

Scheduling

Nicole Megow^{*1}, Benjamin J. Moseley^{*2}, David Shmoys^{*3},
Ola Svensson^{*4}, Sergei Vassilvitskii^{*5}, and Jens Schlöter^{†6}

- 1 Universität Bremen, DE. nicole.megow@uni-bremen.de
- 2 Carnegie Mellon University – Pittsburgh, US. moseleyb@andrew.cmu.edu
- 3 Cornell University – Ithaca, US. david.shmoys@cornell.edu
- 4 EPFL – Lausanne, CH. ola.svensson@epfl.ch
- 5 Google – New York, US. sergeiv@gmail.com
- 6 Universität Bremen, DE. jschloet@uni-bremen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 23061 “Scheduling”. The seminar focused on the emerging models for beyond-worst case algorithm design, in particular, recent approaches that incorporate learning. This includes models for the integration of learning into algorithm design that have been proposed recently and that have already demonstrated advances in the state-of-art for various scheduling applications: (i) scheduling with error-prone learned predictions, (ii) data-driven algorithm design, and (iii) stochastic and Bayesian learning in scheduling.

Seminar February 5–10, 2023 – <https://www.dagstuhl.de/23061>

2012 ACM Subject Classification Theory of computation → Scheduling algorithms; Mathematics of computing → combinatorial optimization; Theory of computation → Approximation algorithms analysis

Keywords and phrases scheduling, mathematical optimization, approximation algorithms, learning methods, uncertainty

Digital Object Identifier 10.4230/DagRep.13.2.1

1 Executive Summary

Nicole Megow (Universität Bremen, DE)

Benjamin J. Moseley (Carnegie Mellon University – Pittsburgh, US)

David Shmoys (Cornell University – Ithaca, US)

Ola Svensson (EPFL – Lausanne, CH)

Sergei Vassilvitskii (Google – New York, US)

License  Creative Commons BY 4.0 International license
© Nicole Megow, Benjamin J. Moseley, David Shmoys, Ola Svensson, and Sergei Vassilvitskii

This Dagstuhl Seminar was the seventh in a series of Dagstuhl “Scheduling” seminars (since 2008). Scheduling is a major research field that is studied from a practical and theoretical perspective in computer science, mathematical optimization, and operations research. Applications range from traditional production scheduling and project planning to the newly arising resource management tasks in the advent of internet technology and shared resources. Despite the remarkable progress on algorithmic theory for fundamental scheduling problems, new questions gain greater prominence due to the rise of new applications and new methodologies.

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Scheduling, *Dagstuhl Reports*, Vol. 13, Issue 2, pp. 1–19

Editors: Nicole Megow, Benjamin J. Moseley, David Shmoys, Ola Svensson, Sergei Vassilvitskii, and Jens Schlöter



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

At this meeting, we focused on the emerging models for **beyond-worst case analysis**, especially for models that incorporate learning. Decades of research have produced algorithmic discoveries along with impossibility results for scheduling. The primary directions have been focused on achieving the best performance in worst-case scenarios, in terms of run time, memory usage, and proximity to the optimum. Unfortunately, it has often been observed that such algorithms do not necessarily perform the best in practice. Closing the gap between theory and practice is often impossible using the standard approach due to lower bound barriers on worst-case instances.

To bridge the gap between theory and practice, there has been an effort to understand algorithms through a new lens, called beyond-worst-case algorithm analysis. Initial work in this area has produced exciting algorithmic insights and has given convincing explanations for why some algorithms work well in practice. Moreover, the area has led to interesting new algorithmic insights and theoretical directions.

Several of the emerging beyond-worst-case analysis models incorporate learning. Developing the understanding of the models have the promise of discovering new algorithmic insights into scheduling problems and improved performance. This seminar focused on the following three themes.

- *Scheduling with Error-Prone Learned Predictions.* This theme focuses on the settings where the algorithm has access to (e.g. machine-learned) predictions. In this model, an algorithm is given access to a prediction about the problem instance. The performance of the algorithm is parameterized by the quality of this prediction. Typically, an algorithm is given access to an error-prone prediction as machine learning is often imperfect. This prediction can then be leveraged to make algorithmic decisions. Ideally, (1) the predictions result in better performance than the best worst-case bound; (2) the algorithm never performs asymptotically worse than the best worst-case algorithm even if the prediction error is large; and (3) in-between, there is a graceful degradation in performance as the error in the prediction becomes worse. For example, the competitive ratio, approximation ratio or running time can be parameterized by the prediction error.
- *Data-Driven Algorithm Design.* Practitioners often do not use algorithms with the best worst-case performance guarantees directly. Rather, they investigate several algorithms and optimize over them to find the one that works the best for the given application. That is, input data is used to determine the best algorithm (empirically) to use for a given application. Motivated by this, recent research seeks to give a theoretical foundation for this kind of data-driven algorithm selection. In this setting, a family or class of algorithms are defined for a problem. This family is constructed by the algorithm designer. The designer has access to prior instances of the problem that are drawn from a fixed but unknown probability distribution over instances. The goal is to choose an algorithm from the family that will have the best expected performance. That is, to learn from prior instances and decide on the best algorithm for future instances. This model is closely related to practice. Indeed, it is often the case that a scheduling system will have access to job instances from prior days when making scheduling decisions on the current day.
- *Stochastic and Bayesian Learning in Scheduling.* Similar to the learning-augmented setting, we wish to understand how to best incorporate uncertain knowledge about the future in algorithm design. The main difference lies in that, while the learning-augmented framework makes no assumptions on the quality of predictions, the stochastic setting assumes that the data is drawn from a probability distribution. The distribution can either be known upfront or can be learned over time. Questions of interest include:

- If we know the distribution of data, how can we best use it in the design of algorithms?
- If the distribution of data is learned over time, how can we best make decisions while being adaptive to new knowledge of the data?

