. **Binary optimization by momentum annealing**. *Physical Review E*, **100**(1):012111, 2019. 33

[98] Aln Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. **Simulated quantum computation of molecular energies**. *Science*, **309**(5741):1704–1707, 2005. 33

[99] Benjamin P. Lanyon, James D. Whitfield, Geoff G. Gillett, Michael E. Goggin, Marcelo P. Almeida, Ivan Kassal, Jacob D. Biamonte, Masoud Mohseni, Ben J. Powell, and Marco Barbieri. **Towards quantum chemistry on a quantum computer**. *Nature chemistry*, **2**(2):106, 2010. 33

[100] Iulia M. Georgescu, Sahel Ashhab, and Franco Nori. **Quantum simulation**. *Reviews of Modern Physics*, **86**(1):153, 2014. 33

[101] Kenji Sugisaki, Satoru Yamamoto, Shigeaki Nakazawa, Kazuo Toyota, Kazunobu Sato, Daisuke Shiomi, and Takeji Takui. **Quantum chemistry on quantum computers: A polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules**. *The Journal of Physical Chemistry A*, **120**(32):6459–6466, 2016. 33

[102] Peter JJ O' Malley, Ryan Babbush, Ian D. Kivlichan, Jonathan Romero, Jarrod R. McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, and Nan Ding. **Scalable quantum simulation of molecular energies**. *Physical Review X*, **6**(3):031007, 2016. 33

[103] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. **Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets**. *Nature*, **549**(7671):242, 2017. 33

[104] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, and Ryan Babbush. **Quantum chemistry calculations on a trapped-ion quantum simulator**. *Physical Review X*, **8**(3):031022, 2018. 33

[105] Andrew D. King, Juan Carrasquilla, Jack Raymond, Isil Ozfidan, Evgeny Andriyash, Andrew Berkley, Mauricio Reis, Trevor Lanting, Richard Harris, and Fabio Altomare. **Observation of topological phenomena in a programmable lattice of 1,800 qubits**. *Nature*, **560**(7719):456, 2018. 33

[106] Kosuke Mitarai, Tennin Yan, and Keisuke Fujii. **Generalization of the Output of a Variational Quantum Eigensolver by Parameter Interpolation with a Low-depth Ansatz**. *Physical Review Applied*, **11**(4):044087, 2019. 33

[107] Lon Bottou and Vladimir Vapnik. **Local learning algorithms**. *Neural computation*, **4**(6):888–900, 1992. 36

[108] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. **" Why should I trust you?" Explaining the predictions of any classifier**. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016. 36

[109] Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. **A framework for evolutionary optimization with approximate fitness functions**. *IEEE Transactions on evolutionary computation*, **6**(5):481–494, 2002. 37

[110] Jrgen Branke, Christian Schmidt, and Hartmut Schmeck. **Efficient fitness estimation in noisy environments**. In *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, pages 243–250, 2001. 37

[111] Stefan Kern, Nikolaus Hansen, and Petros Koumoutsakos. **Local meta-models for optimization using evolution strategies**. In *Parallel Problem Solving from Nature-PPSN IX*, pages 939–948. Springer, 2006. 38

[112] VP Shilo. **The method of global equilibrium search**. *Cybernetics and Systems Analysis*, **35**(1):68–74, 1999. Publisher: Springer. 38

[113] Oleg Shylo, Dmytro Korenkevych, and Panos M Pardalos. **Global equilibrium search algorithms for combinatorial optimization problems**. In *International Conference on Parallel Problem Solving from Nature*, pages 277–286. Springer, 2012. 38

[114] John E Beasley. **Heuristic algorithms for the unconstrained binary quadratic programming problem**. *London, England*, **4**, 1998. 40

[115] Hugo Kubinyi. *QSAR: Hansch analysis and related approaches*, **1**. VcH, 1993. 44

[116] Graham L Patrick. *An introduction to medicinal chemistry*. Oxford university press, 2013. 44

