Overview of Multiobjective Optimization Methods in *in Silico* Metabolic Engineering

Kauthar Mohd Daud¹*, Zalmiyah Zakaria¹, Agus Kunayat², Zuraini Ali Shah¹, Rohayanti Hassan³

¹Artificial Intelligence and Bioinformatics Research Group, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia.
²School of Industrial Engineering, Telkom University, 40257 Bandung, West Java, Indonesia.
³Software Engineering Research Group, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia.

Received 28 June 2018; accepted 5 August 2018, available online 24 August 2018

Abstract: Multiobjective optimization requires finding a trade-off between multiple objectives. However, most of the objectives are contradict towards each other, thus makes it difficult for the traditional approaches to find a solution that satisfies all objectives. Fortunately, the problems are able to solve by the aid of Pareto methods. Meanwhile, in *in silico* Metabolic Engineering, the identification of reaction knockout strategies that produce mutant strains with a permissible growth rate and product rate of desired metabolites is still hindered. Previously, Evolutionary Algorithms (EAs) has been successfully used in determining the reaction knockout strategies. Nevertheless, most methods work by optimizing one objective function, which is growth rate or production rate. Furthermore, in bioprocesses, it involves multiple and conflicting objectives. In this review, we aim to show the different multiobjective evolutionary optimization methods developed for tackling the multiple and conflicting objectives in *in silico* metabolic engineering, as well as the approaches in multiobjective optimization.

Keywords: Multiobjective Evolutionary Algorithm, Pareto methods, *in silico* metabolic engineering, optimization, constraint-based methods.

1. Introduction

With the increased price of petroleum-based fuel, depletion of fuel sources, and environmental issues, biofuel has become the alternative choice. Derived from the biomass conversion from living organisms and rapid development in microbial engineering technology, organisms with profitability and capability in maximizing the production rate of important industrial metabolites have been gaining importance in the last few years. Hence, methods such as Flux Balance Analysis (FBA), Minimization of Metabolic Adjustment (MOMA), and Regulatory On/Off (ROOM), has been developed to simulate the genome-scale metabolic models of organisms, in order to exploit the usefulness of the models.

Various techniques have been proposed, and one of them is genes/reactions knockout. Gene/reaction knockout is simulated by identifying gene/reaction that may possibly increase the biological objective function, mainly production rate or growth rate [1]. Normally, in order to ensure the organisms are viable after perturbations, a bi-level optimization is formulated. In this case, the organism is forced to produce the desired products and at the same time, keeping the viability of the organisms. In computation, however, the bi-level optimization is focused on optimizing a single objective. They also produced one single near-optimal solution of the problem.

As mentioned before, most of the developed methods such as OptKnock, OptFlux, and OptGene, are only focusing on one objective [2–4]. Nevertheless, in bioprocesses, mainly it involves several of other objectives such as growth rate, byproduct formation, desired product yield, and others. Therefore, several methods and techniques have been developed in optimizing more than one objective. Furthermore, multiobjective optimization has significantly shown more benefits compared to single objective optimization [5].

However, unlike single objective optimization, multiobjective optimization involves optimizing multiple conflicting objectives. As an example, in E.coli, the production of succinate acid is at the highest rate when the growth rate is at 0 and vice versa. Thus, it is necessary
to determine a set of trade-off points that represent the near-optimal solutions. Despite that, these points are no preference, usually, the decision makers will scan through the solutions and decided on one final solution based on their own preferences. Generally, multiobjective optimization provides a set of solutions that trade-off between conflicting objectives.

In this paper, we aim to review the methods and techniques of multiobjective optimization in metabolic engineering. The paper is organized as follows: Section 2 describes the general definition of multiobjective optimization. The following Section 3 describes the overview of multiobjective optimization methods and approaches in handling multiple conflicting objectives. Section 4 introduces the multiobjective optimization methods in metabolic engineering. Lastly, Section 5 gives the conclusion, including trend, future directions, and factors that hindered the multiobjective optimization in metabolic engineering.

