TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Informatik Institut für Software- und Multimediatechnik, Lehrstuhl für Softwaretechnologie

From Parameter Tuning to Dynamic Heuristic Selection

Yevhenii Semendiak

Yevhenii.Semendiak@tu-dresden.de
Born on: 7th February 1995 in Izyaslav
Course: Distributed Systems Engineering
Matriculation number: 4733680
Matriculation year: 2020

Master Thesis

to achieve the academic degree

Master of Science (M.Sc.)

Supervisors
MSc. Dmytro Pukhkaiev
Dr. Sebastian Götz

Supervising professor
Prof. Dr. rer. nat habil. Uwe Aßmann

Submitted on: 5th March 2020

Technische Universität Dresden/Fakultät Informatik

Aufgabenstellung für die Masterarbeit

Name, Vorname: Semendiak, Yevhenii

Studiengang: Master DSE Matr. Nr.: 4 7 3 3 6 8 0

Thema:

From Parameter Tuning to Dynamic Heuristic Selection

Zielstellung :

Metaheuristic-based solvers are widely used in solving combinatorial optimization problems. A choice of an underlying metaheuristic is crucial to achieve high quality of the solution and performance. A combination of several metaheuristics in a single hybrid heuristic proved to be a successful design decision. State-of-the-art hybridization approaches consider it as a design time problem, whilst leaving a choice of an optimal heuristics combination and its parameter settings to parameter tuning approaches. The goal of this thesis is to extend a software product line for parameter tuning with dynamic heuristic selection; thus, allowing to adapt heuristics at runtime. The research objective is to investigate whether dynamic selection of an optimization heuristic can positively effect performance and scalability of a metaheuristic-based solver.

For this thesis, the following tasks have to be fulfilled:

- Literature analysis covering closely related work.
- Development of a strategy for online heuristic selection.
- Implementation of the developed strategy.
- Evaluation of the developed approach based on a synthetic benchmark.
- (Optional) Evaluation of the developed approach with a problem of software variant selection and hardware resource allocation.

Betreuer: M.Sc. Dmytro Pukhkaiev, Dr.-Ing. Sebastian Götz

Verantwortlicher Hochschullehrer: Prof. Dr. rer. nat. habil. Uwe Aßmann

Institut: Software- und Multimediatechnik

Beginn am : 01.10.2019

Einzureichen am : 09.03.2020

Unterschrift des verantwortlichen Hochschullehrers

Verteiler: 1 x SCIS, 1x HSL, 1x Betreuer, 1x Student

Contents

List of Figures

List of Tables

0.1 abstract

Abstract will be available in final versions of thesis.

1 Introduction

Intent and content of chapter. This chapter is an self-descriptive, shorten version of thesis.

1.1 Motivation

Structure:

- optimization problem(OP) → exact or approximate (+description to both) → motivation to use **approximate solvers** →
- impact of parameters, their tuning on solvers → motivation of **parameter control** (for on-line solver) →
- but what if we want to solve a class of problems (CoP) → algorithms performance is different →
- user could not determine it [12] → exploration-exploitation balance
- no-free-lunch (NFL) theorem [18] → motivation of the thesis

thesis motivation The most related research field is Hyper-heuristics optimizations [4], that are designed to intelligently choose the right low lewel heuristics (LLH) while solving the problem. But the weak side of hyper-heuristics is the luck of parameter tuning of those LLHs [links]. In the other hand, meta-heuristics often utilize parameter control approaches [links], but they do not select among underlying LLHs. The goal of this thesis is to get the best of both worlds - algorithm selection from the hyper-heuristics and parameter control from the meta-heuristics.

1.2 Research objective

Yevhenii: Rename: Problem definition?

The following steps should be completed in order to reach the desired goal:

Analysis of existing studies of algorithm selection. *(find a problem definition, maybe this will do [12])*

Analysis of existing studies in field of parameter control and algorithm configuration problems *(find a problem definition)* [13]

Formulation and development of combined approach for LLH selection and parameter control.

Evaluation of the developed approach with

Yevhenii: family of problems??? since it is a HH, maybe we should think about it...

.

Research Questions At this point we define a Research Questions (RQ) of the Master thesis.

- RQ 1 Is it possible to select an algorithm and it hyper-parameters while solving an optimization problem *on-line*?
- RQ 2 What is the gain of selecting and tuning algorithm while solving an optimization problem?
- RQ 3? How to solve the problem of algorithm selection and configuration simultaneously?

1.3 Solution overview

Yevhenii: Rename: Problem solution?

