Roman Václavík

Algorithms for Personnel Scheduling Enhanced by Machine Learning Techniques

August 2018

CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Electrical Engineering
Department of Control Engineering

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Control Engineering

Algorithms for Personnel Scheduling Enhanced by Machine Learning Techniques

Doctoral Thesis

Roman Václavík

Prague, August 2018

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Control Engineering and Robotics

Supervisor: doc. Ing. Přemysl Šůcha, Ph.D.
Supervisor specialists: prof. Dr. Ing. Zdeněk Hanzálek

Roman Václavík: *Algorithms for Personnel Scheduling Enhanced by Machine Learning Techniques*, Ph.D. Thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Control Engineering © Prague, August 2018

Declaration

This doctoral thesis is submitted in partial fulfillment of the requirements for the degree of doctor (Ph.D.). The work submitted in this dissertation is the result of my own investigation, except where otherwise stated. I declare that I worked out this thesis independently and I quoted all used sources of information in accord with Methodical instructions about ethical principles for writing academic thesis. Moreover I declare that it has not already been accepted for any degree and is also not being concurrently submitted for any other degree.

Czech Technical University in Prague
Prague, August 2018 *Roman Václavík*

Acknowledgments

I would like to acknowledge my thesis supervisors doc. Ing. Přemysl Šůcha, Ph.D. and prof. Dr. Ing. Zdeněk Hanzálek for their good advices during my Ph.D. studies and their trust in my research topic.

Huge thanks belong to my colleagues Ing. Antonín Novák who designed and implemented the first model of the branch-and-price algorithm for the nurse rostering problem in his master thesis and to Ing. István Módos who worked on a system for adaptive online scheduling of tasks in our team at DEMANES project.

I have to emphasize the perfect working environment in our research group at the Department of Control Engineering. This could not be possible without the friendly colleagues and the excellent research group leadership by prof. Dr. Ing. Zdeněk Hanzálek, doc. Ing. Přemysl Šůcha, Ph.D. and Ing. Michal Sojka, Ph.D.

I would like to express my deepest gratitude to all my family members who supported me not only during the Ph.D. studies but continuously during my whole life. Without them, I would not be where I am nowadays.

This work was supported by ARTEMIS FP7 EU and by the Ministry of Education of the Czech Republic under the project DEMANES 295372, by the Grant Agency of the Czech Republic under the Project GACR FOREST P103-16-23509S and by the Technology Agency of the Czech Republic under the Centre for Applied Cybernetics TE01020197.

Czech Technical University in Prague
Prague, August 2018 *Roman Václavík*

Abstract

Many algorithms solving combinatorial problems can be found in the literature since researchers deal with these problems for many decades. The algorithms are mostly designed following a model-based approach. This means that the focus is on a rigorous description of an algorithm based on a deep understanding of the considered problem. However, this approach does not take into account work with data (i.e., processing of data in order to let data drive the algorithm) even though it generates data flows. This can be seen as a flaw since a lot of arbitrary already calculated data is present inside the algorithm iterations, but it is not used anymore in the future steps of algorithm generally.

In this thesis, our goal is to overcome this flaw and, thus, to improve the classical algorithms regarding speed and robustness. To do that, we introduce a data-driven approach which is combined with the standard model-based approach to get advantages of both approaches and also to negate their disadvantages.

The nurse rostering problem was chosen as a principal problem for the thesis. It is a very-well known NP-hard combinatorial problem assigning shifts to nurses in order to create a roster concerning many various constraints. High-quality rosters then have a positive impact on the entire hospital including its patients.

Three different approaches for the nurse rostering problem are proposed to demonstrate the feasibility of our ideas. First, a neural network classifier was designed to evaluate the solution changes made inside of a tabu search algorithm. Second, a regression model was used to obtain a tight upper bound for a pricing problem in the branch-and-price algorithm. Third, an estimator of algorithm runtime was developed to improve the behavior of a complex client-server scheduling system.

All proposed approaches were verified on real-world benchmark instances. The experiments show that each approach achieved significant improvements. Moreover, the neural network classifier is robust enough to handle small changes in an input instance. Furthermore, the regression model was designed as a problem-independent, i.e., it can be used in other domains without any major modifications. Finally, the scheduling system with the estimator almost equaled the average quality achieved by the scheduling system with the exact knowledge about the instance runtime for some cases.

Keywords: nurse rostering problem, data-driven design, machine learning, neural network, branch-and-price, pricing problem, regression

Abstrakt

V literatuře lze nalézt mnoho algoritmů řešících kombinatorické problémy, jelikož vědci se touto problematikou zabývají po mnoho desetiletí. Algoritmy jsou většinou navrženy pomocí modelového přístupu. To znamená, že důraz je kladen na přesný popis algoritmu založeného na detailním pochopení uvažovaného problému. Tento přístup však nezohledňuje práci s daty (tzn. zpracování dat pro pozdější použití v algoritmu), přestože generuje datové toky. Tato vlastnost může být chápana jako nedostatek, jelikož v jednotlivých iteracích algoritmu je přítomna spousta různých již vypočtených dat, které se ale v dalších krocích algoritmu obecně nevyužívají.

V této dizertační práci je naším cílem překonat tento nedostatek, a tím zlepšit klasické algoritmy z hlediska rychlosti a robustnosti. Za tímto účelem přicházíme s přístupem založeným na datech, který je kombinován se standardním přístupem založeným na modelu, aby získal výhody obou přístupů a zároveň potlačil jejich nevýhody.

Problém rozvrhování zdravotních sester byl zvolen jako nosný problém této práce. Jedná se o velmi dobře známý NP-obtížný kombinatorický problém, který přiřazuje směny sestrám tak, aby vytvořil rozvrh s ohledem na mnoho různých omezení. Kvalitní rozvrhy pak mají pozitivní dopad na celou nemocnici včetně jejich pacientů.

