SIMULATION-BASED ANALYSIS OF FITNESS LANDSCAPE IN OPTIMISATION

IMITĀCIJAS MODELĒŠANĀ BĀZĒTA PIEMĒROTĪBAS AINAVAS ANALĪZE OPTIMIZĀCIJĀ

Galina Merkuryeva, Prof., Dr. habil., Department of Modelling and Simulation, Riga Technical University, 1 Kalku Street, Riga LV-1658, Latvia, e-mail: galina.merkurjeva@rtu.lv

Vitalijs Bolshakovs, Mg.sc.ing, Ph.D. student, Department of Modelling and Simulation, Riga Technical University, 1 Kalku Street, Riga LV-1658, Latvia, e-mail: vitalijs.bolsakovs@rtu.lv
Keywords: fitness landscape, simulation optimisation, statistical measures, information measures, vehicle scheduling

Abstract – Simulation-based analysis of fitness landscape with application to optimisation problems is discussed in the paper. Methods of analysis of fitness landscapes and measures known in literature are reviewed. Procedure for simulation-based analysis of fitness landscape is introduced. Software prototype to perform this analysis is described. Case study for a vehicle scheduling problem with the time window constraints is given and demonstrates the main steps of fitness landscape analysis applied to simulation optimisation problem.

Introduction

Nowadays evolutionary algorithms are often used to solve complex optimisation problems. However, in some cases they may be not efficient. The suitability of an evolutionary algorithm for optimisation problem can be determined with use of analysis of fitness landscape. In this paper fitness landscape analysis techniques and their use for optimisation problems solution are described. Special attention is made to simulation-based fitness landscape analysis, as simulation provide a powerful tool for evaluation of the complex system performance.

Some statistical and information measures for fitness landscape analysis are described in this paper. Development of a software prototype for analysis of fitness landscape is also described. Procedure and algorithms for simulation-based analysis of fitness landscape are introduced. The developed procedure is tested on a vehicle scheduling problem with time windows. Simulation experiments for analysis of the problem fitness landscape and determination of its hardness for evolutionary algorithms are made.

Fitness Landscape Analysis

Definition and Interpretation

A fitness landscape analysis presents a powerful tool for visualising and investigating evolutionary processes. The fitness landscape comprises of three main elements: the set of genotypes, the fitness function that evaluates the genotypes, and the genetics operators that define neighbourhood relations between the genotype sets [1].

The fitness landscape can be represented by a surface in a search space that defines the fitness for each potential solution. In this case, searching an optimal solution can be interpreted as walking on the fitness landscape surface towards the highest hill, with overcoming other hills and valleys.

It is possible to represent graphically a 3D landscape for one fitness function and a genotype with length two (or two-dimensional search space). However, the fitness landscapes always can be described [2] by a directed graph whose vertices are solutions (genotypes) labelled with fitness values, and arcs are neighbour relations between relevant solutions.

Structural Characteristics and Influencing Factors

The fitness landscape can have different structures on its surface. If an optimisation problem has the fitness function to be maximised, than all optimums are represented as hills. Each hill defines a solution that has higher fitness value than neighbour solutions. The highest hill that has the highest fitness defines a global optimum that is the best solution of the optimisation problem. Others optimums are called as local ones.

Fitness landscapes can also have structures such as swamps, or plateaus that are surface areas which have highly similar fitness values. Areas that have smaller fitness values and divide sets of hills are called as valleys [1, 2].

The fitness landscape has many structural characteristics that describe its structures, such as the modality, the epistasis, the ruggedness, the deception and others [3]. The modality evaluates the number of optima in a search space and an optima density. Problems that have more than one optimal solution are called multimodal. Less optima number defines lower landscape modality. The epistasis refers to genotype fitness dependence on multiple genes interaction. It is stated that a fitness function is non-linear function on distribution of the gene values, and the epistasis defines this nonlinearity [3].
The ruggedness is a fitness landscape structure that influences the problem hardness for an evolutionary algorithm. Higher modality and epistasis lead to a more rugged fitness landscape that is harder to search [4, 5].