- described problems solved by HH, highlight problems of existing HHs(off-line, solving a set of homogeneous problems in parallel)
- create / find portfolio of MHs (Low level Heuristics)
- define a search space as combination of LLH and their hyper-parameters (highlight as a contribution)
- solve a problem on-line selecting LLH and tuning hyper-parameters on the fly. (highlight as a contribution? need to analyze it.)

Thesis structure The description of this thesis is organized as follows. First, in chapter 2 we refresh readers background knowledge in the field of problem solving and heuristics. In this chapter we also define the scope of thesis. Afterwards, in chapter 2 we describe the related work and existing systems in defined scope. In Chapter 4 one will find the concept description of dynamic heuristics selection. Chapter 5 contains more detailed information about approach implementation and embedding it to BRISE. The evaluation results and analysis could be found in Chapter 6. Finally, Chapter 7 concludes the thesis and Chapter 8 describe the future work.

2 Background and Related Work Analysis

In this chapter we provide reader with the base knowledge in field of Optimization Problems and the process of their solving. The reader who is an expert in field of Optimization and Search Problems could find this chapter as an obvious discussion of well-known facts. If the notions of *Parameter Tuning* and *Parameter Control* seems like two different names for one thing, we encourage you to read this chapter carefully. We highly recommend for everyone to refresh the knowledge of sections topics and examine the examples of Hyper-Heuristics in 2.2.4 and Systems for parameter tuning in 2.3.3 since we use them later in concept implementation.

In this chapter we...

2.1 Optimization, Search Problems and their Solvers

2.1.1 Definition of the Optimization Problem

While the Search Problem (SP) defines the process of finding a possible Solution for the Computation Problem, an Optimization Problem (OP) is the special case of the SP, focused on the process of finding the "best possible" Solution for Computation Problem [10].

In this thesis we focus on the Optimization Problems – a special case of the Search Problems.

A lot of conducted studies in this field have tried to formalize the concept of OP, but the underlying notion such a vast that it is almost impossible to exclude the application domain from the definition. However, in [1, 3, 9] authors distinguished OP characteristics that overlap through each of these works and we would like to mention them here.

First of all, let us define the subject of the Optimization. In general, it could be imagined as the Target System (TS) displayed on picture 2.1. Analytically it could be represented as the function $Y = f(X)$. Informally it accepts the information with it's *inputs* $\mathbf{X}$ sometimes also called variables or parameters, performs a *Task* and produces the result on it's *outputs* $\mathbf{Y}$.

Pair of X and respective Y form a Solution for Computation Problem. The Solution could also be characterized by the *objective* value(s) - a quantitative measure of TS

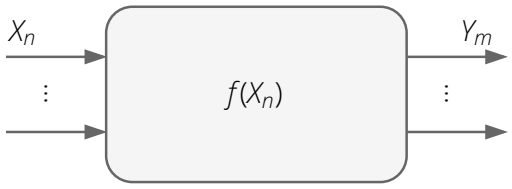

Figure 2.1 Target System

performance that we minimize or maximize. We could obtain those value(s) directly by reading the *Y*, or indirectly for instance, noting the time TS took to produce the output *Y* for given *X*. The Solution objective value(s) form object(s) of Optimization. For the sake of simplicity we here use *Y*, *outputs*, *objectives* and *X*, variables, **parameters(?)**interchangeably.

Next, let us highlight the Target System characteristics. Among mentioned in [1, 3, 9] we found those the most important:

- **Input data types** of *X* is a crucial characteristic. The variables could be either *discrete* where representatives are binary strings, integer-ordered or categorical data, *continuous* where variables are usually a range of real numbers, or *mixed* as the mixture of previous two cases.
- **Constrains** are functional dependencies that describe the relationships among inputs and defile the allowable values for them.
- **Amount of knowledge** TS exposes about the dependencies between $X \rightarrow Y$ or objective values. With respect to this knowledge, the Optimization could be *White Box* - the TS exposes it internals fully, so it is even possible to derive the algebraic model of TS. *Black Box* - the exposed knowledge is mostly negligible. As an example, imagine the
- **Dependencies randomness** One of possible challenges, while obtaining the knowledge about TS is uncertainty of output. Ideal case is the *deterministic* dependency between *X* and *Y*, however in most of real-world challenges engineers tackle with the *stochastic* systems whose output is affected by random processes.
- **Cost of evaluation** is the amount of resources (computational, time, money, etc.) TS will spend to obtain the result for particular input. It vary from very cheap if the TS is a simple algebraic formula and Task is to evaluate it, to very expensive if the TS is a complex Neuron Network and the Task is to train it on data.
- **Number of objectives** could be either *Single*, or *Multiple*. According to the number of objectives, the result of optimization will be either single Solution, or set of non-dominated (Pareto-optimal) Solutions [6].

