Statistical Evaluation of the Usability of Decision-Oriented Graphical Interfaces in Scheduling Applications

Anna Prenzel, Georg Ringwelski

Hochschule Zittau/Görlitz
Fakultät Elektrotechnik und Informatik Obermarkt 17
02826 Görlitz
a.prenzel@hszg.de
g.ringwelski@hszg.de

Abstract: In this paper, we present the results of a usability study in interface design for scheduling systems. We show that providing possibilities to interactively create schedules improves the quality of decision making. The considered ways of interaction are tailored to the needs of human schedulers. They facilitate and optimize planning decisions at all stages of the scheduling procedure. Based on a study with 35 test subjects and overall 105 hours of usability testing we verify that the use of some features improves both quality and practicability of the produced schedules.

1 Introduction

Problem. An important challenge in the design of scheduling systems is to incorporate human factors sufficiently [JWM04]. Normally schedulers require interactive features including

- adapting schedules during execution due to accidents that must be resolved immediately
- adapting future schedules due to expert knowledge which was not included in the model a priori
- evaluating different scenarios for parts of a future schedule.

Therefore schedulers are reluctant to accept pure push-the-button-optimizers, that do not include expert decisions. Scheduling systems should make it easy both to make the right decisions, and to enter the decisions in the system [BBIM96].
Contribution. The goal of this paper is to evaluate different ways to design a user interface that targets these requirements. A fleet scheduling system, which belongs to an information system for water suppliers, serves as an example.
We describe a set of user interface features that can be combined to different human-computer interaction models. We evaluate the models based on an empirical study we

carried out in 105 hours of usability testing with 35 test subjects. Our study shows that the quality of the produced schedules correlates with use and availability of the regarded features.

1.1 The Structure of Scheduling Problems

The main concern of scheduling is the assignment of jobs to resources, for example, items for production, items for transport or shifts in a hospital. Machines, vehicles and employees can be considered as resources. Scheduling systems are expected to solve combinatorial problems such as finding sequences or start times of jobs, good resource utilization, minimal makespan and many more. Solving these problems is complex (often NP-complete) because solutions have to satisfy numerous constraints about the start time of jobs (time windows, non-overlapping constraints, sequencing constraints) and the assignment to resources (capacity constraints, qualifications).
Our case study in fleet scheduling is based on a formal model described in [KLMS05]. In this application jobs have to be carried out at different geographic locations. The scheduling task is to create a route for each vehicle, such that all deadlines can be met. The objective functions to minimize are:

- The total travel time between each two jobs in the schedule (cost function)
- The time between the beginning of the first and the end of the last job in the schedule (execution time)

Furthermore, the workload[1] should be balanced between the resources.

2 Interactive Decision Support

2.1 Requirements for the Decision Support

Expert schedulers know the criteria that make their schedules practicable or impracticable and prefer certain schedules over others [FWW10]. Their preferences can be expressed in terms of additional start time and resource constraints like:

- „start this job not until 10 o'clock"
- „use resource X (not) for this job"

Preferences usually include some uncertainty. It is not clear from the start whether and to what extent they can be incorporated in a feasible solution[2] and how much the quality of

[1] In our case, the workload of a resource is the sum of the durations of all assigned jobs
[2] feasible = no violation of constraints

the overall schedule is affected. The scheduler has to evaluate several possible solutions, in order to decide which one comes closest to his ideas.
Furthermore, it is usually not obvious how to set the weight of multiple optimization goals. Schedulers often derive them from existing schedules and use them for subsequent adaptations. For example, in our case study a compromise between balancing the workload and reducing the overall travel time has to be found.
Subsequent modifications of schedules play a big role in practical scheduling as well ([MT06], [FWW10]). For different reasons there might be unanticipated changes to schedules being carried out. For example, a schedule has to be adapted if a resource breaks down or a new job has to be included in case of an event. Again, there might be preferences about the best way to perform modifications. However, only minimal changes are allowed to schedules, that are already being carried out.

The scheduling system can support these modifications in multiple ways. Above all, automatic optimization capabilities relieve the user from the manual scheduling work [WJM06]. The interface should allow the user to freely choose which part of the schedule is to be optimized automatically and which part is scheduled manually. In addition, it should allow a partial specification of attributes of jobs and resources and resolve the remaining attributes automatically. This way the human operator can focus on the properties of the schedule that are relevant for his decision.
Manual scheduling operations can be supported by the system too. In order to prevent faulty decisions, the system should supervise the compliance with the underlying constraints. In doing so it is not sufficient to show an error message as soon as a constraint is violated. We rather suggest to visualize the scope of action already when the human is about to make a decision. This means highlighting valid properties for the considered jobs and resources with regard to the state of the current schedule. This way the human does not have to make the effort to withdraw a faulty decision.