Objectives of Analysis

The main objective of a fitness landscape analysis is evaluation of the difficulty of the optimisation problem. A problem is hard to solve with an evolutionary algorithm if its fitness landscape has a large number of structures, which disturb search of the global optimal solution. The main landscape feature which influences problems difficulty for an optimisation algorithm is the fitness landscape ruggedness. Difficulty of a problem is also dependent on its modality and a number of epistasis. However these landscape features do not appear to be sufficient for a problem to be difficult for evolutionary algorithms. These features can be used in order to select the best search algorithm [6].

Review of Fitness Analysis Techniques

A number of different techniques have been developed for fitness landscape analysis by evaluating its structural characteristics. These techniques can be divided in two groups:
- statistical analysis,
- information analysis.

Statistical analysis techniques use collection of the statistical data and evaluate correlation between potential solutions on the fitness landscape. Information analysis techniques use measures from information theory to estimate structural features of a fitness landscape.

Fitness landscape analysis techniques do not require information about all problem solutions, but analyze only a part of a fitness landscape data and apply different strategies for data collection. In this paper, a random walk strategy is used which starts from a random solution on the landscape. Then it collects fitness data sequentially moving from one solution to another with application of the mutation operator.

Fitness analysis techniques described in this paper could be applied for all types of evolutionary algorithms.

Statistical Analysis

Within the statistical analysis approach the correlation is used to measure the ruggedness of the landscape. In case of a high correlation between two sets of solutions, or landscape points, these points have similar fitness values, and the fitness landscape is considered smooth or less rugged. Otherwise the fitnesses of neighbour solutions are very different, the correlation is low and the landscape is rugged.

The autocorrelation function indicates the correlation between two sets of points that are separated by a distance \( \Gamma \) [7]:

\[
\rho(\Gamma) \approx \frac{E(f_i f_{i+\gamma}) - E(f_i)E(f_{i+\gamma})}{V(f_i)},
\]

where \( E(f_i) \) and \( V(f_i) \) are the expectation and the variance, respectively, of a sequence of fitness values, \( \{f_i\}_{i=1}^N \), where \( N \) is the length of the sequence. For smooth landscapes the autocorrelation function is close to 1 and tends to zero for rugged landscapes [3].

The ruggedness of a fitness landscape can be also measured with correlation length. It defines a distance beyond which two sets of fitness points becomes uncorrelated. The correlation length is evaluated [3] by:

\[
\tau = \frac{1}{\ln(\rho(1))},
\]

where \( \rho(1) \) is the autocorrelation of the neighbouring points. The magnitude of correlation length indicates the smoothness of landscape. A longer correlation length will indicates a very smooth landscape while a shorter length will indicate a more rugged landscape. Let note that both autocorrelation and correlation length techniques can be used only for a statistically isotropic fitness landscapes. A landscape is statically isotropic if a fitness value sequence obtained with a random walk forms a stationary random process for the assumed joint distribution of the fitness values [4].

Information Analysis

To characterise the structure of a fitness landscape, information measures are used and defined over the sequence of fitness values which is obtained by a random walk on the landscape. The entropy of the structure of fitness landscape is used as a basic concept to quantify the ruggedness of landscape. Four information measures are proposed in literature [4]. They are:
- information content \( H(\varepsilon) \);
- partial information content \( M(\varepsilon) \);
- information stability \( e^* \);
- density-basin information \( h(\varepsilon) \).

The information content is a measure of entropy in the system. It characterizes the amount of information necessary to reconstruct the system. In case of a high entropy and information content, the landscape has big variety of structures and is more rugged. All information measures are calculated with notice to a calculation accuracy which is defined by parameter \( \varepsilon \).
In this paper, application of the fitness landscape and are not usable in many real life problems [8].

