An optimal generic model for multi-parameters and big data optimizing: a laboratory experimental study

To cite this article: D N Utama et al 2018 J. Phys.: Conf. Ser. 978 012045

View the article online for updates and enhancements.
An optimal generic model for multi-parameters and big data optimizing: a laboratory experimental study

D N Utama¹, N Ani¹, M M Iqbal²
¹Computer Science, Universitas Mercu Buana, Jakarta, Indonesia
²Information Systems Department, UIN Syarif Hidayatullah, Jakarta, Indonesia
ditdit.nugeraha@mercubuana.ac.id

Abstract. Optimization is a process for finding parameter (parameters) that is (are) able to deliver an optimal value for an objective function. Seeking an optimal generic model for optimizing is a computer science study that has been being practically conducted by numerous researchers. Generic model is a model that can be technically operated to solve any varieties of optimization problem. By using an object-oriented method, the generic model for optimizing was constructed. Moreover, two types of optimization method, simulated-annealing and hill-climbing, were functioned in constructing the model and compared to find the most optimal one then. The result said that both methods gave the same result for a value of objective function and the hill-climbing based model consumed the shortest running time.

1. Introduction
Efficiency is one factor considered to choose an optimization model. A model’s run time consumption is one of some variables that is able to be taken into account to see the efficiency of a model including the optimization model. Where, to develop a generic optimization model which is efficient is a challenging study. Besides efficient, it is going to be valuable to solve any types of optimization problem without changing any things to become suitable to the case. Also, it can be operated to find the precise decision that something problematic to get [1].

Numerous researchers have conducted research in optimizing field that implemented in several cases. [2] developed an optimization model to minimize the maximum index of the used frequency slots. The model technically constructed based on a combination of heuristic and genetic algorithms. The methods were used to sort connection requests and look for an optimal routing and spectrum assignment scheme. [3] developed two types of signal optimization model for addressing issues in asymmetric continuous flow intersection (CFI). The first model uses two step procedures, and the second one takes mixed-integer-linear-programming. The result revealed that the second model is more flexible in designing phase, however the first one performs better in reducing link queue length. Moreover, [4] constructed an optimization model that realized in decision support model to optimize a placement of stations in urban areas. Two parameters considered in this study; the number of reachable household and the cost of overall e-transportation energy. And [5] used the fuzzy evaluation method to create an optimization model for rationalizing and optimizing the structure of passenger transportation. The model can evaluate the condition of passenger transportation structure of both conditions; in model and real domain.

Furthermore, [6] constructed an optimization model to investigate a performance of underground gas drainage system. Several factors were considered in the study, e.g. pure gas drainage flow, gas
concentration, and energy efficiency. The model was practically applied to a typical modern coal mine.

Then, [7] [8] have developed a decision support system based on optimization model. The model was developed based on the method water flow algorithm (WFA) and a combination between fuzzy-logic with WFA (called FWFA). It used to test the value of objective function of each decision alternatives that considers several parameters, e.g. traffic volume, density, etc.

This paper portrays the development of generic optimization model that is realized in solving the multi parameters and big data optimization case. Two types of optimization method, simulated-annealing (SA) and hill-climbing (HC), are a main method of the model that have been tested to see their efficiency of running time. The introduction section of the paper is followed by following sections; related works, research methodology, results and discussion, and conclusion and further works.

2. Related works
A generic method or model, as a main topic, has been studied by several researchers in many fields. In the field computers and operations research [9] proposed a generic Pareto based local search metaheuristic. It was used for marketing campaign. Here, a Greedy Randomized Neighborhood Structure was designed which was embedded in a multi-objective local search metaheuristic. In the field microprocessors and microsystems, [10] constructed a generic model to optimize the energy consumption of mobile platform. The parameters that taken into account in this study are power, response time, and thermal. Furthermore, in the field computer aided chemical engineering, [11] proposed a generic computer-aided methodology for synthesizing different processing networks. In this study, a superstructure optimization was operated. The methodology was able to handle different network optimization problems in many application fields. And also, in the environmental management field, [12] developed a generic methodology for the design of a self-optimizing controller in sewer systems, to keep the systems close to optimal performance. The result said that the proposed methodology has been successfully implemented and it is promising to contribute to performance increase of sewer systems.

3. Research methodology
In this study, two methods were technically functioned in developing a generic optimization model. They are methods simulated-annealing [13] [14] [15] [16]and hill-climbing [17] [15]. A class diagram, that is a part tool of object oriented method [17-18], was operated to design a schematic view of the constructed model. The laboratory experiments were practically conducted by involving a big data (a hundred thousand records with forty specified parameters, or similar to four million fields). They were done in Laboratory Optimization Models and Systems for Decision Support (UIN Syarif Hidayatullah, Jakarta, Indonesia) by using machine with processor Intel® Celeron® CPU B815 @ 1.60GHz (2 CPUs), memory 2048MB RAM, BIOS InsydeH2O V 1.06, and operating system Windows 8.1 Pro 64-bit (6.3, Build 9600). Those models also were tested in specific case (in a decision support model for renovating and rehabilitating mosques [19]).