The scheduling community has recently made striking contributions to both these questions. When the data distribution is known up front – distribution of processing times in this case – there has been major progress on algorithms for scheduling jobs on multiple machines. This is an interesting practical and theoretical question as the processing time of one job may affect the completion of another. An interesting extension to the above question is when the algorithm is allowed some *adaptivity* in a second stage: after computing a schedule based on the distributional knowledge as above, the real processing times are revealed and the algorithm can adapt by making small changes to the schedule. In some scheduling settings, it is known that near-optimal scheduling algorithms can be efficiently computed under the assumption that jobs arrive according to a well-understood probability distribution, such as a Poisson. In contrast, often in the worst case, there are strong lower bounds showing all online algorithms must be far from optimal. The goal is to achieve better performance under the assumption that jobs arrive according to a known probability distribution.

Organization of the Seminar. The seminar brought together 43 researchers from theoretical computer science, mathematical optimization, and operations research. The participants consisted of both senior and junior researchers, including a number of postdocs and advanced PhD students.

During the five days of the seminar, 29 talks of different lengths took place. Six keynote speakers gave an overview of the state-of-the art of the respective area resp. presented recent highlight results in 60 minutes:

- Siddhartha Banerjee: We Need to Talk About How we Talk About Online Decision-Making
- Sami Davies: Scheduling with Communication Delays
- Silvio Lattanzi: Clustering with Advice
- Claire Mathieu: Stable matching in Practice
- Debmalya Panigrahi: Santa Claus with Predictions
- Sergei Vassilvitskii: Five(ish) Years of Algorithms with Predictions.

The remaining slots were filled with shorter talks of 30 minutes on various topics related to scheduling, resource allocation, and related problems, from the perspective of coping with uncertainty, learning, and applications in practice.

Further, in the beginning of the week, open problem sessions were held. One of the open problem sessions was devoted to problems that would have been of interest to Gerhard Woeginger, pivotal member of the community that passed away in 2022. Throughout the week, a few sessions with spotlight talks of 8 minutes gave participants the chance to announce recent results and invite for discussions. The schedule left ample free time that was actively used for fruitful discussions and joint research.

Outcome. Organizers and participants regard the seminar as a great success. The seminar achieved the goal to bring together the related communities, share the state-of-the art research and discuss the current major challenges. The talks were excellent and stimulating; participants actively met in working groups in the afternoon and evenings. It was remarked positively that a significant number of younger researchers (postdocs and PhD students) participated and integrated well.

Acknowledgements. The organizers wish to express their gratitude towards the Scientific Directorate and the administration of the Dagstuhl Center for their great support for this seminar.

2 Table of Contents

Executive Summary

<i>Nicole Megow, Benjamin J. Moseley, David Shmoys, Ola Svensson, and Sergei Vassilvitskii</i>	1
--	---

Overview of Talks

Paging with Succinct Predictions <i>Antonios Antoniadis</i>	6
Flow Time Scheduling with Uncertain Processing Time <i>Yossi Azar</i>	6
Better Trees for Santa Claus <i>Étienne Bamas</i>	7
We Need to Talk About How we Talk About Online Decision-Making <i>Siddhartha Banerjee</i>	7
Scheduling with communication delays <i>Sami Davies</i>	8
Stochastic Configuration Balancing <i>Franziska Eberle</i>	8
Decentralized Multi-agent Learning in Bipartite Queuing Systems <i>Daniel Freund</i>	9
Online Matching with Reusable Capacities <i>Vineet Goyal</i>	9
Online Load Balancing Beyond ℓ_p Norms <i>Thomas Kesselheim, Marco Molinaro, and Sahil Singla</i>	10
Online Algorithms with Multiple Predictions <i>Amit Kumar</i>	10
Clustering with Advice <i>Silvio Lattanzi</i>	11
Algorithms with Prediction Portfolios <i>Thomas Lavastida</i>	11
Predictions for Online Min-Sum Scheduling Problems <i>Alex Lindermayr</i>	12
DAG Scheduling Problems <i>Alberto Marchetti-Spaccamela</i>	12
Stable matching in practice <i>Claire Mathieu</i>	12
Online Rounding of Bipartite Matchings <i>Seffi Naor</i>	13
Learning-augmented Assignment: Santa Claus does Load Balancing <i>Debmalya Panigrahi</i>	13
Evaluating Stochastic Score Functions <i>Kevin Schewior</i>	14

Learning-Augmented Query Policies for Minimum Spanning Tree with Uncertainty
Jens Schlöter 14

New bounds for approximation algorithms for graph burning
Jiri Sgall 15

Hiking through the complexity landscape with Gerhard – a personal reminiscence
Martin Skutella 15