Furthermore, an analysis of a fitness landscape can be used to evaluate the problem complexity and to select the best evolutionary algorithm. However, evolutionary algorithms also can’t be applicable for a particular case. As for the statistical measures, the information measures can be obtained only for fitness landscapes that are statistically isotropic [4].

**Application to Vehicle Scheduling Problem**

Vehicle scheduling problems (VSP) present a class of problems that are aimed at assigning a set of scheduled trips to a set of vehicles, in such a way, that each trip is associated with one vehicle, and a cost function for all trips is minimized [8, 9]. This problem is often modified with additional constraints, like time windows, different truck capacity, etc.

A number of methods to solve VSP problems are proposed in literature, e.g. integer programming, combinatorial methods, heuristics. However, all ready solutions can be applied only for special problem cases and are not usable in many real life problems [8].

In this paper, application of the fitness landscape analysis to complicated vehicle scheduling problems is proposed. These problems do not have efficient optimisation methods and can be solved by using evolutionary algorithms. However, evolutionary algorithms also can’t be applicable for a particular case. Furthermore, an analysis of a fitness landscape can be used to evaluate the problem complexity and to select the best evolutionary algorithm.

In practice, VSP can also be complicated by stochastic processes existing in the system, e.g. when the length of a trip is a random variable. In this case evaluation of potential solutions can be done with simulation, and simulation optimisation could be used to solve such problems. Simulation technology provides a flexible tool to determine the optimality of each solution. Therefore, a simulation-based fitness landscape analysis that supposes fitness evaluation of the solution with use of simulation becomes an important task.

**Prototyping**

Fitness evaluation of the potential solution is made with a help of simulation. The procedure for the simulation-based fitness landscape analysis comprises of three stages (Figure 1):

- Fitness landscape’s path generation;
- Fitness evaluation of solutions in the path;
- Analysis of the path’s fitness sequence.

To generate landscape path solutions and analyse their fitness sequence, Java applications are developed. For fitness evaluation, AnyLogic 6 Advanced simulation software is used. At first stage a standalone application performs a random walk on the problem fitness landscape. As a result, genotypes of landscape path solutions are obtained. As far as AnyLogic software does not have API for external applications, all data is transferred in the form of tables into MS Excel spreadsheets.

![Fig. 1. Main stages of fitness landscape analysis](image-url)

At the second stage, AnyLogic Parameters Variation experiment type is used to run the simulation model with different parameters, i.e. solutions with different genotypes. As a result, the model generates another spreadsheet which contains sequence of fitness values.

Finally at the last stage, statistical and information fitness landscapes analysis is performed on generated fitness path data simultaneously. For example, within statistical fitness landscape analysis autocorrelation function of neighbour solutions is calculated. I addition, correlation length value and autocorrelation function for sets of points separated with 10 solutions are calculated.

Algorithm of information analysis consists of three main steps:

1) Determination of information stability $\varepsilon^*$;
2) Iterative calculation of information content $H(\varepsilon)$, partial information content $M(\varepsilon)$ and density-basin information $h(\varepsilon)$ for different values of $\varepsilon$.
3) Printout of received data.

To define $\varepsilon^*$, an interval of possible $\varepsilon$ values is divided by 2 in each iteration. A half-interval that contains possible value $\varepsilon^*$ is selected for further analysis in the next iteration. As was mentioned above, information stability is equal to the smallest $\varepsilon$ value, where no any slope can be determined. Measures $H(\varepsilon)$, $M(\varepsilon)$ and $h(\varepsilon)$ are calculated iteratively in the interval $[0, \varepsilon^*]$ with a step 0.05.
Case Study

Problem Definition

A case study based on the vehicle scheduling problem with time windows of logistic enterprise is developed for prototype validation.