2.2 Decision Support Features

We have designed a set of interaction features that can be used to build a scheduling interface providing the recommended decision support:

Full Optimization (FO): A control to optimize the whole schedule. It allows choosing from various built-in cost functions.

Single Job Optimization (SJO): The interface allows to select a single job in the schedule and triggers automatic optimization of its position (using the same cost functions as FO). Remaining jobs in the schedule are kept unchanged.

Resource Optimization (RO): Like SJO. All jobs belonging to the same resource can be selected at once.

Group Optimization (GO): Like SJO. Any group of jobs from different resources can be selected.

Fit-in (FIT): The interface allows the user to define the position of a job within the sequence and looks for a valid start time.

Constraint Highlighting (CH): The interface recognizes the intention to change a property of a job and colours possible values

- red, if they are invalid
- green, if they are valid

with regard to constraints of the individual job.

Enhanced Constraint Highlighting (ECH): Additional to CH: values of properties, that violate constraints in relation to other jobs are coloured

- yellow, if the value can be applied as soon as the properties of conflicting jobs are adapted
- grey, if the value can never be applied in conjunction with the conflicting jobs.

Fixation (FIX): The interface allows the direct input of the desired properties of one or more jobs. They are turned into additional constraints to be considered by all features. For example, the scheduler can fix the start time or resource assignment for certain jobs.

3 Evaluation of the Decision Support

3.1 The Test Interface

Figure 1: Test interface for fleet scheduling. ECH is applied to job 17.

The fleet scheduling interface that is used for the tests is shown in Figure 3.1. It consists of a Gantt Chart, that shows the temporal sequence of jobs (grey) for each resource and

the travel times between them (brown). Full optimization can be accessed via the toolbar, while RO, GO and SJO are available in context menus for selected jobs, groups of jobs or resources. Fixation and Fit-in is possible via context menus, too. The property dialogs of the jobs provide additional possibilities to fix start times and resources.
In our case, CH and ECH are used as follows [DR11]: While dragging a job in the Gantt chart, positions are highlighted green, if they are within its time window, and red, if they are outside the time window or in an invalid resource. ECH additionally indicates, whether a job can be fitted into the tour of a vehicle (yellow positions). Sometimes a job cannot be included in a tour, because there is no feasible sequence together with the other jobs in this tour. In this case the background colour of the resource is grey. For further support we provide a map (not depicted).

3.2 Combining Decision Support Features to Interaction Models

Using the fleet scheduling interface we want to provide an empirical evaluation of

- the suitability of the features for efficiently performing scheduling tasks
- the quality that can be achieved in terms of the cost function.

For this we combine the decision support features to 5 interaction models. They are shown in Table 3.3.2.

Model 1/2: Manual scheduling. Model 1 uses CH, model 2 additionally provides ECH.

Model 3: Full optimization (FO). Manual modifications can only be done after the optimization, as fixation is not allowed.

Model 4: Like model 3, but fixation is allowed. Manual scheduling decisions can be fixed before calling FO. After the optimization, manual changes can be fixed and any number of reoptimization steps is allowed.

Model 5: Optimization of groups of jobs is allowed. Fixation can be achieved indirectly by excluding manually positioned jobs from optimization groups. Several optimization steps with overlapping groups of jobs are possible.

The models are equivalent in that they all can be applied to the same scheduling problems. However, they differ in the decision support provided for creating and modifying schedules. The effort necessary to obtain an optimal and practicable schedule depends on the degree of automation and the assistance with manual operations. Thus we expect that the choice of the model influences the quality of the produced schedules in terms of the considered cost functions. Furthermore, we expect that the availability of some features like ECH or FIX clearly improves the overall user experience which is usually reflected in a higher rate of task success and lower processing times.
In order to compare the models, we have designed several test tasks that were carried out by peer groups with each model respectively.