Stochastic Minimum-Norm Combinatorial Optimization
Chaitanya Swamy 15

Fast Graph Algorithms with Learned Duals
Ali Vakilian 16

Five(ish) Years of Algorithms with Predictions
Sergei Vassilvitskii 17

Search infected nodes in uncertain graphs
Jose Verschae 17

Bayesian Scheduling: Analysis of simple policies
Tjark Vredeveld 17

Minimizing Completion Times for Stochastic Jobs via Batched Free Times
Rudy Zhou 18

Participants 19

3 Overview of Talks

3.1 Paging with Succinct Predictions

Antonios Antoniadis (University of Twente, NL)

License © Creative Commons BY 4.0 International license
© Antonios Antoniadis

Joint work of Antonios Antoniadis, Joan Boyar, Marek Eliás, Lene M. Favrholdt, Ruben Hoeksma, Kim S. Larsen, Adam Polak, Bertrand Simon

Main reference Antonios Antoniadis, Joan Boyar, Marek Eliás, Lene M. Favrholdt, Ruben Hoeksma, Kim S. Larsen, Adam Polak, Bertrand Simon: “Paging with Succinct Predictions”, CoRR, Vol. abs/2210.02775, 2022.

URL <https://doi.org/10.48550/arXiv.2210.02775>

We study learning-augmented paging from the new perspective of requiring the least possible amount of predicted information. More specifically, the predictions obtained alongside each page request are limited to one bit only. We consider two natural such setups: (i) discard predictions, in which the predicted bit denotes whether or not it is “safe” to evict this page, and (ii) phase predictions, where the bit denotes whether the current page will be requested in the next phase (for an appropriate partitioning of the input into phases). We develop learning-augmented algorithms for each of the two setups and establish that our algorithms are essentially best possible. We believe that succinct predictions are of interest for other problems beyond paging.

3.2 Flow Time Scheduling with Uncertain Processing Time

Yossi Azar (Tel Aviv University, IL)

License © Creative Commons BY 4.0 International license
© Yossi Azar

Joint work of Yossi Azar, Stefano Leonardi, Noam Touitou

Main reference Yossi Azar, Stefano Leonardi, Noam Touitou: “Flow Time Scheduling with Uncertain Processing Time”, CoRR, Vol. abs/2103.05604, 2021.

URL <https://arxiv.org/abs/2103.05604>

We consider the problem of online scheduling on a single machine to minimize unweighted and weighted flow time. The existing algorithms for these problems require exact knowledge of the processing time of each job. This assumption is crucial, as even a slight perturbation of the processing time would lead to polynomial competitive ratio. However, this assumption very rarely holds in real-life scenarios. We present a competitive algorithm (the competitive ratio is a function of the distortion) for unweighted flow time that does not require knowledge of the distortion in advance. For the weighted flow time we present competitive algorithms but, in this case, we need to know (an upper bound on) the distortion in advance. This is joint work with Stefano Leonardi and Noam Touitou based on papers that appeared in STOC 21 and SODA 2022.

3.3 Better Trees for Santa Claus

Etienne Bamas (EPFL – Lausanne, CH)

License © Creative Commons BY 4.0 International license

© Étienne Bamas

Joint work of Étienne Bamas, Lars Rohwedder

Main reference Étienne Bamas, Lars Rohwedder: “Better Trees for Santa Claus”, CoRR, Vol. abs/2211.14259, 2022.

URL <https://doi.org/10.48550/arXiv.2211.14259>

A notorious open problem in approximation algorithms is whether there exists a constant factor approximation for MaxMin Fair Allocation of indivisible items (also known as the Santa Claus problem). Bateni, Charikar, and Guruswami [STOC’09] introduced the MaxMin Arborescence problem as an important special case: Given a directed graph with sources and sinks we have to find vertex disjoint arborescences rooted in the sources such that at each non-sink vertex of an arborescence the out-degree is at least k , where k is to be maximized. This problem is of particular interest, since it appears to capture much of the difficulty of the general Santa Claus problem. Indeed, the progress made by Bateni et al. was quickly generalized by Chakrabarty, Chuzhoy, and Khanna [FOCS’09] to the general case. These two results remain the state-of-the-art for both problems, and they yield a polylogarithmic approximation in quasi-polynomial time.

In this talk, I will present the main ideas behind an exponential improvement to this, a $\text{poly}(\log \log n)$ -approximation in quasi-polynomial time for the MaxMin Arborescence problem.

3.4 We Need to Talk About How we Talk About Online Decision-Making

Siddhartha Banerjee (Cornell University – Ithaca, US)

License © Creative Commons BY 4.0 International license

© Siddhartha Banerjee

Scheduling, and other problems involving online allocation of resources, are topics of great interest across many academic communities. However, the huge diversity in underlying models and methodologies means that existing assumptions/algorithms/guarantees are difficult to understand and compare (and often not very useful...).

I will try to present the stochastic control viewpoint on these problems, and discuss how there seems to be a fundamental divide between the view of online algorithms in CS and in controls. I will then present a sample-path coupling technique, which provides a simple way of reasoning about online algorithms, and regret guarantees against any chosen benchmark. I will describe how this framework gives new algorithms and insights for a variety of problems, including (time permitting):

1. *Constant regret* algorithms (i.e., having additive loss compared to the hindsight optimal solution which is *independent of the horizon and budget*) for several widely-studied settings including online packing, load balancing, dynamic pricing, assortment optimization, and online bin packing.
2. Incorporating side information and historical data in these settings (and achieve constant regret with as little as a single data trace).
3. Fundamental tradeoffs in multi-objective settings (in particular, for fairness in online allocation).