K prokázání proveditelnosti našich nápadů byly navrženy tři různé metody pro problém rozvrhování zdravotních sester. Zaprvé, klasifikátor z neuronových sítí byl navržen tak, aby vyhodnotil změny řešení provedené uvnitř tabu search algoritmu. Zadruhé, regresní model byl použit k získání těsného horního omezení pro pricing problém v branch-and-price algoritmu. Zatřetí, nástroj na odhadování běhu algoritmu pro danou instanci byl vyvinut pro zlepšení chování komplexního rozvrhovacího systému na bázi klient-server.

Všechny navrhované přístupy byly ověřeny na instancích reálného světa. Z experimentů vyplývá, že každý přístup dosáhl významných zlepšení. Klasifikátor z neuronové sítě je navíc dostatečně robustní, aby zvládl malé změny ve vstupní instanci. Navíc byl regresní model navržen jako nezávislý na problému, tj. může být použit v jiných doménách bez významnějších modifikací. Nakonec, rozvrhovací systém s nástrojem na odhadování téměř vyrovnal průměrnou kvalitu dosaženou rozvrhovacím systémem s přesnou znalostí běhu instance pro některé případy.

Klíčová slova: problém rozvrhování zdravotních sester, daty řízený návrh, strojové učení, neuronová síť, branch-and-price, pricing problém, regrese

Contents

Thesis Format

The thesis is presented in the form of a set of publications and accepted manuscripts provided with linking text. This format of the thesis is approved at the Faculty of Electrical Engineering of the Czech Technical University in Prague by the directive of the dean from 13.12.2017.

List of Abbreviations

BaP	branch-and-price
CG	column generation
CPU	central processing unit
EDD	earliest due date
ILP	integer linear programming
LP	linear programming
MCT	minimal completion time
ML	machine learning
MP	master problem
MSE	mean-squared error
NRP	nurse rostering problem
OLSM	ordinary least squares method
RCSP	resource constrained shortest path
RF	random forest for regression
RM	regression model
RMP	restricted master problem
SVMs	support vector machines
SVR	support vector machines for regression
TDM	time division multiplexing
TS	tabu search
TSA	tabu search algorithm

Goals and Objectives

This thesis is focused on the domain of the personnel scheduling enhanced by machine learning techniques. Its goals were set as follows:

1. Study the state-of-the-art related to a solving of NP-hard combinatorial problems and identify possible improvements.
2. Design a machine learning based approach to accelerate the evaluation of solution changes made by a local operator in a heuristic for the nurse rostering problem.
3. Analyze and improve a branch-and-price algorithm for the nurse rostering problem by processing the data which occurs and is calculated during the solving process.
4. Propose a method estimating an algorithm runtime of the nurse rostering problem instances.
5. Verify the proposed methods on real-world instances and compare them with the state-of-the-art.

Chapter 1

Introduction

Combinatorial problems are challenging for researchers for many decades. In the literature, one can find a wide spectrum of approaches for solving these problems from heuristic ones through the meta-/hyper- heuristics to the exact ones. There are also general algorithms dealing with different problems without special fine-tuning and tailor-made algorithms which are useable just for the given problem or only for some instances of the problem. All across these approaches, a general scheme (Figure 1.1) can be observed. The input is a problem instance which goes through the main flow of the algorithm in order to obtain a (sub-)optimal solution at the end. Inside the main flow, a **time-consuming** set of actions (i.e., iterations) is performed which is **repeated** until a termination condition (e.g., a maximal number of iterations was achieved, or an optimality condition was met) is not satisfied.

The iterations are a core of the algorithm since they try to find the best solution to the problem instance under the given circumstances. In general, one iteration consists of two main parts (Figure 1.1):

1. Modifying the solution - some algorithm specific operations are performed based on the given parameters to make a change in the solution which should lead to the better result either immediately or in the future steps.

2. Evaluating solutions - newly created solutions/partial solutions have to be evaluated by so-called cost function in order to get the comparison with other solutions and to decide whether the created solution should be considered in the future algorithm steps or not.

Note that an **intermediate data** is arising in the processing of iterations. This data can be arbitrary already calculated data which depends on the used actions in the iteration. For example, in a tabu search algorithm (Glover and Laguna (1998)) the intermediate data can contain successful and unsuccessful changes applied to the solution. In the branch-and-price (BaP) al-