3.3 Setup of the Usability Test

We have formed 5 test groups each consisting of 7 students from different faculties of our institution. The subjects were asked to perform 6 scheduling tasks. The models available for the particular tasks were dependent on the test group. We determine the best model for each task by comparing the average performance and confidence interval in the following metrics: accumulated travel time, task completion, time effort, number of undo operations and number of manual interactions. The tests took 3 hours per participant including a briefing of 30 minutes at the start. The maximum duration for each task was set to 15 minutes.

3.3.1 Design of the Test Tasks

The test tasks were designed to reflect typical scheduling situations, as described in Section 2.1. However, the participants had no experiences in scheduling. Therefore the relevant scheduling preferences that would otherwise arise from the expert knowledge of the scheduler had to be predefined for each task.

1. Schedule a set of jobs such that the total travel time is minimized and the workload is balanced between the resources. For some jobs there are precedence constraints.
2. Schedule a set of jobs such that the total travel time is minimized and the workload is balanced. For some jobs fixed start times and resources are given.
3. An additional vehicle is to be utilized. Change the given schedule such that some suitable jobs are assigned to it.
4. An event occurs and requires an additional job. The working schedule must include the job as early as possible, but it has to remain unchanged until 10 o'clock.
5. Schedule a set of jobs such that the total travel time is minimized and the workload is balanced. Jobs beyond the German-Polish border must be carried out in one piece.
6. Change the current schedule such that vehicle 3 finishes work at 12 o'clock. Remaining jobs have to be assigned to other vehicles.

The tasks are to be carried out with 4 vehicles and about 25 predefined jobs. All jobs have time window and resource constraints. In addition to fulfilling the tasks, the participants also have to strive for a compromise between low travel time and balanced workload (with the former being the most relevant criterion). The scheduling interface used in the tests offers two cost functions for automated optimization: „minimize travel time" and „balance workload". However, due to the schedule specifications for the particular tasks and the required compromise between the cost functions it is not possible to fulfill the tasks only with full optimization (FO). Instead, manual modifications and / or a clever use of group optimization (SJO, GO, RO) are required.

Table 1: Example peer groups and models for task 1.

Models	Features	Persons	Group (Task 1)
-	-	7	1
Model 1	CH	7	2
Model 2	ECH	21	3, 4, 5
Model 3	FO + ECH	7	1
Model 4	FO + FIX + ECH	14	2, 3
Model 5	SJO + GO + RO + FIX + FIT + ECH	14	4, 5

3.3.2 Assignment of Test Groups to Interaction Models

The table below shows the distribution of test persons to different models. The models are divided into two areas: manual optimization (model 1 and 2) and automated optimization (models 3, 4 and 5). The participants first carried out their tasks manually and then repeated them with the help of automatic features. The assignment of models to groups changes from task to task. This ensures that each group deals at least one time with each interaction model. We assigned fewer participants to models that were expected to be very difficult (model 1 and the model without any features) or discouraging for the test subjects.

4 Results

4.1 Usability Metric 1: Travel Time

In Figures 2 - 4 the achieved qualities of the schedules are shown for each particular task. The average qualities are influenced by the number of successfully completed tasks. Both task 6 and task 1 turned out to be insoluble for our testers in 15 minutes if no decision support was provided. Consequently, we cannot present further results.

The schedules created with manual features (model 1 and 2) were worse than those created with automatic optimization (model 3, 4 and 5). It can be concluded that ECH is not sufficient for creating complete schedules. Although ECH provides assistance for creating *feasible* schedules, achieving *optimality* is considerably more difficult. Thus manual scheduling is only applicable for smaller parts of a schedule. As shown in Figure 4, it is frequently used in the other models in order to modify existing schedules or to fix decisions before optimization.

Comparing models 3 and 4, the quality decreases if fixation is not allowed. This suggests that preferences should be incorporated in advance (FIX) rather than after automated optimization. In this case, the overall quality of the schedule can be optimized with regard to the fixations.

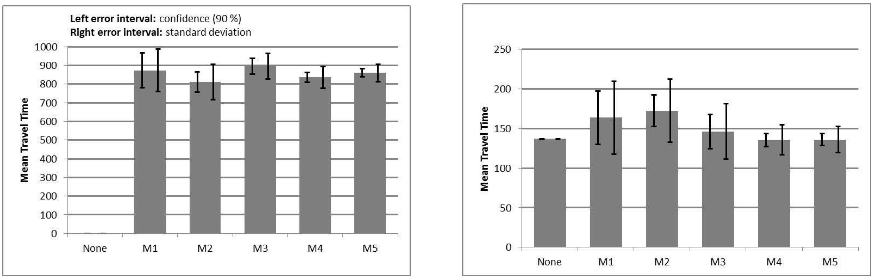

Figure 2: Mean travel time - task 1 and 2.