[117] Louis P Hammett. **Some relations between reaction rates and equilibrium constants.** *Chemical Reviews*, **17**(1):125–136, 1935. Publisher: ACS Publications. 44

[118] Corwin Hansch, Peyton P Maloney, Toshio Fujita, and Robert M Muir. **Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients**. *Nature*, **194**(4824):178–180, 1962. Publisher: Springer. 44

[119] Kamel Mansouri, Nicole Kleinstreuer, Ahmed M Abdelaziz, Domenico Alberga, Vinicius M Alves, Patrik L Andersson, Carolina H Andrade, Fang Bai, Ilya Balabin, Davide Ballabio, and others. **CoMPARA: collaborative modeling project for androgen receptor activity**. *Environmental health perspectives*, **128**(2):027002, 2020. 45

[120] Robert Kavlock, Kelly Chandler, Keith Houck, Sid Hunter, Richard Judson, Nicole Kleinstreuer, Thomas Knudsen, Matt Martin, Stephanie Padilla, David Reif, and others. **Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management**. *Chemical research in toxicology*, **25**(7):1287–1302, 2012. Publisher: ACS Publications. 45

[121] Raymond R Tice, Christopher P Austin, Robert J Kavlock, and John R Bucher. **Improving the human hazard characterization of chemicals: a Tox21 update**. *Environmental health perspectives*, **121**(7):756–765, 2013. Publisher: National Institute of Environmental Health Sciences. 45

[122] Francesca Grisoni, Viviana Consonni, and Davide Ballabio. **Machine learning consensus to predict the binding to the androgen receptor within the CoMPARA project**. *Journal of chemical information and modeling*, **59**(5):1839–1848, 2019. Publisher: ACS Publications. 45, 46

[123] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences, 2017. 45

[124] Roberto Todeschini, Davide Ballabio, Matteo Cassotti, and Viviana Consonni. **N3 and BNN: two new similarity based classification methods in comparison with other classifiers**. *Journal of chemical information and modeling*, **55**(11):2365–2374, 2015. Publisher: ACS Publications. 45

[125] Jiazhong Li, Beilei Lei, Huanxiang Liu, Shuyan Li, Xiaojun Yao, Mancang Liu, and Paola Gramatica. **QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling**. *Journal of computational chemistry*, **29**(16):2636–2647, 2008. Publisher: Wiley Online Library. 46

REFERENCES

[126] N Sukumar, Ganesh Prabhu, and Pinaki Saha. **Applications of genetic algorithms in QSAR/QSPR modeling**. In *Applications of Metaheuristics in Process Engineering*, pages 315–324. Springer, 2014. 46

[127] Ahmed Fawzy Gad. **pygad PyPI**. 46

[128] Nicholas Chancellor. **Modernizing quantum annealing using local searches**. *New Journal of Physics*, **19**(2):023024, 2017. 52

[129] Robert H Swendsen and Jian-Sheng Wang. **Replica Monte Carlo simulation of spin-glasses**. *Physical review letters*, **57**(21):2607, 1986. 52

[130] K Hukushima and Y Iba. **The Monte Carlo Method in the Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm**. 2003. 52

[131] Wenlong Wang, Jonathan Machta, and Helmut G Katzgraber. **Population annealing: Theory and application in spin glasses**. *Physical Review E*, **92**(6):063307, 2015. 52

[132] Maliheh Aramon, Gili Rosenberg, Elisabetta Valiante, Toshiyuki Miyazawa, Hirotaka Tamura, and Helmut G Katzgraber. **Physics-inspired optimization for quadratic unconstrained problems using a digital annealer**. *Frontiers in Physics*, **7**:48, 2019. 52

[133] Hayato Goto. **Bifurcation-based adiabatic quantum computation with a nonlinear oscillator network**. *Scientific reports*, **6**(1):1–8, 2016. 52

Appendix A

Benchmark results on CUBO

A.1 Results on all of the problems

All the results of benchmarking localFMQA on 5 CUBO problems, which were mentioned in chapter 3.2, are shown in figure A.1. For each instance and each method, we conducted 10 times of optimization trials to measure the average performance. A line and its surrounding colored area indicate the mean and the standard deviation calculated by the 10 trials of each configuration.