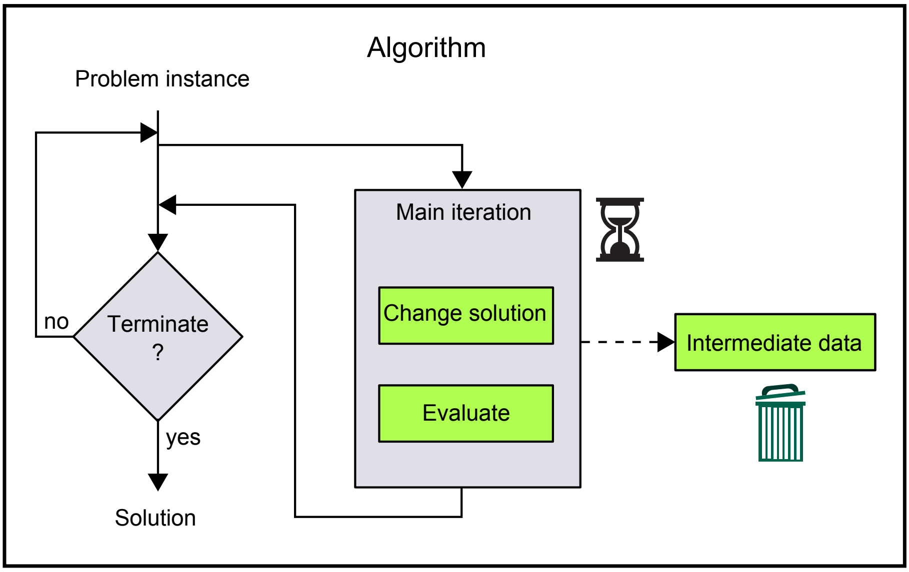

Figure 1.1: A general scheme of an algorithm solving combinatorial problems

gorithm (Barnhart et al. (1998); Feillet (2010)), dual solutions make up the intermediate data. Most of the approaches (Cheang et al. (2003); Burke et al. (2004); Kumar and Panneerselvam (2012); Di Gaspero and Schaerf (2000)) throw away this data and do not use it anymore in the future steps of an algorithm. Therefore, in this thesis, we propose methods exploiting the intermediate data in order to improve speed or convergence of combinatorial algorithms and, thus, to fill the gaps in the literature and to present the possible future directions in solving combinatorial problems.

1.1 Main Motivation

The complexity of the real-world problems (Curtois (2018); De Causmaecker et al. (2004)) becomes larger. Not only the problem size is bigger (e.g., more people to schedule, more cities to visit by a van, more tasks to be performed on machines, etc.). But, at the same time, the number of constraints also increases a lot, and their complexity grows as well. All these facts arise very high demands on algorithms and give a clear statement: **the combinatorial problems have to be solved quickly and robustly**.

1.1.1 Algorithms Speed

Speed was and still is the crucial aspect of every algorithm. The time demands can significantly vary in different domains. For example, the message scheduling in real-time systems has to operate in hundreds of milliseconds while the production plan for a car manufacturing can be solved for many hours since the result is not needed promptly and is calculated, e.g., only once per month. Nevertheless, speeding up the algorithms for both example domains would be desirable. The message scheduling could use the additional time to incorporate more difficult constraints or to increase safety by adding duplicated messages to achieve a certain level of redundancy. The production planning system would gain a time to generate not only one solution but more solutions which differ in constraints weights. This would allow the planner (i.e., the person using the planning system) to choose the most suitable variant with respect to the current needs of the factory.

Unfortunately, the real-world combinatorial problems are often NP-hard and even though the computational hardware is better every year it still does not give enough power to solve all instances of these problems.

1.1.2 Algorithms Robustness

As was already mentioned in the introduction, the wide spectrum of approaches for solving combinatorial problems exists in the literature. Each problem has hundreds of possible approaches which can solve it with varying results. This is not a big issue since it is quite logical that each problem requires more or less different approach. However, the backbone of algorithms across domains remains the same. The biggest changes can be found in specific operations for modifying the solution and in setting up several parameters and constants of the algorithm. Sometimes, this setting is not a straightforward process and, thus, it has to be set up empirically.

The bigger issue occurs with a changing of the algorithm itself or its parameters for even slightly different instances of the same combinatorial problem. It is not very unusual that many algorithms in the literature are tailor-made for the instances of the given problem and usage of the algorithm on a little bit different instance is possible only if major changes are implemented to the algorithm. One can say, that these algorithms are not robust at all.

1.2 Principal Problem

Personnel scheduling/rostering, namely nurse rostering problem (NRP) (De Causmaecker and Vanden Berghe (2011)), was chosen as a principal problem for this thesis. The NRP is very-well known NP-hard combinatorial problem where an algorithm tries to find a solution (so-called a *roster*) in which *shifts* are assigned to *employees (nurses)* for each day in a *planning period* in the best possible way. The assignment cannot be arbitrary, but it has to take into account various constraints. Two kinds of constraints are considered. *Hard constraints* cannot be violated at all since the result would be an infeasible roster (e.g., a nurse can work maximally at one shift per day). Soft constraints can be violated, but a weighted penalty for such a violation has to be added to the value of *objective function*. Therefore, the goal is to find the roster with the minimal value of objective function.

High-quality rosters have a significant impact on the entire hospital including its patients. First of all, a head nurse, who was responsible for creating rosters manually, have now more time for her normal work since the algorithm generates rosters automatically and the head nurse just verify the result and, if needed, she only performs small local changes in the final roster. Secondly, the created roster is righteous to all nurses and, thus, e.g., everybody has more or less the same workload, the same number of free weekends and the same number of not so popular shifts. This leads to overall better satisfaction of nurses with their work which is then reflected also in the happiness of patients with the provided services. Finally, the hospital saves money due to the more effective utilization of nurses.

For the NRP, it is typical that it is solved for the planning period of one month. Months usually do not differ much, at most in small units. For example, a new shift is introduced, or some constraint is removed/modified. From the algorithm perspective, the problem is the same, but instances are slightly different which can cause certain troubles for some algorithms which are fine-tuned for the specific instances. Moreover, rosters should be generated in a reasonable time especially when the date, on which the roster has to be publicly available for all nurses, is approaching.

Finally, we have chosen the NRP as the principal problem because several divers real-world benchmark instances for the NRP exist in the literature at one place and, thus, it is possible to convincingly demonstrate our approaches in comparison with the results from other researchers.

1.3 Key Ideas

Section 1.1 outlines the issues with solving of combinatorial problems in the current state-of-the-art. In the following subsections, an overview of key ideas on how to overcome these issues is presented.

1.3.1 Algorithm Design

For many decades, the *model-based methodology* (Paterno (2012)) is used for designing the algorithms for combinatorial problems. It is not a big surprise since it is a perfect tool to rigorously describe the complex mathematical systems based on a deep understanding of the problem. Moreover, the visual part of this design is readable for a wide audience and, thus, almost everyone can quickly understand the concept of the system from the high-level point of view. Furthermore, the design (or the parts of the design) can be quite easily reused for the similar combinatorial problems.

