Chapter 1

Introduction

1.1 Motivation

Considering human preferences and integrating their unique domain knowledge in scheduling processes improves efficiency and increases human well-being. By aligning priorities of the management with the priorities and experience of the workforce, organizations can ameliorate productivity and the use of human expertise as well as enhance employee satisfaction and motivation. However, careful planning to ensure effective and employee-friendly use of human resources is both essential and challenging.

Incorporating employee preference in scheduling decisions has several advantages for an organization. First, considering employee preferences increases job satisfaction and therefore plays a central role in reducing additional staff turnover and increasing job attractiveness (Abbink et al. 2005), which is of particular importance in light of the ongoing shortage of skilled workers. Second, organizations are not only able to foster higher job satisfaction but also improve productivity by taking employee preferences into account for workforce scheduling (Loch and Wu 2008, Kesavan et al. 2022). Third, employees whose preferences are considered tend to experience lower stress levels, which positively impacts their mental and physical health, resulting in fewer sick days and a healthier, more engaged workforce (Borndörfer et al. 2015).

Similarly, organizations can benefit from leveraging human knowledge for scheduling decisions. In many contexts, such as crew scheduling or healthcare scheduling, humans possess a large amount of non-quantifiable data, unique domain knowledge, and years of experience that are of high value for decision-making (Ibrahim and Kim 2019). This expertise often includes nuanced understanding as well as contextual and situational awareness that cannot be quantified easily and fed into an algorithm. By combining this tacit knowledge with data-driven optimization and machine learning, organizations can improve outcomes and simultaneously make targeted use of the scarce human resource (Kesavan and Kushwaha 2020, Syntetos et al. 2016).

The current challenge lies in determining how and when to incorporate human preferences and knowledge into scheduling processes. When considering employee preferences, capturing and including these preferences early and effectively in processes like scheduling is often challenging. Balancing several factors is essential to improve organizational productivity while addressing the preferences of the workforce. In times of severe staff shortages, especially in sectors like healthcare, it has become crucial to integrate human expertise only when it adds significant value to ensure that the already high workload of the staff is not further exacerbated.

This thesis aims at improving the scheduling processes as well as the well-being of the employees. Leveraging data-driven approaches, such as optimization and machine learning, and collaborating closely with experts from industry, we provide both methodologically sound and practically relevant insights. Motivated by challenges in research and practice, we present novel methods based on optimization and machine learning to incorporate human preferences and human knowledge in two different scheduling contexts: First, we consider the crew scheduling problem at a large European railway freight carrier. Second, we consider the surgery scheduling problem at a university hospital. In addition to the methodological and data-based contributions, we present a field study where we test one approach in practice. The results reveal several advantages for both human well-being and productivity, and provide insights as well as important implications for both research and practice.

1.2 Outline and Contribution

This thesis is structured into four main chapters (Chapters 2 - 5), and a conclusion (Chapter 6) that summarizes the results and provides an outlook for future research. All projects share the goal of improving scheduling processes and the well-being of the humans involved in and affected by these processes. Each of the four main chapters features one research project. In these chapters, we analyze real-world data sets to develop data-driven approaches which integrate human preferences and human expertise in scheduling decisions. Each chapter is written as an independent paper and consists of an abstract, a literature review, a methodological part, a data description part, a results section and a conclusion.

Chapter 2 presents a combined approach of machine learning and combinatorial optimization to include the preferences of planners in a crew scheduling optimization algorithm.[1] We collect a data set of 16,000 train driver duties from a large European railway freight carrier that were labeled with thumbs up ("favorable") or thumbs down ("unfavorable"). This way, planners could indicate whether each duty met their preferences or not. Using these labels, we train a machine learning model to predict these preferences and integrate the resulting model in a column generation algorithm which generates feasible train driver duties. We observe that this combination of machine learning and combinatorial optimization increases the share of duties that meet the preferences of the planners by over 12% while at the same time improving productivity of the generated duties. As a result, the cost level of the new schedules is similar to the cost level of the schedules generated without machine learning. This

[1] Chapter 2 is based on the paper by Theresa Gattermann-Itschert, Laura Maria Poreschack, and Ulrich Thonemann that was published in *Transportation Science* (Gattermann-Itschert et al. 2023). Theresa Gattermann-Itschert and Laura Maria Poreschack both contributed equally to the description of the problem setting, development of the solution approach, and the analysis of the results. The data collection and training of the prediction model was mainly done by Theresa Gattermann-Itschert. The literature review, the integration of the prediction model in the column generation algorithm and the empirical study was mainly done by Laura Maria Poreschack. Overall, the work was divided fairly between Theresa Gattermann-Itschert and Laura Maria Poreschack. Ulrich Thonemann participated in discussions about the prediction model, and gave input for the integration method and the positioning of the paper. The paper further benefited from the comments of two anonymous referees, an associate editor and an area editor of Transportation Science.

demonstrates that including preferences does not lead to an increase in cost, and can even enhance productivity.