2.1.2 Classes of Solvers

Some (but not only) literature: [2]

The characteristics of Target System could restrict and sometimes strictly define the possible approach to perform an Optimization. As instance, for Black Box TS one will not be able to apply any Mixed Integer Programming Optimization approaches, however Heuristics still will work. In opposite case, if the TS exposes algebraic dependencies between input variables and output objective, the MIP optimization, in particular MILP

or MINLP approaches could be used to derive mathematically proved optimal Solution [3].

Exact Solvers

Approximate Solvers

Motivation of Approximate Solvers

Pros and cons of both [11]

2.2 Approximate Solvers for Optimization Problems

TSP as the running example. I guess, I will introduce it as an example of perturbation problems in previous section.??

2.2.1 Heuristics

Definition

Examples

2.2.2 Meta-Heuristics

Definition

Classification

Examples

We distinguish following examples among all existing meta-heuristics, since later we use them as the LLH in developed hyper-heuristic.

GA

SA

ES

2.2.3 Hybrid-Heuristics

Definition

Examples

Guided Loca Search (GLS) + Fast Local Search [17]

Direct Global + Local search [16]

Simulated Annealing + Local Search [15]

No-Free-Lunch Theorem

NFL is the problem of heuristics[18]

Exploration-Exploitation Balance

Conclusion

Proper assignment of hyper-parameters has great impact on exploration-exploitation balance and those on (meta) -heuristic performance.

2.2.4 Hyper-Heuristics

Definition

Classification

Search Space: heuristic selection, heuristic generation

Learning time: on-line learning hyper-heuristics, off-line learning hyper-heuristics, no-learning hyper-heuristics

Other classification characteristics from [12], [5], mb smth else. For instance, hyperparameter tuning

Examples

[7] (Online algorithm selection at page 27); [12]

2.2.5 Conclusion on Approximate Solvers

Pros and cons of heuristics - Heuristics are strictly problem dependent and each time require adaptations.

Pros and cons of meta-heuristics - no LLH selection, strict to one problem

Pros and cons of hybrid-heuristics - no LLH selection, strict to one problem ?

Pros and cons of hyper-heuristics - no parameter control?

2.3 Parameter Tuning as a Search Problem

The goal of section: analysis of existing systems for hyper-parameter optimization (tuning), weaknesses and strength of each of the system

2.3.1 Parameter Tuning Problem Definition

2.3.2 Approaches for Parameter Tuning

Grid Search

Random Search

Model Based Search

2.3.3 Systems for Model Based Parameter Tuning

IRACE

approach [14]

pros and cons

SMAC

approach description

BOHB

approach description

AUTO-SKLEARN

CASH (Combined Algorithm Selection and Hyperparameter optimization) problem

pros and cons (on-line or off-line, problems to solve, extensibility) [8]

BRISEv2

approach description

Yevhenii: Other systems?

2.4 Parameter control as an Optimization Problem

2.4.1 Parameter Control Definition

2.4.2 Examples and Reported Impact

impact of parameter control based on other's evaluation

2.5 Conclusion

The meta-heuristic systems designers reported positive impact of parameter control embedding. However, as the outcome of the no-free-lunch theorem, those systems can not tolerate broad range of problems, for instance, problem classes. In other hand, hyper-heuristics are designed with an aim to select the low level heuristics and those propose a possible solution of problem, stated in no-free-lunch theorem, but the lack of parameter control could dramatically decrease the performance of LLH (probably, I need to find a prove of this, or rephrase).

Scope of thesis defined. In this thesis we try to achieve the best of both worlds applying the best fitting LLH and tuning it's parameters while solving the problem on-line.

3 Concept Description

In this chapter we describe the concept of developed selection Hyper-Heuristic with parameter control, not diving deep into the implementation details.

The structure of this chapter is as follows.

Yevhenii: maybe we should not highlight the structure of such a small chapter

First, in in section 3.1 we define the Search Space entity requirements and structure. It should bound the world of Low Level Heuristics and the world of Hyper-parameters of those heuristics.

Next, we describe the Prediction process within the previously defined Search Space in section 3.2. Here we highlight an importance of a prediction model decoupling from the previously defined Search Space structure. Doing so, we provide certain level of flexibility for user in the usage of different prediction models or developing his own.

Finally, in section 3.3 we gather our attention onto the Low Level Heuristics - a working horse of the hyper-heuristic. Here we highlight the requirements for LLH in terms of features that will be used by HH.