It may seem that this methodology is flawless. However, even the model-based approach has some limitations. The model cannot handle the arbitrary complexity. To be more precise, the complexity is bounded by a capacity of a human brain. More importantly, this approach does not take into account work with data. Obviously, it generates data flows, but it does not process data in order to let data drive the algorithm.

The data-driven approach (Qin (2012)) offers directions to grasp this unexplored area of processing the intermediate data. This approach is basically the opposite of the model-based approach. The whole process is driven by data which means that machine learning (ML) techniques should be involved in order to discover the desired knowledge or even to find out the completely new connections and correlations which one may not even know. The data-driven approach started to be more visible recently with the rise of deep learning (LeCun et al. (2015)), big data (Chen et al. (2014)) and autonomous driving systems (Chen et al. (2015)).

The data-driven approach is very powerful and can handle very complex problems (e.g., currently level 3 of automated driving (Casner et al. (2016))) which are not possible to describe by model-based approaches. However, the data-driven approaches usually require many data to receive some generality of the system. Moreover, it is very hard (in some cases it is not even possible) to verify/certify such systems nowadays.

Therefore, this thesis takes the best from both approaches, combines their advantages and negates their disadvantages. The algorithms in the thesis are designed by the model-based approach while the intermediate data from the main iteration are handled by the data-driven approach. The correctness of the algorithm is ensured by mechanisms outside the main iteration. The algorithm speedup and robustness is achieved by machine learning techniques applied inside the main iteration.

1.3.2 Machine Learning Enhancement

With the usage of the data-driven approach, ML techniques have to be used at some point. The ML techniques can be divided into many categories based on the purpose. In this thesis, we divide ML techniques to two high-level categories: *offline and online learning.*

The main differences between offline and learning are:

1. **Data manipulation**
   - Online learning usually does not store the incoming data. It uses it to update a learning model and then data is forgotten.
   - Offline learning uses the database where every incoming data is stored.
2. **Learning time**
   - Online learning mostly requires negligible time for learning/re-learning since it is done at runtime and it is not desirable to slow down the execution time of the overall system.
   - Offline learning does not have any requirements for the learning time. It is quite common that the learning takes even days.
3. **Re-learning / Adaptability**
   - Online learning utilizes the re-learning procedure which allows it to dynamically react to sudden changes in the incoming data.
   - Offline learning generally does not provide re-learning on time. Re-learning can be, of course, performed but it can take much time which is not suitable to do during the overall system runtime.
4. **Generalization**
   - Online learning mostly tends to follow the actual trends instead of global ones.
   - Offline learning finds the generality, so it is not influenced by the temporal trends.
5. **Accuracy**
   - Online learning can suffer from a lack of data which can negatively influence the accuracy of the learning model.
   - Offline learning achieves good accuracy. However, one has to be aware of over-fitting which can also significantly lower the generalization capabilities.

From the comparison mentioned above, it follows that both learning categories have different properties. Therefore, one has to decide which category is more suitable based on what is needed and required by the system where some ML technique will be incorporated. In this thesis, both categories are used for different purposes.

1.4 Related Work

This section briefly describes the main works related to the thesis. The detailed literature review can be found in Chapters 2 (*Section 1.1*), 4 (*Section 1.2*) and 3 (*Section 1.1*).

1.4.1 Neural Networks

The first studies that applied learning techniques to scheduling problems appeared in the 1990s. In general, neural networks are a suitable tool for pattern recognition (Bishop et al. (1995); Ripley (2007)) because they can learn from experience and deal with noise in the input data almost as well as human beings. A search-control policy for the resource-constrained scheduling problems was described by Zhang and Dietterich (2000) where a neural network is used as an approximator for a future value of the resource dilation factor in the next iteration of a scheduling algorithm. This approach, which was verified using the NASA space shuttle payload processing problem, outperformed the best known non-learning search algorithm at that time. A neural network and logistic regression were used by Li et al. (2011) to determine the correct order for executing low-level heuristics in a hyper-heuristic method. These techniques work as classifiers, and they decide whether the execution order for low-level heuristics is good or not. Experiments based on exam timetabling problems showed that both methods could significantly speed up the original algorithm. Li et al. (2012) also recommended the use of pattern recognition to evaluate the quality of the solutions, where a neural network distinguishes good and bad solutions based on the structure of the overall solution. The theoretical results indicated that speedup was achieved for the NRP and the educational timetabling problem.

1.4.2 Repetitive Pricing Problem

The BaP algorithm has been successfully employed to solve many combinatorial optimization problems, but authors have used various techniques to solve the pricing problem in particular. In the nurse rostering domain, the resource constrained shortest path (RCSP) Problem is often used to formalize the pricing problem. For example, Maenhout and Vanhoucke (2010) address an exact method based on the BaP algorithm for the NRP with hard constraints. Moreover, the authors explored various branching strategies and verified them on a custom artificial dataset. Another method based on the BaP algorithm for the NRP was presented in Burke and Curtois (2014). The method is not exact because several real-world benchmark instances (Curtois (2018)) were not solved to optimality even though timeout did not occur. In all the cases mentioned above, the pricing problem represented a considerable portion of the overall computation time. The need to solve repetitive and time consuming pricing problems was noticed by several researchers. Banerjee and Roy (2015) deal with solving repeated integer linear programming (ILP) problems within the branch-and-bound algorithm. First, their approach determines whether the ILP problems are similar. Then, a learned, boosting tree-based regression technique can be used to map the objective function and the constraints to the decision variables of a similar ILP problem. Even though the authors provide theoretical performance guarantees, the method is not exact. A dynamic RCSP as a pricing problem in a BaP algorithm was identified as a critical element of the algorithm in Zhu and Wilhelm (2013) because only the arc costs or arcs forbidding/prescribing change. A three-stage approach was proposed to deal with this re-optimization issue. The necessity of accelerating the solving process for the pricing problem is addressed in Ghoniem et al. (2015). The authors use a specialized dynamic programming procedure for the RCSP to solve a runway aircraft sequencing problem.