Enterprise trucks (vehicles) with various parameters have to deliver various types of goods from one distribution centre to many shops all over the country. Distribution routes for vehicles are known. The following parameters are defined for each route: a sequence of shops (route points), average time intervals for a vehicle moving between route points, loading and unloading average time and type of goods that can be carried in this route. However, shops do not accept goods all daytime, but only in predefined time windows. An average demand of goods of each type is defined for each shop. Truck capacities are limited and known.

The problem is aimed to assign trucks to routes in order to minimise the total idle time of all trucks. The truck idle time is defined as sum of time periods, when a truck is waiting for next trip in the parking place.

Problem Formal Description

As far as sequences of shops in routes are fixed, and average time intervals for truck moving between route points are known, the only possibility for optimisation of the vehicle schedule is assignment of trucks to routes, and definition of start time for each route. Thus, the following decision variables are introduced in the problem, correspondingly:

\[ \text{trip } x \_\text{truck } i, \text{trip } x \_\text{time}, \]

where \( x \) is a route number, and \( i \) is a truck number.

The objective function \( f \) is aimed to minimise the total idle time for all trucks, i.e.:

\[ f = \sum_{i=1}^{N} T_{idle}^i \rightarrow \min, \]

where \( T_{idle}^i \) is the total idle time for truck \( i \); and \( N \) is a number of trucks.

The problem constraints are divided in three groups:

1) Truck capacity constraints, which define that no truck can ship more cargo, than its capacity;
2) Delivery time constraints, i.e. goods could be delivered to shops only within defined time windows;
3) Gate capacity constraints, i.e. a number of trucks that can be loaded in a warehouse simultaneously can not exceed number of gates.

Experimental analysis of formulated problem shows that the problem could have many solutions, which could be not feasible with restrictions. This makes computational process non-efficient. To avoid this situation, the objective function (3) was modified taking into account an amount of constraints that were not satisfied by potential solutions:

\[ f^* = \sum T_{idle} + k_1 T_c + k_2 T_m + k_3 T_0 + k_4 N_{ol} + k_5 N_{ot}, \]

where \( f^* \) is modified objective function; \( T_c \) defines the total duration of overlapping trips for one vehicle; \( T_m \) defines the total time of window mismatches; \( T_0 \) and \( N_{ol} \) determine the total time and a number of vehicles that have overdone 24 working hours; and \( N_{ot} \) is a number of vehicles that are overloaded. In (4), all indexes for unsatisfied constraints are multiplied with coefficients \( k_i > 1, i = 1,...,5 \) that artificially increase a value of objective function and make fitness of solutions worse.

Simulation with AnyLogic

To determine fitness of potential vehicle schedule solutions discrete event simulation model in AnyLogic is developed, in which each vehicle is simulated as a particular active object. Specific processes related to the distribution centre operations are simulated. Vehicle simulation state chart is given in Figure 2, which defines vehicle’s possible states (e.g. parking, moving, unloading) and transitions between them.

During simulation process constraint violations such as time window mismatch in delivery (wrong delivery time), shortage of truck and gate capacity, are determined.

Optimisation Scenarios

As inputs 37 routes and corresponding 17 trucks, and 36 shops are selected. Specific parameters of trucks, shops and routes are defined in the model. Input data are interpreted as deterministic and then as stochastic, as loading, transportation and unloading times are...
random in practice. As simulation output the total idle
time for all trucks is calculated.

A number of decision variables, that define a vehicle
schedule is equal to 74. Thus, exploring of the decision
search space requires evaluation of thousands possible
solutions. Finally, function (4) is used for fitness
evaluation of the solutions simulated.

To solve the problem, two optimisation scenarios are
analysed.

Scenario 1. Problem solution with OptQuest
As simulation model was developed in AnyLogic
software, the powerful optimisation software add-on
OptQuest [10] included in the commercial product was
used in the first scenario. However, the optimization
problem has been solved only for a simplified
scheduling problem (Figure 3).