Chapter 3 expands on the insights, the methodology and the model developed in Chapter 2.[2] We test the machine learning model in the field and interview planners from all planning regions of the European railway freight carrier. We observe that the trained machine learning model is effective in imitating the planners' preferences. Moreover, the interviews reveal that there are regional differences in preferences leading to challenges in scheduling. This can be mitigated using our machine learning approach and developing additional regional features. The "regional" model, which incorporates regional features, demonstrates a 3% improvement in AUC compared to the "global" model that excludes these features. This enhancement translates to the regional model correctly classifying 70% of duties across all regions, with performance reaching up to 73% in specific regions. Importantly, this improvement in classification accuracy does not lead to an increase in duty costs, which remain comparable to the global model.

Furthermore, an analysis of planner feedback and a comparison with the model's most influential features indicate strong alignment between the planners' priorities and the rating rationale of the machine learning model.

Chapter 4 analyzes the potential of segmenting prediction tasks to leverage human expertise and alleviate human workload at the same time.[3] We consider three different work modes: In one mode, a human solves the task alone. In a second mode, a human and an algorithm solve the task together. In the third mode, an algorithm solves the task alone. We use a data set with over 70,000 surgeries of a large university hospital and focus on the task of surgery duration prediction. Employing an ex-post task segmentation approach, we

[2]Chapter 3 is single-authored and benefited from discussions with Ulrich Thonemann.

[3]Chapter 4 is based on the paper by Dominik Walzner, Laura Maria Poreschack, Andreas Fügener, Sebastian Schiffels and Christof Denz that was published in the *Proceeding of the International Conference on Information Systems (ICIS)* (Walzner et al. 2023). To the application study, the analysis of the results and the discussion both Dominik Walzner and Laura Maria Poreschack contributed equally. The literature review was mainly done by Laura Maria Poreschack. The theory development was mainly done by Dominik Walzner. Andreas Fügener and Sebastian Schiffels participated in discussions on the application study, and provided support for positioning the paper. Christof Denz collected the data. The paper further benefited from the comments of three anonymous referees and an associate editor of the conference proceedings.

observe that we can improve task performance while simultaneously decreasing human workload, which is of particular interest in industries with staff shortages, such as the healthcare sector. This chapter provides evidence that it is not optimal to employ a "one size fits all" solution, and simply replace humans by algorithms to alleviate their workload. We argue for a more nuanced approach, which addresses performance and workload requirements simultaneously, and demonstrate its effectiveness.

Chapter 5 investigates how to dynamically manage the effort-value tradeoff for integrating human judgment into algorithmic predictions.[4] Integrating human expertise into algorithmic models for surgery duration prediction has been shown to improve accuracy compared to algorithms alone. However, in healthcare, where human resources are limited and opportunity costs are high, such involvement must be carefully managed to ensure time is focused on critical medical tasks. This study presents an ex-ante framework for selectively incorporating human expertise based on its cost and algorithmic uncertainty.

Using a data set of over 70,000 surgeries from a university hospital, we train a meta model to predict the value of integrating human expertise in the algorithmic prediction process. Based on this, the framework can decide when human input is worth the effort and when algorithms should predict the duration on their own. Results show that selectively integrating human expertise not only enhances prediction performance but also reduces human workload by 30% to 80% as well as planning costs by 27% to 35% compared to the current approach. This demonstrates that a cost-sensitive strategy can optimize the use of human expertise, reserving it for cases where it adds the most value. This chapter exhibits how the potential analyzed in Chapter 4 can be realized in practice using an ex-ante approach to decide which task to assign to which work mode. As most features that indicate whether human input is of value are algorithmic meta

[4]Chapter 5 is joint work with Dominik Walzner, Andreas Fügener, Ulrich Thonemann and Christof Denz. The methodology, the application study, the training of the meta model, the development of the framework and the analysis of the results was mainly done by Laura Maria Poreschack. Dominik Walzner and Laura Maria Poreschack contributed equally to the literature review. The prediction models were trained by Dominik Walzner. Andreas Fügener and Ulrich Thonemann participated in discussions about the framework design and the positioning of the paper. Christof Denz collected the data.

features capturing algorithmic uncertainty, this approach is highly generalizable to contexts beyond surgery duration prediction.

Overall, our research contributes toward a better understanding of how human preferences and expertise can be integrated into scheduling decisions and how this affects productivity and human well-being. Using real-world data sets, we can contribute both academically and practically relevant insights.

Chapter 2 contributes to the literature as well as industry practice by developing a novel approach to integrate planners' preferences into a crew scheduling optimization model. We apply machine learning to learn these preferences and train a classifier that predicts whether a duty meets the preferences of planners or not. We incorporate these predictions in the optimization model. By learning preferences, our approach integrates expert knowledge without the need to formulate it mathematically. Different from other studies, we do not only apply machine learning, but also use interpretability measures to better understand the model predictions. The most influential features turn out to be related to the duty structure and include duty productivity, which provides valuable insights for practitioners. Consequently, we improve the scheduling process for both the employees as well as the organization: First, the preferences of employees are considered which improves their job satisfaction and well-being. Second, we improve productivity and efficiency of the duties for the railway freight carrier.