1.4.3 Processing Time Estimation

Some simple methods for estimating the processing time functions of the tasks were proposed in the literature such as using the average of the last n values of the processing time. However, these methods cannot be used if the processing times of the tasks are significantly different. A more successful approach is to use statistical methods or methods from machine learning (Page et al. (2008); Hutter et al. (2014)). In comparison with single resource environment, estimation of the processing time of the tasks in heterogeneous systems is even more complicated, i.e., the same task may have different processing times on the resources with different processing power. Therefore, the same estimation cannot be directly used for different resources. Iverson et al. (1999) solve this problem by benchmarking the resources. The benchmark result is appended to a feature vector of a task, and the k-nearest neighbor method is used to find similar observations from which an estimation of the processing time is computed. An alternative approach to appending a benchmark result to a feature vector is to scale the processing time of a task by the benchmark result of the resource on which the task is processed (Page et al. (2008)). However, none of these works deal with the scheduling of tasks solved by anytime algorithms.

1.5 Contributions

The main contributions of the thesis grouped by the theme are as follows:

1. The heuristic for the NRP enhanced by offline learning (mainly based on points 1, 4 and 5 from Section 1.3.2) ⇒ A neural network classifier is used to accelerate the evaluation phase in a tabu search algorithm. The classifier determines whether the changed solution is better than the original one or not.

    - The original design of a classifier determining a roster's quality is proposed.
    - The practical use of the classifier in a tabu search algorithm is shown.
    - The proposed approach achieves a better speedup and classification rate in comparison with Li et al. (2012).
    - The classifier is robust enough to handle changes in the number of employees to a certain extent.
    - The novel design of weak classifiers is introduced.

2. The exact algorithm for the NRP enhanced by online learning (mainly based on points 2 and 3 from Section 1.3.2) ⇒ For some combinatorial problems, more than 90% of the total central processing unit (CPU) processing time is consumed by repetitive solving the pricing problem in the BaP algorithm. A method utilizing the knowledge gained from previous executions of the pricing problem to reduce the solution space of pricing problems solved in future iterations is proposed. The method is based on a regression model which predicts a tight upper bound for the current iteration of the pricing problem.

    - Branch-and-price algorithm is analyzed in order to identify places which can be enhanced by ML techniques.
    - The regression model predicting an upper bound is designed to accelerate a solving process of the pricing problem.
    - The approach is verified in two different case studies: the NRP and the scheduling of time-division multiplexing (Minaeva et al. (2016)).
    - The improvement of the state-of-the-art Branch-and-price approaches for the general NRP (Burke and Curtois (2014)) and for the time division multiplexing schedulin (Minaeva et al. (2016)) is achieved.

3. Algorithm runtime estimation by online learning (mainly based on points 3 and 4 from Section 1.3.2) $\Rightarrow$ For some client-server scheduling systems, it is crucial to know the processing time of the incoming instances in order to correctly utilize the server and to keep the reasonable response time of the server. A regression method is used to estimate this time based on the well-chosen features of an instance.

    - The estimator providing the estimation of the whole processing time function is designed to bring accuracy to the scheduling system.
    - The scheduling system with the estimator is evaluated on a real client-server application from the domain of personnel rostering.

1.6 Outline

The following three chapters are dedicated to the author's relevant research results. Namely, the classifier evaluating the changed solution in tabu search algorithm is described in Chapter 2, the regression model predicting an upper bound for the pricing problem in BaP algorithm is addressed in Chapter 3, and the estimator of an algorithm runtime for the NRP instances is introduced in Chapter 4. Every chapter has a short summary consisting of: a full citation of the publication, a journal ranking and an online link to the publication. Each publication is presented in the original formatting given by the respective journal.

Finally, Chapter 5 concludes the thesis by summarizing the achieved results, discussing future directions and evaluating the fulfillment of the stated thesis goals.

Chapter 2

Roster Evaluation Based on Classifiers for the Nurse Rostering Problem

Full citation:
Roman Václavík, Přemysl Šůcha, and Zdeněk Hanzálek. Roster evaluation based on classifiers for the nurse rostering problem. *Journal of Heuristics*, 22 (5):667–697, Oct 2016. ISSN 1572-9397. DOI: 10.1007/s10732-016-9314-9.

Journal statistics according to the Journal Citation Report (2017):

Rank:	Q3 (Computer science, artificial intelligence)
	Q3 (Computer science, theory & methods)
Total Cites:	1122
Impact Factor:	1.129
5-Year Impact Factor:	1.864
Immediacy Index:	0.200
Citable Items:	20
Cited Half-life:	8.7
Citing Half-life:	>10

This paper is available at https://doi.org/10.1007/s10732-016-9314-9.

Chapter 3

Accelerating the Branch-and-Price Algorithm Using Machine Learning

Full citation:
Roman Václavík, Antonín Novák, Přemysl Šůcha, and Zdeněk Hanzálek. Accelerating the branch-and-price algorithm using machine learning. *European Journal of Operational Research*, 271(3):1055 – 1069, 2018. ISSN 0377-2217. DOI: 10.1016/j.ejor.2018.05.046.