2. Multiobjective Optimization

Optimization is defined as maximizing or minimizing a function from a set of decision variables, that is restricted by a series of constraints [6]. Optimization can be divided into two problems depending on the number of objective function being optimized; (1) single objective optimization (SO) and (2) multiobjective optimization (MO). The former optimization is related to optimize a single objective function, whereas the latter involves more than one objective being optimized. However, the objective functions being optimized are always conflicting to each other, thus a trade-off among the solutions need to consider.

The trade-off is losing a thing in order to gain another thing. For a solution, therefore, it may be good for one function, but it may be bad for another function. The mathematical expressions for MO problems can be expressed as follows:

\[
\min/\max Z = Z(x*) = [Z_1(x*), Z_2(x*), ..., Z_k(x*)]\]

Subject to:

\[
g_j(x*) = b_j (j = 1, 2, ..., m)\]

where \(k\) is the number of objectives to optimized, \(m\) is the number of constraints and \(g\) is the constraints being imposed to the solution space. Fig. 1 below illustrates the multiobjective optimization problems.

Fig. 1: Illustration of multiobjective optimization.

Based on Fig 1, there are an \(n\)-dimensional decision variable vectors \((x_1, x_2, ..., x_n)\) initialized in the solution space \(X\), and find a vector of \(x^*\) that optimizes the set of \(k\) objective functions \(Z(x*)\). The solution space is restricted by a series of constraints, \(b_j\). In MO, an acceptable, in this case feasible, solutions are the solutions that satisfy the objectives without being dominated by any other solution. Furthermore, in MO, dominance is crucial in determining a goodness of a solution. A solution \(x\) is said to dominate another feasible solution \(y\) when: (1) the solution \(x\) is no worse than the solution \(y\) in all objectives and (2) solution \(x\) is better than solution \(y\) in at least one objective. In this case, solution \(x\) is said to be a non-dominated solution, whereas solution \(y\) is a dominated solution.

These solutions are known as Pareto solutions and when mapped to a graph, it is known as Pareto graph [7]. In MO, the obtained non-dominated solutions should as close as possible to the true Pareto solution and uniformly distributed over the Pareto graph while capturing the whole graph. Once the non-dominated solutions are found, the decision makers may decide on the final solution based on their own preference according to the optimization problems.

3. Approaches in Multiobjective Optimization

3.1 Overview of Multiobjective Optimization Approaches

The MO can be tackled in two approaches: (1) traditional approach and (2) Pareto methods, as shown in Fig. 2. The first approach involves analytical and numerical method. Among them are scalarization methods, which include weighted sum approach, goal attainment, and lexicographic method; and non-Pareto methods, which include \(\varepsilon\)-constraint. However, the numerical method requires mathematical equations including defining the iteration [8]. However, the former approach is able to generate one solution at each iteration and they are sensitive to the shape of Pareto curve, although they have fast convergence and high searching efficiency.

Therefore, in order to overcome the limitations of traditional approaches, the Pareto-based approach has been introduced and developed. The Pareto-based approach can be further divided into non-evolutionary algorithms and evolutionary algorithms. The method in non-evolutionary Pareto-based approach is Normal Boundary Intersection (NBI). However, NBI is only suitable for maximum two objectives and the generated non-dominated solutions are not guaranteed to be a near-optimal [9].

Nowadays, intelligent algorithms such as those inspired by true-nature events are well-known in solving the optimization algorithms. Thus, evolutionary based algorithms such as Genetic Algorithm (GA), Harmony Search Algorithm (HS), Differential Search Algorithm (DSA), and others, has been used in solving the MO [10–13]. The self-adaptation and flexibility of these Pareto-based evolutionary algorithms have successfully solved various MOP including electrical flow, scheduling, an engineering problem, and others.
3.2 Differences in Multiobjective Evolutionary Algorithms (MOEA)

The framework in MOEA can be distinguished into three categories, which are (1) fitness assignment, (2) elitism, and (3) diversification. Each of these categories corresponds to the different goals of multiobjective optimization. Fig. 3 represents the methods in each category.