Chapter 3 contributes to the literature and practice in several ways. First, we develop a model specifically designed to capture planners' preferences at a regional level. Unlike previous studies that primarily emphasize cost minimization or generic preference integration, our approach incorporates regional variations, enabling a more tailored and accurate crew scheduling process aligned with the localized needs and priorities of railway operations. Second, through in-depth interviews with planners, we gain qualitative insights into the planning process that extend beyond quantitative data. Third, we validate our model through a field experiment, providing empirical evidence of its effectiveness in a real-world context. This validation demonstrates the model's applicability in day-to-day railway operations.

Chapter 4 contributes to the literature on digitization in the healthcare sector by considering algorithmic automation and the integration of human expertise in algorithmic predictions, and by evaluating to which extent physicians could be relieved from the task of surgery duration predictions. Additionally, we provide a more precise prediction methodology, which is able to outperform the current status quo, i.e., the human. We also contribute to the literature on human-AI collaboration by juxtaposing the effort and value of human involvement. Our results demonstrate the potential of task segmentation in human-AI collaboration.

Chapter 5 contributes to two literature streams: First, we contribute to the literature on human judgment integration by presenting an integration framework that enables a more nuanced use of human judgment for applications where human resources are scarce. Our framework allows to determine a cost-sensitive integration policy that is adaptable to different opportunity costs of human involvement. Therewith, we put a strong focus on the costs of involving humans in the prediction task, an important aspect which has not received much attention in the literature where the focus was mostly on improving performance. For example, frameworks that present judgmental adjustment integration policies focused mostly on deciding whether human judgment increases or decreases the prediction performance, see e.g., Chen et al. (2023). Second, we contribute to the healthcare operations management literature by approaching the problems of reliable surgery duration predictions and scarce human resources simultaneously. We provide a prediction methodology, which outperforms existing approaches, such as human predictions and full automation. Consequently, we alleviate the workload of the medical staff in two dimensions: First, they face less overtime due to more reliable schedules. Second, they spend less time on the administrative scheduling task, and can concentrate more on their medical tasks.

Overall, our findings contribute to the literature on operations management, scheduling, and human-AI collaboration, and in particular to the areas of human preferences and human judgment integration. Our developed methods enable us to leverage the strengths of both humans and algorithms to improve the well-being of the humans affected by the scheduling processes. Additionally, we derive insights for both researchers and practitioners on how to incorporate human preferences and expertise.

Chapter 2

Using Machine Learning to Include Planners' Preferences in Railway Crew Scheduling Optimization

In crew scheduling, optimization models can become complex when a large number of penalty terms is included in the objective function to take planners' preferences into account. Planners' preferences often include non-monetary aspects for which both the mathematical formulation and the assignment of appropriate penalty costs can be difficult. We address this problem by using machine learning to learn and predict planners' preferences. We train a random forest classifier on planner feedback regarding duties from their daily work in railway crew scheduling. Our data set contains over 16,000 duties that planners labeled as good or bad. The trained model predicts the probability that a duty is perceived as bad by the planners. We present a novel approach to replace the large construct of penalty terms in a crew scheduling optimization model by a single term which penalizes duties proportionally to the predicted probability of being assessed as unfavorable by a planner. By integrating this probability into the optimization model, we generate schedules which include more duties with preferred characteristics. We increase the mean planner acceptance probability by more than 12% while only facing a marginal increase in costs compared to the original approach that utilizes a set of multiple penalty terms. Our approach combines machine learning to detect complex patterns regarding favorable duty characteristics and optimization to create feasible and cost efficient crew schedules.

2.1 Introduction

Railway crew scheduling is a complex optimization task. The objective is to combine a large number of trips to feasible train driver duties at low cost. A sequence of train driver activities performed by one train driver is called a duty. Such a sequence can consist of trips, repositionings and breaks, and forms one or two days of work for an anonymous crew member. Duty feasibility is subject to various physical restrictions and regulations such as the maximum working time, operational requirements and the fact that a duty must start and end at the same depot. Duties can be split up into four different groups. Duties with a duty length of below five hours are considered as *very short duties*. Duties with a duty length between five and twelve hours are considered as *regular duties*, and duties with a duty length between twelve and fourteen hours are considered as *very long duties*. If a duty is so long that labor regulations require a hotel stay for the driver to rest, a duty is called a *hotel duty*.

Typically, such crew scheduling problems are solved with column generation, iterating between a sub problem to generate feasible duties and a master problem to select the cost-optimal set of duties (Barnhart et al. 1998).

In practice, objectives besides cost minimization are often relevant. Planners have preferences regarding the duty characteristics and include train drivers' preferences into their considerations. For example, train drivers tend to dislike duties with several train changes, very long duties, and duties with a high number of repositionings. Planners can take such preferences into account to improve the train drivers' satisfaction. High train driver satisfaction is important for railway companies, for example to avoid strikes (Abbink et al. 2005). Planners can accept small increases in overall costs if, in turn, certain preferences are better met. Hence, they are not necessarily interested in a cost-optimal solution, but in a solution that meets a variety of preferences at reasonable costs. Such preferences can be incorporated in the optimization problem by including corresponding penalty terms in the objective function.