Journal statistics according to the Journal Citation Report (2017):

Rank:	Q1 (Operations research & management science)
Total Cites:	43505
Impact Factor:	3.428
5-Year Impact Factor:	3.960
Immediacy Index:	0.817
Citable Items:	711
Cited Half-life:	9.7
Citing Half-life:	>10

This paper is available at https://doi.org/10.1016/j.ejor.2018.05.046.

Chapter 4

Adaptive Online Scheduling of Tasks With Anytime Property on Heterogeneous Resources

Full citation:

István Módos, Přemysl Šůcha, Roman Václavík, Jan Smejkal, and Zdeněk Hanzálek. Adaptive online scheduling of tasks with anytime property on heterogeneous resources. *Computers & Operations Research*, 76:95 – 117, 2016. ISSN 0305-0548. DOI: 10.1016/j.cor.2016.06.008.

Journal statistics according to the Journal Citation Report (2017):

Rank:	Q1 (Computer science, interdisciplinary applications)
	Q1 (Engineering, industrial)
	Q1 (Operations research & management science)
Total Cites:	11297
Impact Factor:	2.962
5-Year Impact Factor:	3.174
Immediacy Index:	0.809
Citable Items:	257
Cited Half-life:	8.2
Citing Half-life:	9.3

This paper is available at https://doi.org/10.1016/j.cor.2016.06.008.

Chapter 5

Conclusion

The thesis is focused on the domain of the personnel scheduling enhanced by machine learning techniques. These techniques bring a new view to the design of scheduling algorithms. The common model-based approach is combined with the data-driven approach which is used for handling the intermediate data. This allows achieving the speedup and the robustness of classical algorithms.

Three different approaches for the NRP were addressed to demonstrate the feasibility of our ideas. First, a neural network classifier for evaluating changes in a solution made by a local operator in tabu search algorithm shows very good speedups and resistance to minor changes in the consider instances. Second, a regression model predicting a tight upper bound for the pricing problem in the BaP algorithm also achieves noticeable speedups. Furthermore, it is designed in a general way and, thus, it can be used not only for the NRP but other combinatorial problems without any significant modifications as well. Finally, an algorithm runtime estimator helps the client-server scheduling system to reach the desired behavior which is comparable to the system using the exact knowledge about the algorithm runtime.

5.1 Fulfillment of Goals and Objectives

1. Study the state-of-the-art related to a solving of NP-hard combinatorial problems and identify possible improvements.

   ⇒ This goal was fully accomplished by providing the general overview in Chapter 1 while more details are in Chapters 2 (*Section 1*), 4 (*Section 1*) and 3 (*Section 1*). We have identified that the most algorithms are designed by a model-based approach and, thus, intermediate data which arises during a solving process is basically ignored in future algorithm steps. There is only a little research about exploiting this data in the literature. Therefore, we have proposed three different methods how this state can be improved with the data-driven approach.

2. Design a machine learning based approach to accelerate the evaluation of solution changes made by a local operator in a heuristic for the nurse rostering problem.

   ⇒ This goal was fully accomplished by designing a neural network method accelerating the evaluation phase in heuristic approaches. The method is presented in Chapter 2 and is based on a classifier, which can determine whether the changed solution (more precisely, the changed part of the solution) is better than the original or not. This decision is made much faster than a standard cost-oriented evaluation process. Moreover, our approach is illustrated on a tabu search algorithm where it rejects the majority of the potentially bad solutions and the remaining solutions are then evaluated in a standard manner.

3. Analyze and improve a branch-and-price algorithm for the nurse rostering problem by processing the data which occurs and is calculated during the solving process.

   ⇒ This goal was fully accomplished by identifying weak points in the BaP algorithm and subsequently implementing the proposed improvements to this algorithm. The analysis of the considered case-studies in Chapter 3 shows that more than 90% of the total CPU processing time of the BaP algorithm is consumed by solving the pricing problem. The pricing problem is repetitive and it solves the same problem from scratch differing only in the input dual prices. We demonstrate how to utilize the knowledge gained from previous executions of the pricing problem to reduce the solution space of pricing problems solved in future iterations by using a very fast regression model predicting a tight upper bound for the current iteration of the pricing problem.

4. Propose a method estimating an algorithm runtime of the nurse rostering problem instances.

   ⇒ This goal was fully accomplished by developing a general procedure estimating an instance difficulty for a system for an adaptive online scheduling of tasks with anytime property on heterogeneous resources. For the correct working of the system described in Chapter 4, it is crucial to know the instance difficulty, i.e., a relationship between the processing time and the quality of the instance solution. Therefore, we propose a regression method for estimating these relationships using information obtained from the already executed instances.

5. Verify the proposed methods on real-world instances and compare them with the state-of-the-art.

⇒ This goal was fully accomplished by performing extensive experiments which confirmed that the proposed methods are valid and the obtained results are better compare to the literature.

⇒ The classifiers were verified in Chapter 2 (*Section 4*) on the real-world NRP instances where the average speedup was four times in comparison with the algorithm without the classifier while the solution quality remains more or less the same. The algorithm with the classifier was also compared with the relevant literature result (Li et al. (2012)) and the speedup of three times was achieved.

⇒ The BaP algorithm with the regression model was tested on two different case studies in Chapter 3 (*Section 6*). For the NRP domain, the CPU processing time was reduced by 40% on average for the real-world instances and by 40% on average in comparison with the state-of-the-art approach (Burke and Curtois (2014)) as well. In case of the scheduling of time-division multiplexing for multi-core platforms, the CPU processing time was reduced by 22% on average in comparison with Minaeva et al. (2016).

⇒ Experiments with the regression method estimating the NRP instances difficulty in Chapter 4 (*Section 6*) showed that the proposed estimator is very helpful for the overall scheduling system. The system with the estimator achieved almost the same average quality across all instances (tasks) as the system with the exact knowledge about the instance difficulty, and it was superior to the system with a linear estimation.