According to [14], most researchers developed their algorithms by adapting strategies from these differences. Each of these methods has been thoroughly reviewed in [14]. Fitness assignment is used to obtain non-dominated solutions that are near to the true Pareto. There are three differences, which are weighted sum approach, altering the objective function and Pareto ranking. In weighted sum, a weight is assigned to each objective function, and the sum of total weight used is equal to 1. A weighted sum is a classical approach that has been applied in WBGA-MO, MOGA, RWGA, and others; due to the simplicity of implementation and computationally efficient [15,16].

Fig. 3: Differences in Multiobjective Evolutionary Algorithms.

Meanwhile, elitism is used to ensure the obtained non-dominated solutions are able to cover the whole range of true Pareto. There are two approaches, either maintaining the elitist solutions or store the elitist solutions in an external archive. The first approach is easy to implement, but not suitable for a large number of non-dominated solutions, while the latter is time inefficient, although it is able to keep the previous non-dominated solutions without being replaced by new non-dominated solutions.

The third category is diversification, which is important in allowing the solutions to be uniformly distributed along the Pareto graph. There are three approaches, crowding distance, cell-based density and fitness sharing. The most popular approach is crowding distance as it does not require a user-defined parameter. Moreover, crowding distance can be used as a parameter to determine the density of a solution. However, it does not suitable for a small number of the population [18].
algorithms including PESA, SPEA2, SPEA, and PAES [19, 20].

For fitness sharing, the solutions in a densely populated area are assigned a penalty to its fitness in order to search for unexplored sections. However, this method is computationally expensive as it requires niche count defined by the user. Table 1 shows the strategies and approaches used in multiobjective evolutionary algorithms (MOEA).

Table 1: List of Multiobjective Evolutionary Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fitness Assignment</th>
<th>Diversification</th>
<th>E</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGA [21]</td>
<td>Altering objective function</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MOGA [22]</td>
<td>Pareto ranking</td>
<td>Fitness sharing</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>NPGA [23]</td>
<td>X</td>
<td>Fitness sharing</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>WBGA [15]</td>
<td>Weighted Sum</td>
<td>Fitness sharing</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RWGA [16]</td>
<td>Weighted Sum</td>
<td>Fitness sharing</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NSGA [24]</td>
<td>Pareto ranking</td>
<td>Fitness sharing</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SPEA [25]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SPEA2 [20]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PAES [17]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PESA [19]</td>
<td>X</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PESA-II [19]</td>
<td>X</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NSGA-II [26]</td>
<td>Pareto ranking</td>
<td>Crowding distance</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>MEA [27]</td>
<td>Pareto ranking</td>
<td>Fitness sharing</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Micro-GA [28]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RDGA [29]</td>
<td>Solve MO as single objective</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DMDOEA [30]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: Checkmark (✓) represents it being used in the algorithm and cross mark (X) represents not being used in the algorithm.

A – Archive, E - Elitism

4. Multiobjective Optimization Algorithms in Metabolic Engineering

Metabolic engineering is a process to increase the production of certain metabolites by optimizing the metabolic and biosynthetic pathways of an organism [31]. The aim of metabolic engineering is to improve the design of organisms by means of (1) gene/reaction knockout, (2) modification of specific regions in metabolic network that may contribute in enhancing the production yield, (3) manipulation of metabolic networks using various existing network reconstruction tools and manipulation of biological molecule using biological molecule manipulation tools, and (4) integrating new non-native pathway into the host. To date, most researchers focused on improving the metabolic network despite simplicity yet full information resides and can be gained from the manipulation of the metabolic network.

Due to this, several methods and tools have been developed, including constraint-based methods. Furthermore, the constraint-based methods such as FBA, ROOM, and MOMA, has been coupled together with an optimization algorithm, due to the nature of constraint-based, which are only able to find the flux values and not optimizing the production. Therefore, there are new methods developed, including Flower Pollination-Clonal Selection Algorithm, IdealKnock, OptGene, RobustKnock and others [3, 32–34]. These methods are able to find mutants with a high value of production rate and growth rate. Furthermore, the aforementioned methods work by identifying reaction knockout that may improve the production of desired metabolites while keeping the organism viable.