As we can see from the plots, greedy search is always the fastest of the four methods at the first few hundred steps, but ends up in a local optimum. On the other hand, FMQA seems to be able to find more optimal solutions in the end, though the improvement at the beginning of the optimization is poorer than the greedy search. The scale of gap varies depending on the problem. However, the localFMQA method constantly shows the hybrid property of the greedy search and FMQA methods as stated in the chapter 2. Unless we are sure that only the very small number of evaluations is allowed, the localFMQA is probably the better choice than greedy search or FMQA.

A.2 How the optimization proceeded

Visualizing the sequence of the evaluated configurations gives us how the optimization proceeds. In figure A.2, the 3000 sequencial evaluations by greedy, FMQA, and localFMQA methods are plot. The characteristic description can be made on each method:

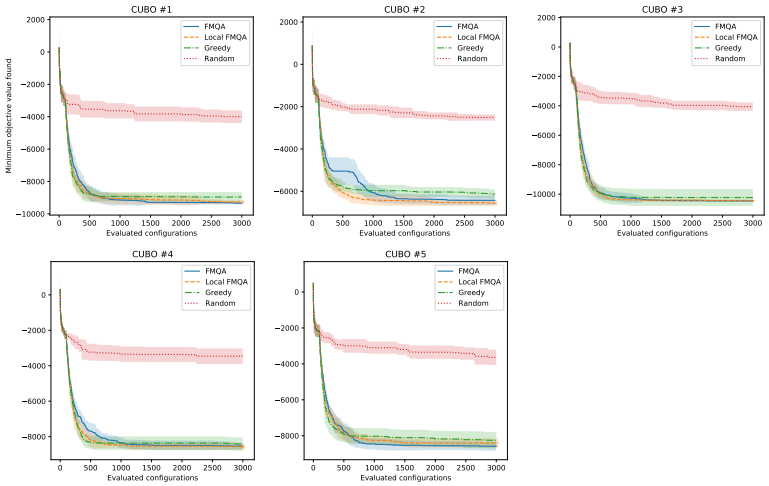


Figure A.1: Results of optimization for all the 5 CUBO problems.

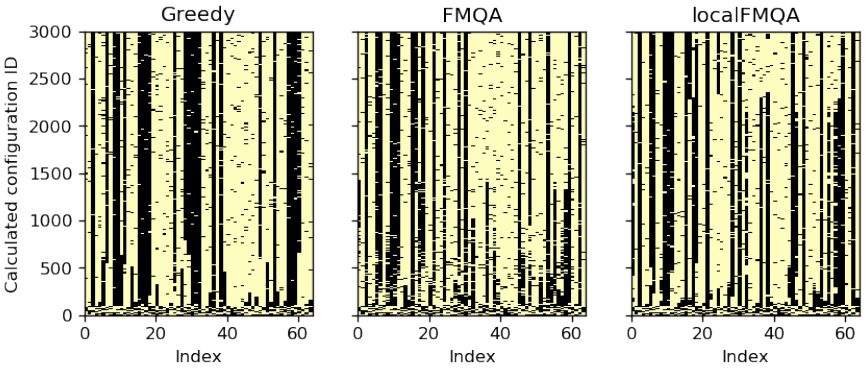


Figure A.2: All the configurations obtained through optimizations.

- In greedy method, the configuration is almost fixed in earlier steps
- In FMQA method, the sequence is highly variable and the plot looks noisier, but finaly find a good solution.
- In localFMQA method, the sequence looks stable like greedy, but drastic changes occur even at later steps.

The observation supports our assumption that the localFMQA method is the hybrid of the greedy-like search method and FMQA. In addition, localFMQA changes few configuration bits at a time, which makes more sense for human than seemingly-noisy FMQA.