5.2 Concluding Remarks

We believe that the data-driven design is a very important topic for the future of combinatorial algorithms since much unused various data is present during a solving process. So far, there is not too much attention in the literature about this topic, but in the recent years, one can see a growing trend. Fortunately, there are quite new domains (e.g., autonomous driving) showing how to combine model-based and data-driven approaches. Moreover, these domains are rapidly pushing the used technologies like a deep learning towards since they require better accuracies and faster learning/evaluating times. In a long-term, the domain of combinatorial optimization can benefit from this research and naturally improve its algorithms by applying the findings from these new domains.

Bibliography

Banerjee, A.G., Roy, N., 2015. Efficiently Solving Repeated Integer Linear Programming Problems by Learning Solutions of Similar Linear Programming Problems using Boosting Trees. Technical Report MIT-CSAIL-TR-2015-001. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Cambridge, MA 02139, USA.

Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W., Vance, P.H., 1998. Branch-and-price: Column generation for solving huge integer programs. Operations research 46, 316–329.

Bishop, C., Bishop, C.M., et al., 1995. Neural networks for pattern recognition. Oxford university press, New York.

Burke, E.K., Curtois, T., 2014. New approaches to nurse rostering benchmark instances. European Journal of Operational Research 237, 71–81.

Burke, E.K., De Causmaecker, P., Berghe, G.V., Van Landeghem, H., 2004. The state of the art of nurse rostering. Journal of scheduling 7, 441–499.

Casner, S.M., Hutchins, E.L., Norman, D., 2016. The challenges of partially automated driving. Communications of the ACM 59, 70–77.

Cheang, B., Li, H., Lim, A., Rodrigues, B., 2003. Nurse rostering problems—-a bibliographic survey. European Journal of Operational Research 151, 447–460.

Chen, C., Seff, A., Kornhauser, A., Xiao, J., 2015. Deepdriving: Learning affordance for direct perception in autonomous driving, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 2722–2730.

Chen, M., Mao, S., Liu, Y., 2014. Big data: A survey. Mobile networks and applications 19, 171–209.

Curtois, T., 2018. Employee scheduling benchmark data sets. http://www.schedulingbenchmarks.org/. Accessed: 2018-08-02.

De Causmaecker, P., Demeester, P., Berghe, G.V., Verbeke, B., 2004. Analysis of real-world personnel scheduling problems, in: Proceedings of the 5th international conference on practice and theory of automated timetabling, Pittsburgh, pp. 183–197.

De Causmaecker, P., Vanden Berghe, G., 2011. A categorisation of nurse rostering problems. Journal of Scheduling 14, 3–16.

Di Gaspero, L., Schaerf, A., 2000. Tabu search techniques for examination timetabling, in: International Conference on the Practice and Theory of Automated Timetabling, Springer. pp. 104–117.

Feillet, D., 2010. A tutorial on column generation and branch-and-price for vehicle routing problems. 4or 8, 407–424.

Ghoniem, A., Farhadi, F., Reihaneh, M., 2015. An accelerated branch-and-price algorithm for multiple-runway aircraft sequencing problems. European Journal of Operational Research 246, 34–43.

Glover, F., Laguna, M., 1998. Tabu search, in: Handbook of combinatorial optimization. Springer, pp. 2093–2229.

Hutter, F., Xu, L., Hoos, H.H., Leyton-Brown, K., 2014. Algorithm runtime prediction: Methods & evaluation. Artificial Intelligence 206, 79–111.

Iverson, M.A., Ozguner, F., Potter, L.C., 1999. Statistical prediction of task execution times through analytic benchmarking for scheduling in a heterogeneous environment, in: Heterogeneous Computing Workshop, 1999.(HCW'99) Proceedings. Eighth, IEEE. pp. 99–111.

Kumar, S.N., Panneerselvam, R., 2012. A survey on the vehicle routing problem and its variants. Intelligent Information Management 4, 66.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, J., Burke, E.K., Qu, R., 2011. Integrating neural networks and logistic regression to underpin hyper-heuristic search. Knowledge-Based Systems 24, 322–330.

Li, J., Burke, E.K., Qu, R., 2012. A pattern recognition based intelligent search method and two assignment problem case studies. Applied Intelligence 36, 442–453.

Maenhout, B., Vanhoucke, M., 2010. Branching strategies in a branch-and-price approach for a multiple objective nurse scheduling problem. Journal of Scheduling 13, 77–93.

Minaeva, A., Šůcha, P., Akesson, B., Hanzálek, Z., 2016. Scalable and efficient configuration of time-division multiplexed resources. Journal of Systems and Software 113, 44–58.

Page, A.J., Keane, T.M., Naughton, T.J., 2008. Scheduling in a dynamic heterogeneous distributed system using estimation error. Journal of Parallel and Distributed Computing 68, 1452–1462.

Paterno, F., 2012. Model-based design and evaluation of interactive applications. Springer Science & Business Media, Berlin.

Qin, S.J., 2012. Survey on data-driven industrial process monitoring and diagnosis. Annual reviews in control 36, 220–234.

Ripley, B.D., 2007. Pattern recognition and neural networks. Cambridge university press, New York.

Zhang, W., Dietterich, T.G., 2000. Solving combinatorial optimization tasks by reinforcement learning: A general methodology applied to resource-constrained scheduling. Journal of Artificial Intelligence Reseach 1, 1–38.

Zhu, X., Wilhelm, W.E., 2013. Implementation of a three-stage approach for the dynamic resource-constrained shortest-path sub-problem in branch-and-price. Computers & Operations Research 40, 385–394.