However, previous research in in silico metabolic engineering are only focusing on optimizing one single objective, majority production rate. Yet, in bioprocesses, it involves multiple and conflicting objectives such as production rate of desired metabolites, growth rate, and byproduct rate. Therefore, current focus has shifted towards multiobjective optimization. Not only in this domain but other domain as well [35–37]. Nevertheless, the multiobjective optimization in in silico metabolic engineering is still new.

In metabolic engineering, the important factors that need to be considered are production rate and growth rate. This is because the target of the mutant is not only producing the promising amount of desired metabolites but also viable after the extreme perturbations. Organisms that largely manipulate in large scale are Escherichia coli and Saccharomyces cerevisiae. Considering that their metabolic and biological information are studied tediously and most updated, therefore most research used these organisms to manipulate for producing products in bulk forms such as ethanol, succinic acid, and acetic acid.

Roughly, the developed MOEAs are mostly due to the limitation of FBA that only limited to single objective function. The earliest multiobjective optimization in enhancing the production of succinic acid is carried out by [38]. The authors applied Strength Pareto Evolutionary Algorithm 2 (SPEA2) and Non-Dominated Sorting Genetic Algorithm II (NSGA-II) in identifying reaction knockout strategies in E.coli to optimize the production rate and growth rate. This finding has kick-start for other developed methods, including LPPFBA, NISE and FBA, and Metaboflux [39–42]. Table 2 shows the list of MOEA in in silico metabolic engineering.

Table 2: List of MOEA in Metabolic Engineering

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fitness Assignment</th>
<th>Diversification</th>
<th>Elitism</th>
<th>Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPPFBA [39]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>NISE+FBA [40]</td>
<td>Weighted sum approach</td>
<td>Cell-based density</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>NSGA-II [38]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SPEA2 [38]</td>
<td>Pareto ranking</td>
<td>Cell-based density</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Metaboflux [42]</td>
<td>Altering objective function</td>
<td>Fitness sharing</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: Checkmark (✓) represents it being used in the algorithm and cross mark (X) represents not being used in the algorithm.
Linear Physical Programming-based Flux Balance Analysis (LPPFBA) is developed due to the limitation of FBA, which only focus on single objective function. LPPFBA is applied to hepatocyte function in a bioartificial liver system for determining a set of optimal solutions for various pairs of urea secretion, albumin, NADPH, and glutathione syntheses. Although LPPFBA is suitable for more than 2 objectives, however, the user needs to define degrees of significance for each objective function.

Following that, Noninferior set estimation (NISE) has been applied with FBA to improve the production of poly (3-hydroxybutyrate) in *E. Coli*. NISE method is used to estimate the non-dominated near-optimal solutions. Furthermore, NISE is able to give a good approximation of Pareto set, however, it does not consider enzymatic information. Meanwhile, Metaboflux is developed for exploiting the metabolic network of an organism, thus allows the incorporation of multipurpose characteristics of a cell. Eventually, it contributes to the significance of a model, although it is time inefficient.

Furthermore, there is another research that finding the combination of reactions for the knockout and simulate them in the experimental laboratory [43]. Using *E.coli* strains [44], they focus on increasing the production of target organic acids, including acetic acid, lactic and succinic acids, while minimizing the formation of byproducts. By using Flux Balance Analysis and based on criteria defined, they obtained 4 mutants for different target organic acids.

5. Conclusion

Most real-world problems are centered upon multiobjective. This include designing, scheduling, controlling, and others in various areas such as economics, financial, electrical power systems, and others. Not so long ago, the traditional chemical synthesis processes have been shifting towards computational simulation due to the benefits in terms of time, raw materials source, and prior knowledge. Additionally, the multiobjective optimization has extended in computational biology and bioinformatics. Even though, customizing together these approaches and strategies may introduce a new multiobjective algorithm.

At last, we focus on multiobjective optimization problem in solving the metabolic engineering problem. Several algorithms that have been developed are reviewed as well, together with advantages and disadvantages. As mentioned before, there are three goals associated with MOP, and each goal is distinct with their strategies, approaches, and functionalities. Regardless of the difference, customizing together these approaches and strategies may introduce a new multiobjective algorithm. Still, further validation is still needed in both biological and computational.

References