3.5 Scheduling with communication delays

Sami Davies (Northwestern University – Evanston, US)

License  Creative Commons BY 4.0 International license
 Sami Davies

Joint work of Sami Davies, Janardhan Kulkarni, Thomas Rothvoss, Sai Sandeep, Jakub Tarnawski, Yiaho Zhang

I'll discuss progress on scheduling with communication delays. In this setting, if two dependent jobs are scheduled on different machines, a delay must pass between their execution times. The question of whether constant factor approximation algorithms exist in this setting was one of the biggest open problems in scheduling theory. We effectively answered this question by (1) finding polylog approximation algorithms when the delay is uniform between dependent jobs and (2) proving super-constant hardness when the delay is non-uniform between dependent jobs.

3.6 Stochastic Configuration Balancing

Franziska Eberle (London School of Economics, GB)

License  Creative Commons BY 4.0 International license
 Franziska Eberle

Joint work of Franziska Eberle, Anupam Gupta, Nicole Megow, Benjamin Moseley, Rudy Zhou

Main reference Franziska Eberle, Anupam Gupta, Nicole Megow, Benjamin Moseley, Rudy Zhou: “Configuration Balancing for Stochastic Requests”, CoRR, Vol. abs/2208.13702, 2022.

URL <https://doi.org/10.48550/arXiv.2208.13702>

The configuration balancing problem with stochastic requests generalizes many well-studied resource allocation problems such as load balancing and virtual circuit routing. In it, we have m resources and n requests. Each request has multiple possible configurations, each of which increases the load of each resource by some amount. The goal is to select one configuration for each request to minimize the makespan: the load of the most-loaded resource. In our work, we focus on a stochastic setting, where we only know the distribution for how each configuration increases the resource loads, learning the realized value only after a configuration is chosen. We develop both offline and online algorithms for configuration balancing with stochastic requests. When the requests are known offline, we give a non-adaptive policy for configuration balancing with stochastic requests that $O(\frac{\log m}{\log \log m})$ -approximates the optimal adaptive policy. In particular, this closes the adaptivity gap for this problem as there is an asymptotically matching lower bound even for the very special case of load balancing on identical machines. When requests arrive online in a list, we give a non-adaptive policy that is $O(\log m)$ competitive. Again, this result is asymptotically tight due to information-theoretic lower bounds for very special cases (e.g., for load balancing on unrelated machines). Finally, we show how to leverage adaptivity in the special case of load balancing on related machines to obtain a constant-factor approximation offline and an $O(\log \log m)$ -approximation online. A crucial technical ingredient in all of our results is a new structural characterization of the optimal adaptive policy that allows us to limit the correlations between its decisions.

3.7 Decentralized Multi-agent Learning in Bipartite Queuing Systems

Daniel Freund (MIT – Cambridge, US)

License © Creative Commons BY 4.0 International license
© Daniel Freund

Joint work of Daniel Freund, Thodoris Lykouris, Wentao Weng

Main reference Daniel Freund, Thodoris Lykouris, Wentao Weng: “Efficient decentralized multi-agent learning in asymmetric queuing systems”, CoRR, Vol. abs/2206.03324, 2022.

URL <https://doi.org/10.48550/arXiv.2206.03324>

Learning in multi-agent systems often poses significant challenges due to interference between agents. In particular, unlike classical stochastic systems, the performance of an agent’s action is not drawn i.i.d. from some distribution but is directly affected by the (unobserved) actions of the other agents. This is the reason why most collaborative multi-agent learning approaches aim to globally coordinate all agents’ actions to evade this interference.

In this talk, we focus on agents in a decentralized bipartite queuing system, where N agents request service from K servers. Prior decentralized approaches aim to globally identify a coordinated schedule or do not take advantage of the bipartite structure, which leads to significant shortcomings: performance that degrades exponentially in the number of servers, requirement of shared randomness and unique identifiers, and computationally demanding algorithms. In contrast, we provide a low-complexity algorithm that is run decentrally by each agent, avoids the shortcomings of “global coordination” and leads the queuing system to have efficient performance in asymmetric bipartite queuing systems while also having additional robustness properties.

3.8 Online Matching with Reusable Capacities

Vineet Goyal (Columbia University – New York, US)

License © Creative Commons BY 4.0 International license
© Vineet Goyal

Joint work of Vineet Goyal, Garud Iyengar, Rajan Udhwani

Main reference Vineet Goyal, Garud Iyengar, Rajan Udhwani: “Asymptotically Optimal Competitive Ratio for Online Allocation of Reusable Resources”. CoRR aVol. bs/2002.02430 (2022)

URL <https://doi.org/10.48550/arXiv.2002.02430>

Online matching problems are at the heart of resource allocation given the inherent demand uncertainty. For instance, in a typical setting users arrive sequentially to a platform and the platform needs to make irrevocable matching decisions in an online manner. We consider fundamental generalizations of the classical variants in order to incorporate some of the natural stochasticity in resource usage that arises in many applications.