Curriculum Vitae

Roman Václavík was born in Hranice, Czech Republic, in 1987. He received his bachelor's degree in Web and multimedia from the Czech Technical University in Prague (CTU) in 2009. He continued at the CTU to study Software engineering where he received his master of science degree with honors in 2011 when he had defended his master thesis focused on the algorithms for the Nurse Rostering Problem. Subsequently, he has started his Ph.D. studies at the Department of Control Engineering on the topic of algorithms for personnel scheduling enhanced by machine learning techniques. During his studies, he participated on the projects GACR FOREST, DEMANES and Centre for Applied Cybernetics. He is interested in combinatorial optimization, scheduling, and machine learning.

His teaching activities in CTU involved course of Combinatorial Optimization, where he also participated in the preparation of the educational material. He has supervised several students' projects and diploma theses. He has also helped with an organization of two conferences in Prague (EUROSYS 2013, MISTA 2015).

Research results of Roman Václavík were presented in several international conferences (EU/ME 2013, POSTER 2013, MISTA 2013, PATAT 2014, MISTA 2015). Moreover, his results were published in the impacted international journals: Journal of Heuristics, Computers & Operations Research and European Journal of Operational Research. Finally, he is a coauthor of a chapter in book Runtime Reconfiguration in Networked Embedded Systems.

Czech Technical University in Prague
Prague, August 2018 *Roman Václavík*

List of Author's Publications

All of the author's publications are directly related to the topic of the thesis. They are separated into four groups as follows.

Publications in Journals with Impact Factor

Roman Václavík, Antonín Novák, Přemysl Šůcha, and Zdeněk Hanzálek. Accelerating the branch-and-price algorithm using machine learning. *European Journal of Operational Research*, 271(3):1055 – 1069, 2018. ISSN 0377-2217. DOI: 10.1016/j.ejor.2018.05.046. **Co-authorship 25 %, journal rank Q1, 1 citation**.

Roman Václavík, Přemysl Šůcha, and Zdeněk Hanzálek. Roster evaluation based on classifiers for the nurse rostering problem. *Journal of Heuristics*, 22(5):667–697, Oct 2016. ISSN 1572-9397. DOI: 10.1007/s10732-016-9314-9. **Co-authorship 33.3 %, journal rank Q3, indexed in Web of Science**.

István Módos, Přemysl Šůcha, Roman Václavík, Jan Smejkal, and Zdeněk Hanzálek. Adaptive online scheduling of tasks with anytime property on heterogeneous resources. *Computers & Operations Research*, 76:95 – 117, 2016. ISSN 0305-0548. DOI: 10.1016/j.cor.2016.06.008. **Co-authorship 20 %, journal rank Q1, indexed in Web of Science**.

Book Chapters

Roshan Kotian, Stefano Galzarano, Claudio Bacchiani, Aly A. Syed, Přemysl Šucha, Roman Václavík, and Andrei Pruteanu. Runtime validation framework. In Zoltan Papp and George Exarchakos, editors, *Runtime Reconfiguration in Networked Embedded Systems: Design and Testing Practices*, pages 93–112. Springer Singapore, Singapore, 2016. ISBN 978-981-10-0715-6. DOI: 10.1007/978-981-10-0715-6_4. **Co-authorship 14.3 %**.

International Conferences and Workshops

Antonín Novák, Roman Václavík, Přemysl Šůcha, and Zdeněk Hanzálek. Nurse Rostering Problem: Tighter Upper Bound for Pricing Problem in Branch and Price Approach. In *7th Multidisciplinary International Conference on Scheduling: Theory and Applications*, pages 759–764. MISTA, Prague, Czech Republic, 2015. **Co-authorship 25 %**.

Přemysl Šůcha, István Módos, Roman Václavík, Jan Smejkal, and Zdeněk Hanzálek. Online Scheduling System for Server Based Personnel Rostering Applications. In *10th International Conference of the Practice and Theory of Automated Timetabling*, pages 561–564. PATAT, York, United Kingdom, 2014. **Co-authorship 20 %**.

Roman Václavík, Přemysl Šůcha, and Zdeněk Hanzálek. A Low Time-Consuming Rosters Evaluation in Personnel Scheduling Problems Based on Pattern Learning. In *14th EU/ME workshop*, pages 38–43. EU/ME, Hamburg, Germany, 2013. **Co-authorship 33.3 %**.

Roman Václavík, Přemysl Šůcha, and Zdeněk Hanzálek. Roster Evaluation Based on the Classifiers for the Nurse Rostering Problem. In *6th Multidisciplinary International Conference on Scheduling: Theory and Applications*, pages 744–747. MISTA, Ghent, Belgium, 2013. **Co-authorship 33.3 %**.

Roman Václavík. A Boosting Algorithm for the Classifiers in the Nurse Rostering Problems. In *17th International Student Conference on Electrical Engineering*, pages 1–5. Czech Technical University in Prague, Prague, Czech Republic, 2013. **Co-authorship 100 %**.

Other Publications

Roman Václavík. Preliminary Doctoral Thesis: Roster Evaluation Based on the Classifiers for the Nurse Rostering Problem. Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic, 2013. **Co-authorship 100 %**.

Czech Technical University in Prague
Prague, August 2018 *Roman Václavík*

This thesis is focused on the domain of the personnel scheduling enhanced by machine learning techniques. Its goals were set as follows:

1. Study the state-of-the-art related to a solving of NP-hard combinatorial problems and identify possible improvements.

2. Design a machine learning based approach to accelerate the evaluation of solution changes made by a local operator in a heuristic for the nurse rostering problem.

3. Analyze and improve a branch-and-price algorithm for the nurse rostering problem by processing the data which occurs and is calculated during the solving process.

4. Propose a method estimating an algorithm runtime of the nurse rostering problem instances.

5. Verify the proposed methods on real-world instances and compare them with the state-of-the-art.