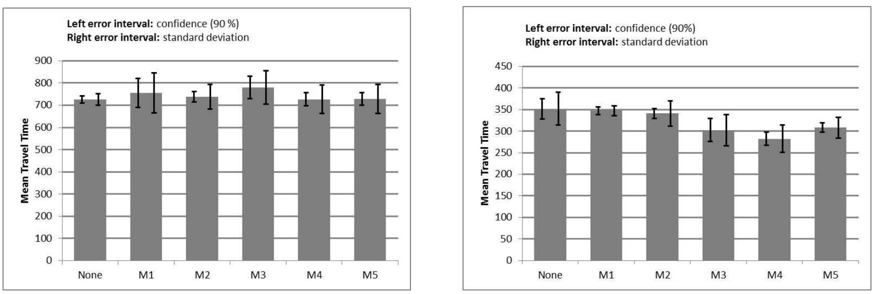

Figure 3: Mean travel time - task 3 and 4.

The overall ranking of the models is shown in Figure 5 (1 is the best, 6 the worst rank). It shows that models 4 and 5 generally produce the shortest travel time. In contrast to the remaining models, models 4 and 5 allow the scheduler to have manually modified schedules completed by automatic optimization. For this purpose it turns out to be irrelevant, whether group optimization (model 5) or full optimization in conjunction with fixation (model 4) is used. The results prove, that manual and automatic features should be used interactively rather than separately from one another.

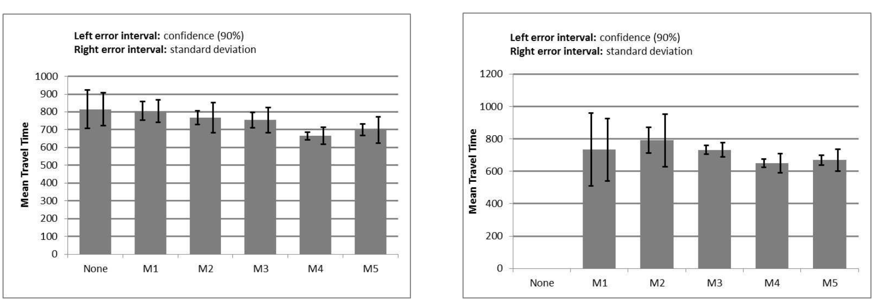

Figure 4: Mean travel time - task 5 and 6.

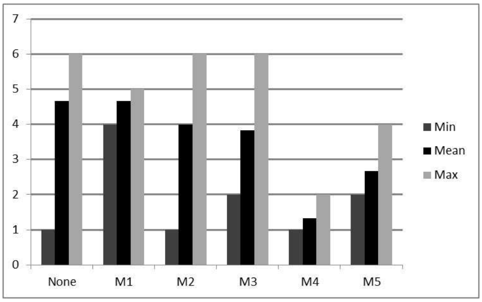

Figure 5: Ranking of the models averaged over the tasks.

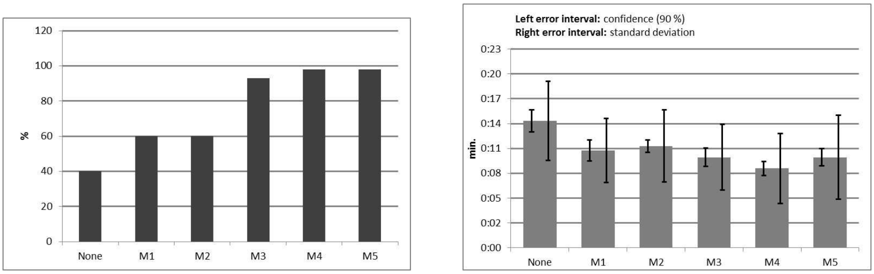

Figure 6: Left: rate of successful task completion, right: average task duration.

4.2 Usability Metric 2: Task Success

The number of participants that have managed to obtain a solution is shown in Figure 6. A task was considered successful, if the schedule did not violate any time window or resource constraints and the scheduling preferences were fulfilled.
With models 1, 2 and ,,None" many participants ran into dead-ends, where they were not able to insert further jobs in the clipboard. In this case model 2 merely depicted a grey Gantt chart background. They would have to manually backtrack former decisions. However, testers would rather give up at this point.