This talk mainly focuses on understanding the impact of reusability of resource capacities – a key aspect of resource allocation in applications such as cloud computing and sharing economies. Here allocated resources are required by users for some a priori unknown (stochastic) durations. Resources are returned after use and are available for re-allocation. We propose a new policy that achieves the best possible guarantee of $(1-1/e - o(1))$ under reasonable assumptions. Further, in the process of analyzing this policy we develop a novel framework of analysis that is useful more broadly in other settings with stochastic resource consumption.

3.9 Online Load Balancing Beyond ℓ_p Norms

Thomas Kesselheim (Universität Bonn, DE), Marco Molinaro, and Sahil Singla

License  Creative Commons BY 4.0 International license
 © Thomas Kesselheim, Marco Molinaro, and Sahil Singla

Joint work of Thomas Kesselheim, Marco Molinaro, Sahil Singla

Main reference Thomas Kesselheim, Marco Molinaro, Sahil Singla: “Online and Bandit Algorithms Beyond p Norms”, in Proc. of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023, pp. 1566–1593, SIAM, 2023.

URL <https://doi.org/10.1137/1.9781611977554.ch58>

The classic problem of online makespan minimization can be understood as minimizing the ℓ_∞ norm of the vector of machine loads. It was extended to ℓ_p norms already more than 25 years ago. We study the problem beyond ℓ_p norms.

We show that general norms admit good algorithms as long as the norm can be approximated by a function that is “gradient-stable”, a notion that we introduce. Roughly it says that the gradient of the function should not drastically decrease in any component as we increase the input vector.

In particular, we give the first $O(\log^2 m)$ -competitive algorithm for online load balancing with respect to an arbitrary monotone symmetric norm. Our techniques extend to applications beyond symmetric norms as well, e.g., to Online Vector Scheduling and to Online Generalized Assignment with Convex Costs.

The set of techniques can also be applied in stochastic settings, e.g., in which the sequence of jobs arrives in random order. Given that they can be understood as based on duality, it would be interesting to see if they also can be used with error-prone predictions.

3.10 Online Algorithms with Multiple Predictions

Amit Kumar (Indian Institute of Technology – New Dehli, IN)

License  Creative Commons BY 4.0 International license
 © Amit Kumar

Joint work of Keerti Anand, Rong Ge, Amit Kumar, Debmalya Panigrahi

Main reference Keerti Anand, Rong Ge, Amit Kumar, Debmalya Panigrahi: “Online Algorithms with Multiple Predictions”, CoRR, Vol. abs/2205.03921, 2022.

URL <https://doi.org/10.48550/arXiv.2205.03921>

This paper studies online algorithms augmented with multiple machine-learned predictions. While online algorithms augmented with a single prediction have been extensively studied in recent years, the literature for the multiple predictions setting is sparse. In this paper, we give a generic algorithmic framework for online covering problems with multiple predictions that obtains an online solution that is competitive against the performance of the best predictor. Our algorithm incorporates the use of predictions in the classic potential-based analysis of online algorithms. We apply our algorithmic framework to solve classical problems such as online set cover, (weighted) caching, and online facility location in the multiple predictions setting. Our algorithm can also be robustified, i.e., the algorithm can be simultaneously made competitive against the best prediction and the performance of the best online algorithm (without prediction).

3.11 Clustering with Advice

Silvio Lattanzi (Google – Barcelona, ES)

License © Creative Commons BY 4.0 International license
© Silvio Lattanzi

In this talk, starting from practical questions, we motivate different models of machine learning advice and present new algorithms that leverage the additional information to obtain stronger guarantees. In particular, we start by describing the semi-supervised active clustering framework and how one can recover convex and non-convex clusters in this setting. Then we describe how partial knowledge about input instances can be leveraged to obtain better guarantees in online correlation clustering and the classic k-means problem.

3.12 Algorithms with Prediction Portfolios

Thomas Lavastida (University of Texas – Dallas, US)

License © Creative Commons BY 4.0 International license
© Thomas Lavastida

Joint work of Michael Dinitz, Sungjin Im, Thomas Lavastida, Benjamin Moseley, Sergei Vassilvitskii

Main reference Michael Dinitz, Sungjin Im, Thomas Lavastida, Benjamin Moseley, Sergei Vassilvitskii: “Algorithms with Prediction Portfolios”, in Proc. of the Advances in Neural Information Processing Systems, 2022.

URL https://openreview.net/forum?id=wwWCZ7sER_C

The research area of algorithms with predictions has seen recent success showing how to incorporate machine learning into algorithm design to improve performance when the predictions are correct, while retaining worst-case guarantees when they are not. Most previous work has assumed that the algorithm has access to a single predictor. However, in practice, there are many machine learning methods available, often with incomparable generalization guarantees, making it hard to pick a best method a priori. In this work we consider scenarios where multiple predictors are available to the algorithm and the question is how to best utilize them.

Ideally, we would like the algorithm’s performance to depend on the quality of the best predictor. However, utilizing more predictions comes with a cost, since we now have to identify which prediction is the best. We study the use of multiple predictors for a number of fundamental problems, including matching, load balancing, and non-clairvoyant scheduling, which have been well-studied in the single predictor setting. For each of these problems we introduce new algorithms that take advantage of multiple predictors, and prove bounds on the resulting performance.