Appendix B

Tutorial

Our implementation of FMQA in Python language is available at

https://github.com/tsudalab/fmqa.

The simple usage of the package is shown here.

B.1 Installation

To install the package, clone the repository from the github and move into its root directory at first. We assume a POSIX shell environment in the following instructions.

```
$ git clone https://github.com/tsudalab/fmqa
$ mv fmqa
```

All the installation process such as extracting files and resolving dependencies are managed by the Python's `setuptools` utilities. The package's information is contained in the `setup.py` script and the installation command is in one line as:

```
$ python setup.py install
```

If no problem occurred, the package is ready to use.

We use a simulated annealing solver in this tutorial. Use `pip` command to install it.

```
$ pip install dwave-neal
```

B.2 Example

As an example problem, we would generate 7x7 binary image with 13 pixels being 1 and with minimum edge length. This makes the pixels concentrate rather than scatter.

We consider the edge length as a black-box objective function, and define it as a function `surface` in Python. We also define a function for visualizing the configurations.

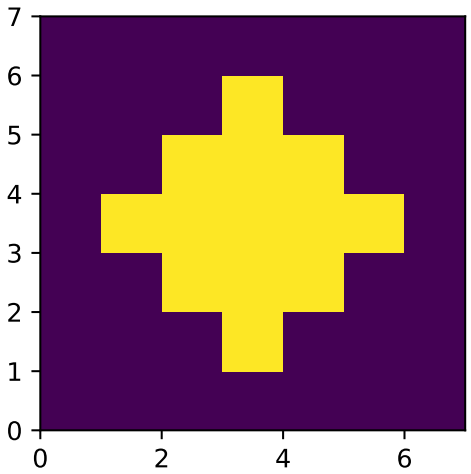
```
import numpy as np
import matplotlib.pyplot as plt

def plot_7x7(x):
    plt.figure(figsize=(3,3))
    plt.pcolor(np.reshape(x, (7,7)))

def surface(x):
    xx = np.reshape(x, (7,7))
    return np.sum(xx[1:,:] ^ xx[:-1,:]) + \
           np.sum(xx[:,1:] ^ xx[:,:-1]) + \
           np.sum(xx[ :, 0]) + \
           np.sum(xx[ :,-1]) + \
           np.sum(xx[ 0, :]) + \
           np.sum(xx[-1, :])
```

An example image below has surface length of 20.

```
x = np.array([
    [0, 0, 0, 0, 0, 0, 0],
    [0, 0, 0, 1, 0, 0, 0],
    [0, 0, 1, 1, 1, 0, 0],
    [0, 1, 1, 1, 1, 1, 0],
    [0, 0, 1, 1, 1, 0, 0],
    [0, 0, 0, 1, 0, 0, 0],
    [0, 0, 0, 0, 0, 0, 0],
]).flatten()
```

```
print("Length = ", surface(x))
plot_7x7(x)
```

```
Length = 20
```



To train our initial FM model, we generate 100 random configurations `xs` and evaluate all of them (`ys`).

```
np.random.seed(12345)
xs = np.random.randint(2, size=(100, 49))
ys = np.apply_along_axis(surface, 1, xs)
```

Our model is an instance of `FMBQM` class defined in the `fmqa` library. A new instance is initialized based on the dataset at the creation time:

```
from fmqa import FMBQM
model = FMBQM.from_data(xs, ys)
```

Now the `model` is ready to be feeded to the optimization solvers. We select the next configuration by simulated annealing algorithm here. The implementation is available in the `neal` package from D-Wave Systems Inc.

```
from neal import SimulatedAnnealingSampler
sa_sampler = SimulatedAnnealingSampler()
```

We haven't coded the constraint that the 13 pixels should take value 1, yet. The constraint can be represented by QUBO problem, too.

$$\min_{\mathbf{x}}\Big(\sum_{i=1}^{d} x_i - 13\Big)^2 = \min_{\mathbf{x}}\Big\{\sum_{i<j} 2x_i x_j - \sum_{i=1}^{d} 25x_i\Big\} \tag{B.1}$$