The objective function used in the experiments is generically formulated in equation (1), where $n$ represents a number of parameter (in the experiments, 40 used to define $n$), $c_i$ is a $i$th parameter coefficient, and $P$ is a value of $i$th parameter. It means that the objective function is to obtain a maximum value of a summation function of forty coefficients and parameters multiplication.

$$\text{Max } f(x) = \sum_{i=1}^{n} c_i P_i$$  \hspace{1cm} (1)

4. Results and discussion

4.1. The constructed model
Schematically, the constructed model is configured by Figure 1. This configuration is technically depicted by class diagram. Here, the constructed model fundamentally consists of two main classes: OptimizationModel and SelectedCase. The class OptimizationModel has two types of model, SAOptimizationModel and HCOptimizationModel; where they have attributes and operations that commonly belong to class OptimizationModel; only several attributes and operations that specifically belong to them. The class SelectedCase indicates that the model is generic, it can be operated in any types of case. The case totally depends on parameters and parameters’ coefficient going to be considered. The objective function of optimization itself is defined in operation definingOF (defining objective function) in class OptimizationModel. For the explanations of other part can be seen in Figure 1 clearly.

![Class Diagram](image)

**Figure 1.** The generic class diagram of constructed model.

The generic pseudocode of procedure for SA and HC methods are showed in Code 1 and 2 respectively. Basically, SA was proposed by [13] and also by [14]. It is a probabilistic method for finding the global minimum of an objective function. In principle, it can possibly deliver several local minimum [20] after optimizing process. The idea of SA itself firstly came from the metallurgy field. The objective function is treated as the energy function of a molten metal exposed to an artificial temperature which is gradually cooled [21]. The temperature controls the probability that the optimizing process moves to a neighbor with a worse value of the objective function (as opposed to the HC method when such a move can never occur). The process will finish when the adjusted temperature reaches the optimal temperature. This scenario can still deliver a local optimum which is not the global optimum, but the algorithm has a chance to escape from this trap as long as the temperature is still high enough.

The variable delta is a basic conception in the SA method. As, when curVal (current value) is better than bestNeighbor, the delta is realistically operated to check the probability to move or to stop by calculating “exp(-delta/baseTemp)” and then comparing it to a value of random (0...1). This condition makes SA more flexible than HC. Can be concluded here that the other specific parameters used in simulated annealing models are baseTemp and optTemp.
Procedure SimulatedAnnealing(parameters)

Begin
  ...
  // randomizing new parameter combination and getting value of
  // objective function
  para <-- random()
  curVal <-- objFunction(para)

  // looping until local optimum is found
  While(baseTemp>=optTemp)
  
  
  End while

End

Procedure HillClimbing(parameters)

Begin
  ...

End

Furthermore, HC method is a local search method. It was offered by [15]. Straightforwardly stated, it contains a loop that continually moves in one direction of an increasing value (uphill), and never goes downhill. It is terminated when it reaches a ‘peak’, where there is no neighbor having a higher value. The probability to be trapped in one local maximum (one hill) is high. A local maximum is a peak that is higher than its neighbors but lower than the global maximum [22].

In Code 2, it is obviously mentioned that a looping process for seeking the optimum value can be terminated when the local optimum value is discovered. This occurs when the variable curVal is better than the value of the variable bestNeighbor. This means, no parameter combination that belongs to its neighbors can give a better value of the objective function than the variable curVal.

Several specific parameters required by HC method can be listed here. The parameters curVal, bestNeighbor, and bestVal are parameters necessitated in this method. They are respectively current value, best value of neighbor, and best value (should be as a local or global best value).

Figure 2 describes distributed data of execution experiment result for each models. The blue dots are depicting distributed data coming from SA based model’s running time, while the read ones illustrate data of HC based model’s running time. The experiments were conducted by manipulating a number of searching looping, 1 – 1000 looping times. Based on a number of searching looping, the running time increases linearly.

\[
f(x) = 0.0007x - 0.0259 \tag{2}
\]

\[
f(x) = 0.0003x - 0.0205 \tag{3}
\]

SA based model gives a linear increase pattern (trend-line) of running time in equation (2), where \( x \) represents a number of searching looping, with \( R^2 \) is 0.8085; while, HC based model produces a pattern with equation (3) with \( R^2 \) is 0.6293. The equation (2) indicates that SA based model consumes longer running time than HC based model. It naturally happens, since in SA method, the optimal value searching process is still conducted when it has not reached an optimal temperature yet; even though the search starting process in the peak of hill. Conversely, in HC based model, the searching process is directly terminated when it gets the optimal value (peak of hill).

4.2. Discussion

The model practically was implemented to tackle the specific problem [19]. However, in that study, only 196 records were used and the numbers of searching looping used are 100, 500, 1,000, 5,000,
10,000, 50,000, and 100,000 times. On the other hand, in this study, we generated 100,000 records (with forty valued parameters) by testing sequentially a number of searching looping until 1,000 times. Based on the result of laboratory experiments for the running time of model execution, both studies exposed a similar pattern for both methods.

5. Conclusion and further works
The generic optimization model to solve a big data problem was successfully constructed. The model consists of two types of optimizing method; SA and HC. The laboratory experiments were practically conducted to test the running time of 1 – 1,000 searching loops for both model. The final conclusion revealed that both models produced a linear pattern of running time, and the model based on HC has shorter running time. This it is more optimal than SA based model with similar capability in searching. Both models also have been implemented in specific case to recommend the best value of decision.

In addition, for the further study, the models essentially can be operated to solve several types of problem of decision making case. The models can be rationally implemented in bigger records as well, e.g. one million or billion records with more than forty parameters taken into account.

References
[4] Yi Z and Bauer P H 2016 Optimization models for placement of an energy-aware electric vehicle charging infrastructure Transportation Research Part E: Logistics and Transportation Review 91 pp 227-244
Microsystems 40 pp 74-87