3.13 Predictions for Online Min-Sum Scheduling Problems

Alex Lindermayr (Universität Bremen, DE)

License © Creative Commons BY 4.0 International license
© Alex Lindermayr

- Joint work of** Alexandra Lassota, Nicole Megow, Martin Rapp, Jens Schlöter
- Main reference** Alexandra Lassota, Alexander Lindermayr, Nicole Megow, Jens Schlöter: “Minimalistic Predictions to Schedule Jobs with Online Precedence Constraints”, CoRR, Vol. abs/2301.12863, 2023.
- URL** <https://doi.org/10.48550/arXiv.2301.12863>
- Main reference** Alexander Lindermayr, Nicole Megow: “Permutation Predictions for Non-Clairvoyant Scheduling”, in Proc. of the SPAA '22: 34th ACM Symposium on Parallelism in Algorithms and Architectures, Philadelphia, PA, USA, July 11 – 14, 2022, pp. 357–368, ACM, 2022.
- URL** <https://doi.org/10.1145/3490148.3538579>
- Main reference** Alexander Lindermayr, Nicole Megow, Martin Rapp: “Speed-Oblivious Online Scheduling: Knowing (Precise) Speeds is not Necessary”, CoRR, Vol. abs/2302.00985, 2023.
- URL** <https://doi.org/10.48550/arXiv.2302.00985>

We consider the problem of scheduling jobs to minimize the total weighted completion time, and present models for various uncertainties. For uncertain processing requirements, also called non-clairvoyant scheduling, we present and recall predictions and results to improve the lower bound of 2 on the competitive ratio on a single machine. For online precedence constraints, we present bounds for different prediction models which try to overcome a simple $\Omega(n)$ lower bound in the worst-case setting. For uncertain processing speeds of unrelated machines, lower bounds rule out competitive ratios better than $\Omega(m)$ for m machines. Here we present two models and algorithms, which overcome this lower bound: speed predictions give predictions on the uncertain speeds, and speed-ordered machines ensure a single machine order which all job-dependent speeds respect, i.e., for any job j , the speed of j on machine i no less than on machine i' if and only if i comes before i' in the order.

3.14 DAG Scheduling Problems

Alberto Marchetti-Spaccamela (Sapienza University of Rome, IT)

License © Creative Commons BY 4.0 International license
© Alberto Marchetti-Spaccamela

- Joint work of** Sanjoy Baruah, Vincenzo Bonifaci, Alberto Marchetti-Spaccamela, Nicole Megow, Jens Schlöter, Martin Skutella, Sebastian Stiller, Leen Stougie, Andreas Wiese

The Directed Acyclic Graph (DAGs) is a popular representation to describe the structure of parallel applications and to model the execution of multi-threaded programs that is widely used in cloud computing and in real-time systems. In this talk I review recent results on DAG scheduling considering different models and focusing on complexity and approximation.

3.15 Stable matching in practice

Claire Mathieu (CNRS – Paris, FR)

License © Creative Commons BY 4.0 International license
© Claire Mathieu

Stable matching methods, based on the algorithm designed by Gale and Shapley, are used around the world in many applications. the college assigned to the applicant in their preference list; robustness; running time; etc.

After a brief theoretical review, we present issues arising in practice, (1) in the context of college admissions in France since 2018, and (2) in the context of upcoming medical studies specialization in France starting in 2024.

3.16 Online Rounding of Bipartite Matchings

Seffi Naor (Technion – Haifa, IL)

License © Creative Commons BY 4.0 International license
© Seffi Naor

Joint work of Niv Buchbinder, Seffi Naor, Aravind Srinivasan, David Wajc

Main reference Niv Buchbinder, Joseph Naor, David Wajc: “A Randomness Threshold for Online Bipartite Matching, via Lossless Online Rounding”, CoRR, Vol. abs/2106.04863, 2021.

URL <https://doi.org/10.48550/arXiv.2106.04863>

Two complementary facets of the online bipartite matching problem are discussed. (1) For numerous online bipartite matching problems, such as edge-weighted matching and matching under two-sided vertex arrivals, state-of-the-art fractional algorithms outperform their randomized integral counterparts. Thus, a natural question is whether we can achieve lossless online rounding of fractional solutions in this setting. Even though lossless online rounding is impossible in general, randomized algorithms do induce fractional algorithms of the same competitive ratio, which by definition are losslessly roundable online. This motivates the addition of constraints that decrease the “online integrality gap”, thus allowing for lossless online rounding. We characterize a set of non-convex constraints which allow for such lossless online rounding and allow for better competitive ratios than yielded by deterministic algorithms. (2) In a different vein, we study the problem of rounding fractional bipartite matchings in online settings. We assume that a fractional solution is already generated for us online by a black box (via a fractional algorithm, or some machine-learned advice) and provided as part of the input, which we then wish to round. We provide improved bounds on the rounding ratio and discuss several applications.

3.17 Learning-augmented Assignment: Santa Claus does Load Balancing

Debmalya Panigrahi (Duke University – Durham, US)

License © Creative Commons BY 4.0 International license
© Debmalya Panigrahi

Assignment problems are among the most well-studied in online algorithms. In these problems, a sequence of items arriving online must be assigned among a set of agents so as to optimize a given objective. This encompasses scheduling problems for minimizing makespan, p-norms, and other objectives, as well as fair division problems such as the Santa Claus problem and Nash welfare maximization. One common feature is that many of these problems are characterized by strong worst-case lower bounds in the online setting. To circumvent these impossibility results, recent research has focused on using additional (learned) information about the problem instance and this has led to dramatic improvements in the competitive ratio over the worst case. In this talk, I will first survey some of this literature (Lattanzi et al., SODA 20; Li and Xian, ICML 21; Banerjee et al., SODA 22; Barman et al., AAAI 22) that addresses specific problems in this domain. I will then proceed to describe recent work

with Ilan Cohen that brings these problems under one umbrella: we give a single algorithmic framework for learning-augmented online assignment for a large class of maximization and minimization objectives.