We encode it to an instance of `BQM` class in D-Wave's `dimod` package.

```
from dimod import BQM
constraint = BQM.from_qubo(np.triu(np.ones((49,49))*2, k=1) - \
                           np.diag(np.ones(49)*25))
```

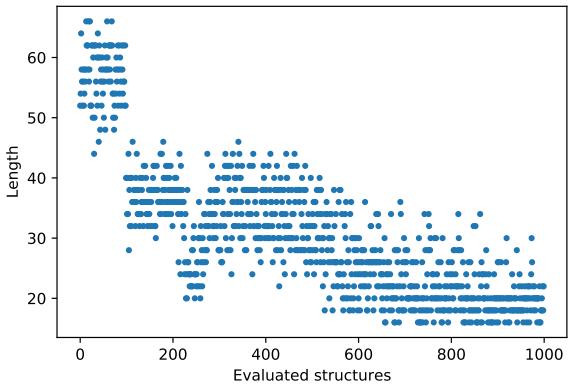
The `constraint` is added to the `model` to generate the solution which satisfies the constraint while minimizing the predictive value.

```
def contains(xs, x):
    ''' The utility function to check if
    the solution x is already contained
    in the dataset xs. '''
    return np.any(np.all(xs == new_x[None,:], axis=1))

# Main loop
for n in range(900):
    # Add constraint with weight 0.25
    model_constrained = BQM.from_qubo(model.to_qubo()[0])
    for i, qii in constraint.iter_linear(): # Linear term
        model_constrained.add_variable(i, qii * 0.25)
    for i, j, qij in constraint.iter_quadratic(): # Quadratic term
        model_constrained.add_interaction(i,j, qij * 0.25)

    # Sampling (optimization) by SA
    new_x = sa_sampler.sample(model_constrained).record['sample'][0]
```

```
    # Apply random editting if the predicted
    # minimizer new_x is already contained in xs
    while contains(xs, new_x):
        new_x[np.random.randint(49)] ^= 1

    # Evaluation of the new_x
    new_y = surface(new_x)

    # Append the new data to the dataset
    # and update the model
    xs = np.r_[xs, [new_x]]
    ys = np.r_[ys, [new_y]]
    model.train(xs, ys, num_epoch=100)
```

Here we plot the result of 1000 times of evaluations.

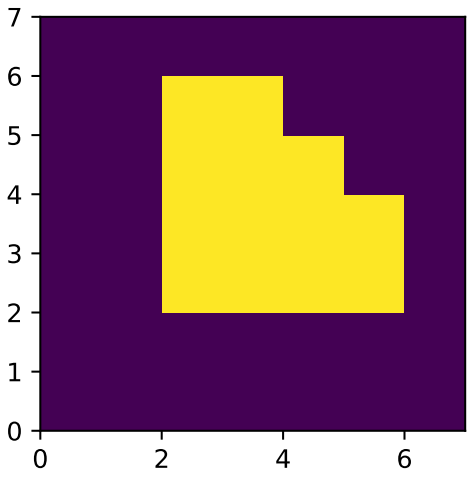
```
plt.plot(ys, '.')
plt.ylabel('Length')
plt.xlabel('Evaluated structures')
```



We can see that after the 100 times of random samplings, the algorithm selectively sampled the configurations with relatively less lengths. Among the all of the 1000 evaluations, the best found configuration is shown below.

```
satisfied = constraint.energies(xs) == -169
best_conf = xs[satisfied][np.argmin(ys[satisfied])]
plot_7x7(best_conf)
print("Length = ", surface(best_conf))
```

```
Length =  16
```



The actual minimum value 16 is found. Note that the solution is solery found on the BBO setting. No information about the objective function such as geometry of the pixels or methematical expression is used.

Declaration

I herewith declare that I have produced this paper on my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research.

The thesis work was done under the guidance of Professor Koji Tsuda at The University of Tokyo.