4.3 Usability Metric 3: Task Duration

The average time users required to solve the tasks (deadline was 15 minutes) is shown in Figure 6. Although the time needed with no model is particularly high, in general the models have a high variance in their execution time. How much time a test person spent to fulfill a task was strongly dependent on his motivation and ideas to improve the schedule. The runtime of the system to solve the scheduling problem was negligible.

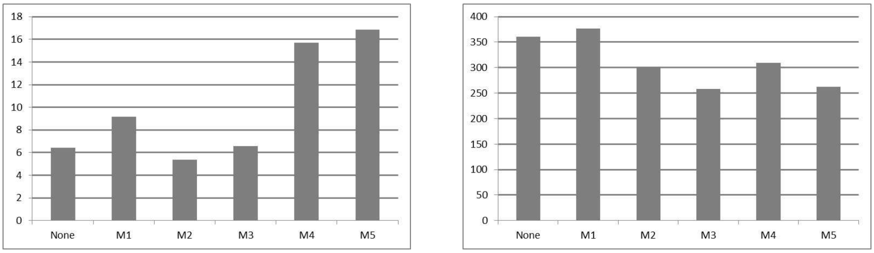

Figure 7: Left: average number of undo operations, right: average number of manual operations.

4.4 Usability Metric 4: Interaction Frequency

Figure 7 (left) shows the number of undo operations averaged over the number of participants. Models 4 and 5 have a strikingly high occurrence of undo, which refers to the general behaviour in the design phase, if there are high-level scheduling features. It consists of alternately applying and reversing automated scheduling features until a satisfying solution is found. Model 1 has a small peak in undo-operations, as there is no aid to predict if an operation will be feasible. Model 2 compensates for this with the color grey.
Figure 7 (right) shows the average number of manual operations (drag and drop of jobs). As expected the manual effort is the higher, the less support is provided. However, manual scheduling is not completely replaced by automated features. In the tests it was required in order to fulfill the schedule specifications given in the test tasks. For this the subjects either modified the optimized schedules subsequently or they created and fixed parts of a schedule before each optimization step.

5 Conclusions

We proposed 8 decision support features to enhance human interaction in scheduling. These features were evaluated in a quantitative study (usability test) with regard to 4 relevant metrics. The results are:

1. The practicability of resulting schedules improves with features to manually fixate, reorder and optimize groups of jobs. In our case, practicability means a cost-optimal integration of expert decisions.

2. Both success rate (solved tasks in given time) and quality (travel time) are highly influenced by the availability of automated scheduling features.

3. Automated scheduling features encourage the user to explore his scope of action on the basis of trial and error (optimize - undo).

References

[BBIM96] Mark H. Burstein, Bolt Beranek, Newman Inc, and Drew V. Mcdermott. Issues in the development of human-computer mixed-initiative planning. In *Cognitive Technology*, pages 285–303. Elsevier, 1996.

[DR11] Steve Dillan and Georg Ringwelski. Using Constraint Propagation to Improve the Usability of an Intelligent Decision Support System. In *Proc. International Workshop on Innovative Scheduling and other Applications using CP-AI-OR (ISA 2011)*. Computer Science Report 01/11 of Brandenburg University of Technology Cottbus, 2011.

[FWW10] Jan C. Fransoo, Toni Waefler, and John R. Wilson, editors. *Behavioral Operations in Planning and Scheduling*, chapter The Unsung Contribution of Production Planners and Schedulers at Production and Sales Interfaces. Springer, 2010.

[JWM04] Sarah Jackson, John R. Wilson, and Bart L. MacCarthy. A New Model of Scheduling in Manufacturing: Tasks, Roles, and Monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46:533, 2004.

[KLMS05] B. Kallehauge, J. Larsen, O. Madsen, and M. Solomon. *Column Generation*, chapter Vehicle Routing Problem with Time Windows, pages 67–98. Springer US, 2005.

[MT06] Katsumi Morikawa and Katsuhiko Takahashi. Modeling Planning and Scheduling Tasks Based on Interviews. In *Proceedings of the 7th Asia Pacific Industrial Engineering and Management Systems Conference*, 2006.

[WJM06] Wout Van Wezel, Rene Jorna, and Alexander Meystel, editors. *Planning in Intelligent Systems*, chapter Planner-Oriented Design of Algorithms for Train Shunting Scheduling. Wiley-Interscience, 2006.