3.18 Evaluating Stochastic Score Functions

Kevin Schewior (University of Southern Denmark – Odense, DK)

License © Creative Commons BY 4.0 International license
© Kevin Schewior

Joint work of Kevin Schewior, Benedikt M. Plank

Main reference Benedikt M. Plank, Kevin Schewior: “Simple Algorithms for Stochastic Score Classification with Small Approximation Ratios”, CoRR, Vol. abs/2211.14082, 2022.

URL <https://doi.org/10.48550/arXiv.2211.14082>

We revisit the Stochastic Score Classification (SSC) problem introduced by Gkenosis et al. (ESA 2018): There are n tests. Each test j can be conducted at cost c_j , and it succeeds independently with probability p_j . Further, a partition of the (integer) interval $\{0, \dots, n\}$ into a number of smaller intervals is known. The goal is to conduct tests so as to determine that interval from the partition in which the number of successful tests lies while minimizing the expected cost. Ghuge et al. (IPCO 2022) independently showed that a polynomial-time constant-factor approximation algorithm exists. We present a simple polynomial-time $(3 + 2 * \sqrt{2})$ -approximation algorithm and highlight (potential) connections to scheduling problems. This is joint work with Benedikt Plank.

3.19 Learning-Augmented Query Policies for Minimum Spanning Tree with Uncertainty

Jens Schlöter (Universität Bremen, DE)

License © Creative Commons BY 4.0 International license
© Jens Schlöter

Joint work of Thomas Erlebach, Murilo Santos de Lima, Nicole Megow, Jens Schlöter

Main reference Thomas Erlebach, Murilo Santos de Lima, Nicole Megow, Jens Schlöter: “Learning-Augmented Query Policies for Minimum Spanning Tree with Uncertainty”, in Proc. of the 30th Annual European Symposium on Algorithms, ESA 2022, September 5-9, 2022, Berlin/Potsdam, Germany, LIPIcs, Vol. 244, pp. 49:1–49:18, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

URL <https://doi.org/10.4230/LIPIcs.ESA.2022.49>

We study how to utilize (possibly erroneous) predictions in a model for computing under uncertainty in which an algorithm can query unknown data. Our aim is to minimize the number of queries needed to solve the minimum spanning tree problem, a fundamental combinatorial optimization problem that has been central also to the research area of explorable uncertainty. For all integral $\gamma \geq 2$, we present algorithms that are γ -robust and $(1 + \frac{1}{\gamma})$ -consistent, meaning that they use at most γOPT queries if the predictions are arbitrarily wrong and at most $(1 + \frac{1}{\gamma})OPT$ queries if the predictions are correct, where OPT is the optimal number of queries for the given instance. Moreover, we show that this trade-off is best possible. Furthermore, we argue that a suitably defined hop distance is a useful measure for the amount of prediction error and design algorithms with performance guarantees that degrade smoothly with the hop distance. Our results demonstrate that untrusted predictions can circumvent the known lower bound of 2, without any degradation of the worst-case ratio. To obtain our results, we provide new structural insights for the

minimum spanning tree problem that might be useful in the context of query-based algorithms regardless of predictions. In particular, we generalize the concept of witness sets – the key to lower-bounding the optimum – by proposing novel global witness set structures and completely new ways of adaptively using those.

3.20 New bounds for approximation algorithms for graph burning

Jiri Sgall (Charles University – Prague, CZ)

License © Creative Commons BY 4.0 International license
© Jiri Sgall

Joint work of Jiri Sgall, Matej Lieskovsky, Andreas E. Feldmann

Graph burning is the following process: We start with a graph with no node burned. At time $t=1$ we choose a node and burn it. At each time step $t > 1$, all neighbours of already burned nodes are also burned, and we choose one arbitrary additional node to burn. The process stops when all nodes of the graph have been burned. The burning number is the minimal number of steps needed for all nodes of a graph to be burned.

We sketch a randomized 2.32-approximation algorithm and a lower bound of $4/3$ for burning number of arbitrary graphs. This improves the previous upper bound of 3, for lower bound only NP-hardness was known.

3.21 Hiking through the complexity landscape with Gerhard – a personal reminiscence

Martin Skutella (TU Berlin, DE)

License © Creative Commons BY 4.0 International license
© Martin Skutella

This talk is dedicated to the memory of my friend and esteemed colleague Gerhard Woeginger, in thankful admiration for his outstanding scientific contributions, his fine taste of problems, his unfailingly inspiring lectures, his valuable advice over the course of 25 years, and, last but not least, for his great sense of humor.

3.22 Stochastic Minimum-Norm Combinatorial Optimization

Chaitanya Swamy (University of Waterloo, CA)

License © Creative Commons BY 4.0 International license
© Chaitanya Swamy

Joint work of Sharat Irahimpur, Chaitanya Swamy

Main reference Sharat Irahimpur, Chaitanya Swamy: “Approximation Algorithms for Stochastic Minimum-Norm Combinatorial Optimization”, in Proc. of the 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020, pp. 966–977, IEEE, 2020.