![Fig. 3. Optimisation progress for simplified problem](image)

As an initial solution in a search process a current
vehicle schedule of the distribution centre was defined
(Figure 3). The best solution that has been found in this
scenario allowed decreasing a number of routes in a
schedule (Figure 4).

![Fig. 4. Optimisation progress for simplified problem](image)

Another alternative for problem solving is use of
external optimization tool. However, an external
optimisation tool connection to AnyLogic is not a
simple task, so first it is important to check if it
provides efficient techniques to solve this problem.

Scenario 2. Problem solution based on fitness
landscape analysis
Various series of experiments of fitness landscape
analysis for the problem formulated were performed
with developed software prototype. In each series, 5
experiments with landscape’s 100 solutions’ long path
were made. In the first series of experiments, the results
of fitness landscape analysis for the problem with
stochastic input data (stochastic times for moving,
unloading, etc.) and deterministic data expressed by
average values were compared.

Information measures of the problem fitness
landscape received from simulation experiments are
given in Table 1, and statistical ones in Table 2.

<table>
<thead>
<tr>
<th>Input data</th>
<th>$H(0.1)$</th>
<th>$M(0.1)$</th>
<th>$h(0.1)$</th>
<th>$\varepsilon^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic</td>
<td>0.66</td>
<td>0.20</td>
<td>0.49</td>
<td>0.40</td>
</tr>
<tr>
<td>Deterministic</td>
<td>0.60</td>
<td>0.17</td>
<td>0.37</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Information content $H(0.1)$ of analysed fitness
landscape is relatively high (i.e. is more than 0.5). This
means that fitness landscape of the optimisation
problem is relatively rugged and even more rugged for
the problem with stochastic data. Then partial data
content $M(0.1)$ is low, and as a result, modality of
fitness landscape is low. Also modality is lower for the
problem with determined data. Density-basin
information $h(0.1)$ indicates that peaks have high
density and their density is higher for a stochastic
problem. Information stability value is bigger for a
stochastic problem, which means that peaks are higher
for stochastic problem.

<table>
<thead>
<tr>
<th>Input data</th>
<th>$\rho(1)$</th>
<th>$\rho(10)$</th>
<th>$\tau(0.1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic</td>
<td>0.84</td>
<td>0.21</td>
<td>7.24</td>
</tr>
<tr>
<td>Deterministic</td>
<td>0.89</td>
<td>0.32</td>
<td>8.75</td>
</tr>
</tbody>
</table>

Statistical measures of fitness landscape also lead to
the conclusion that landscape is more rugged for
problem with stochastic data as autocorrelation value
$\rho(1)$ is smaller. But opposite to information measures,
statistical measures indicate that landscape is relatively
smooth. Correlation length is almost 8 solutions.

In the second series of experiments, dependence of
informational measures from calculation accuracy $\varepsilon$
was analysed. Dependence of information content partial
information content from $\varepsilon$ is shown in Figure 5, 6,
correspondingly.

![Fig. 5. Dependence of information content from accuracy $\varepsilon$](image)

Let note, that information content is highest not for
small $\varepsilon$, when every little slope is taken into account,
but for high $\varepsilon$ values, when part of landscape is smoothed. Partial information content value decreases if when $\varepsilon$ is increased (see, Figure 6).

![Figure 6. Dependence of partial information content from accuracy $\varepsilon$](image)

The results of fitness landscape analysis in scenario 2 lead to the conclusion that the application problem is not hard for evolutionary algorithms.

**Conclusions**

Analysis of fitness landscape allows evaluating hardness of the optimisation problem for evolutionary algorithms. For complex problems fitness landscape evaluation requires application of simulation. The results of fitness landscape analysis show that evolutionary algorithms provide an efficient tool for solving vehicle scheduling problems with time windows. Future research will be devoted to designing simulation optimisation experiments to solve the problem formulated with evolutionary algorithms.

**References**