3.1 Search Space

Importance explanation

Required structure feature-tree structured

3.2 Prediction Process

Importance explanation

Requirements generality, top-down approach of optimization – different views of same Configuration (level-dependent) - filtering, transformation – consider problem features? while selecting meta-heuristic [12] page 6

3.3 Low Level Heuristics

Importance explanation

3.4 Conclusion of concept

to be done...

4 Implementation Details

In this chapter we dive into the implementation details of the selection hyper-heuristic with parameter control.

The best practice in software engineering is to minimize an effort for the implementation and reuse already existing and well-tested code. With this idea in mind we had decided to reuse one of existing (and highlighted by us in 2.3) open-source hyper-parameter tuning systems as the code basis and those turn it into the core of hyper-heuristic. Do to so we analyze the existing systems and highlight important non-functional characteristics from the implementation perspective in section 4.1. Since the selected code base system is not the ideal in terms of such features as Search Space entity abilities and the prediction process, we consider some adaptations in sections 4.2 and 4.3 respectively. We also reuse the set of Low Level Heuristics in section 4.5.2.

4.1 Hyper-Heuristics Code Base Selection

A.k.a. "brain". Need to find a better way to call this part of HH..

4.1.1 Requirements

4.1.2 Parameter Tuning Frameworks

SMAC

BOHB

IRACE

BRISEv2

Yevhenii: Maybe, smth else..

4.1.3 Conclusion

BRISEv2 is the best system for code basis, however it has to be changed as we describe in following sections.

4.2 Search Space

4.2.1 Base Version Description

What is the problem with the current Search Space?

The Scope Refinement Work Throw away and write a new one :D

4.2.2 Implementation

Description

Motivation of structure

Class diagram - i think, I will put it into the appendix

4.3 Prediction Logic

4.3.1 Base Version Description

The Scope Refinement Work prediction should be done in feature-tree structured search space. Most models could handle only flat search space and we would like to enable reuse of those existing models. Though we decouple the structure of Search Space in entity Predictor, while actual prediction process is done in underlying models, that Predictor uses.

4.3.2 Predictor

to decouple prediction from structure of search space.

4.3.3 Prediction Models

Tree parzen estimator

Multi Armed Bandit

Sklearn linear regression wrapper

4.4 Data Preprocessing

4.4.1 Heterogeneous Data

description and motivation of data preprocessing notions

4.4.2 Base Version Description

and Scope of work analysis

4.4.3 Wrapper for Scikit-learn Preprocessors

4.5 Low Level Heuristics

4.5.1 Requirements

4.5.2 Code Base Selection

Available Meta-heuristics with description of their current state With the aim of effort reuse, the code base should be selected for implementation of the designed hyper-heuristic approach.

SOLID

MLRose

OR-tools

pyTSP

LocalSolver

jMetalPy

4.5.3 Scope of work analysis

opened PR

4.6 Conclusion

5 Evaluation

5.1 Evaluation Plan

5.1.1 Optimization Problems Definition

Traveling Salesman Problem

tsplib95 benchmark set which problem I want to solve with hyper-heuristic

N-Queens Problem?

Knapsack Problem?

5.1.2 Hyper-Heuristic Settings

To evaluate the performance of developed system we first need to compare it with the base line. In our case it is the simple meta-heuristic that is solving the problem with static hyper-parameters.

In order to organize the evaluation plan, we distinguish two stages of setup, where different approaches could be applied. At the first stage we select Low Level Heuristic, while at the second one we select hyper-parameters for LLH. The approaches for each step are represented in table 5.1.

Table 5.1 System settings for benchmark

Low Level Heuristics selection	LLH Hyper-parameters selection
1. Random 2. Multi Armed Bandit 3. Sklearn Bayesian Optimization 4. Static selection of SA, GA, ES	1. Default 2. Tuned beforehand 3. Random 4. Tree Parzen Estimator 5. Sklearn Bayesian Optimization

For instance, mentioned above baseline could be described as *Settings*4.1. for meta-heuristics with default hyper-parameters and as *Settings*4.2. for meta-heuristics with tuned beforehand hyper-parameters.

For our benchmark we selected following settings sets:

- *Baseline:* 4.1, 4.2;

- *Random Hyper-heuristic:* 1.1, 1.2, 1.3, 4.3;
- *Parameter control:* 4.4, 4.5;
- *Selection Hyper-Heuristic:* 2.1, 3.1, 2.2, 3.2;
- *Selection Hyper-Heuristic with Parameter Control:* 2.4, 2.5, 3.4, 3.5;

Each of this settings will be discussed in details in following section.

5.1.3 Selected for Evaluation Hyper-Heuristic Settings

Baseline

Hyper-heuristic With Random Switching of Low Level Heuristics

Parameter control

Selection Only Hyper-Heuristic

Selection Hyper-Heuristic with Parameter Control

5.2 Results Discussion

5.2.1 Baseline Evaluation

Meta-Heuristics With Default Hyper-Parameters

Meta-Heuristics With Tuned Hyper-Parameters

Results Description and Explanation

5.2.2 Hyper-Heuristic With Random Switching of Low Level Heuristics

Results Description and Explanation

5.2.3 Parameter Control

Results Description and Explanation

5.2.4 Selection Only Hyper-Heuristic

Results Description and Explanation

5.2.5 Selection Hyper-Heuristic with Parameter Control

Results Description and Explanation

5.3 Conclusion

6 Conclusion

Reviewer: answer research questions

7 Future work

add more sophisticated models

dependencies / constraints in search space

add new class of problem (jmetalpy easly allows it)

evaluation on different types and classes

Reviewer: consider merging with conclusion, if too short

Bibliography

[1] Satyajith Amaran et al. “Simulation optimization: a review of algorithms and applications”. In: *Annals of Operations Research* 240.1 (2016), pp. 351–380.

[2] James S Bergstra et al. “Algorithms for hyper-parameter optimization”. In: *Advances in neural information processing systems*. 2011, pp. 2546–2554.

[3] Lorenz T Biegler and Ignacio E Grossmann. “Retrospective on optimization”. In: *Computers & Chemical Engineering* 28.8 (2004), pp. 1169–1192.

[4] Edmund Burke et al. “Hyper-heuristics: An emerging direction in modern search technology”. In: *Handbook of metaheuristics*. Springer, 2003, pp. 457–474.

[5] Edmund K Burke et al. “A classification of hyper-heuristic approaches: revisited”. In: *Handbook of Metaheuristics*. Springer, 2019, pp. 453–477.

[6] Kalyanmoy Deb. “Multi-objective optimization”. In: *Search methodologies*. Springer, 2014, pp. 403–449.

[7] John H Drake et al. “Recent advances in selection hyper-heuristics”. In: *European Journal of Operational Research* (2019).

[8] Matthias Feurer et al. “Efficient and robust automated machine learning”. In: *Advances in neural information processing systems*. 2015, pp. 2962–2970.

[9] Goncalo Figueira and Bernardo Almada-Lobo. “Hybrid simulation–optimization methods: A taxonomy and discussion”. In: *Simulation Modelling Practice and Theory* 46 (Aug. 2014).

[10] Oded Goldreich. *P, NP, and NP-Completeness: The basics of computational complexity*. Cambridge University Press, 2010.

[11] Juraj Hromkovič. *Algorithmics for hard problems: introduction to combinatorial optimization, randomization, approximation, and heuristics*. Springer Science & Business Media, 2013.

[12] Pascal Kerschke et al. “Automated algorithm selection: Survey and perspectives”. In: *Evolutionary computation* 27.1 (2019), pp. 3–45.

[13] Niklas Lavesson and Paul Davidsson. “Quantifying the impact of learning algorithm parameter tuning”. In: *AAAI*. Vol. 6. 2006, pp. 395–400.

[14] Manuel López-Ibáñez et al. “The irace package: Iterated racing for automatic algorithm configuration”. In: *Operations Research Perspectives* 3 (2016), pp. 43–58.

[15] Olivier C Martin and Steve W Otto. “Combining simulated annealing with local search heuristics”. In: *Annals of Operations Research* 63.1 (1996), pp. 57–75.

[16] Michael Syrjakow and Helena Szczerbicka. “Efficient parameter optimization based on combination of direct global and local search methods”. In: *Evolutionary Algorithms*. Springer. 1999, pp. 227–249.

[17] Edward Tsang and Chris Voudouris. “Fast local search and guided local search and their application to British Telecom’s workforce scheduling problem”. In: *Operations Research Letters* 20.3 (1997), pp. 119–127.

[18] David H Wolpert and William G Macready. “No free lunch theorems for optimization”. In: *IEEE transactions on evolutionary computation* 1.1 (1997), pp. 67–82.

Statement of authorship

I hereby certify that I have authored this Master Thesis entitled *From Parameter Tuning to Dynamic Heuristic Selection* independently and without undue assistance from third parties. No other than the resources and references indicated in this thesis have been used. I have marked both literal and accordingly adopted quotations as such. There were no additional persons involved in the intellectual preparation of the present thesis. I am aware that violations of this declaration may lead to subsequent withdrawal of the degree.

Dresden, 5th March 2020

Yevhenii Semendiak