URL <https://doi.org/10.1109/FOCS46700.2020.00094>

We develop a framework for designing approximation algorithms for a wide class of (1-stage) stochastic-optimization problems with norm-based objective functions. We introduce the model of stochastic minimum-norm combinatorial optimization, wherein the costs involved are random variables with given distributions, and we are given a monotone, symmetric

norm f . Each feasible solution induces a random multidimensional cost vector whose entries are independent random variables, and the goal is to find a solution that minimizes the expected f -norm of the induced cost vector. This is a very rich class of objectives, containing all ℓ_p norms, as also Top- l norms (sum of l largest coordinates in absolute value), which enjoys various closure properties.

Our chief contribution is a framework for designing approximation algorithms for stochastic minimum-norm optimization, which has two key components: (i) A reduction showing that one can control the expected f -norm by simultaneously controlling a (small) collection of expected Top- l norms; and (ii) Showing how to tackle the minimization of a single expected Top- l -norm by leveraging techniques used to deal with minimizing the expected maximum, circumventing the difficulties posed by the non-separable nature of Top- l norms.

We apply our framework to obtain strong approximation guarantees for two concrete problem settings: (1) stochastic load balancing, wherein jobs have random processing times and the induced cost vector is the machine-load vector; and (2) stochastic spanning tree, where edges have stochastic weights and the cost vector consists of the edge-weight variables of edges in the spanning tree returned.

This is joint work with Sharat Irahimpur.

3.23 Fast Graph Algorithms with Learned Duals

Ali Vakilian (TTIC – Chicago, US)

License © Creative Commons BY 4.0 International license
© Ali Vakilian

Joint work of Justin Y. Chen, Sandeep Silwal, Ali Vakilian, Fred Zhang

Main reference Justin Y. Chen, Sandeep Silwal, Ali Vakilian, Fred Zhang: “Faster Fundamental Graph Algorithms via Learned Predictions”, CoRR, Vol. abs/2204.12055, 2022.

URL <https://doi.org/10.48550/arXiv.2204.12055>

We consider the question of speeding up classic graph algorithms with machine-learned predictions. In this model, algorithms are furnished with extra advice learned from past or similar instances. Given the additional information, we aim to improve upon the traditional worst-case run-time guarantees. Our contributions are the following:

- (i) We give a faster algorithm for minimum-weight bipartite matching via learned duals, improving the recent result by Dinitz, Im, Lavastida, Moseley and Vassilvitskii (NeurIPS, 2021);
- (ii) We extend the learned dual approach to the single-source shortest path problem (with negative edge lengths), achieving an almost linear runtime given sufficiently accurate predictions which improves upon the classic fastest algorithm due to Goldberg (SIAM J. Comput., 1995);
- (iii) We provide a general reduction-based framework for learning-based graph algorithms, leading to new algorithms for degree-constrained subgraph and minimum-cost 0-1 flow, based on reductions to bipartite matching and the shortest path problem.

Finally, we give a set of general learnability theorems, showing that the predictions required by our algorithms can be efficiently learned in a PAC fashion.

3.24 Five(ish) Years of Algorithms with Predictions

Sergei Vassilvitskii (Google – New York, US)

License  Creative Commons BY 4.0 International license
© Sergei Vassilvitskii

Worst-case analysis has proven invaluable for understanding aspects of both the complexity and practicality of algorithms. In some cases, however, we do not face worst-case scenarios, and the question arises of how we can tune our algorithms to work even better on the kinds of instances we are likely to see, while keeping a rigorous formal framework similar to what we have developed through worst-case analysis.

We give an overview of a recent trend that develops algorithms parameterized by additional parameters which capture “the kinds of instances we are likely to see,” and obtains a finer grained analysis of algorithms’ performance. We will give examples of re-analyzing classical algorithms through this lens, as well as developing new algorithms that expose new structural insights about the problems.

3.25 Search infected nodes in uncertain graphs

Jose Verschae (PUC – Santiago de Chile, CL)

License  Creative Commons BY 4.0 International license
© Jose Verschae

Joint work of Jose Verschae, I. S. Beaudry, José Baboun, M. Castro, A. Jara, Benjamín Rubio

In the context of COVID-19, methods for traceability and search of active cases played a vital role. One such method uses PCR samples in the sewage network, aiming to locate the appearance of new infections quickly.

Given a representation of the sewer network as an acyclic directed graph, we seek to design a search strategy that finds new infected nodes while minimizing the number of samples needed in the worst case. This problem has been extensively studied in the context of perfect information, that is, if the graph is known. However, in reality, we find much information missing from the network. In this talk, we will propose a model to address the problem when the network is only partially known and present some partial results. In particular, we give tight bounds on the number of samples needed when the network is a planar graph of bounded degree.

3.26 Bayesian Scheduling: Analysis of simple policies

Tjark Vredeveld (Maastricht Univ. School of Business & Economics, NL)

License  Creative Commons BY 4.0 International license
© Tjark Vredeveld

Joint work of Tjark Vredeveld, Sebastian Marban, Cyriel Rutten

Bayesian scheduling is an extension of stochastic scheduling in which there is uncertainty about the system parameters. By processing jobs, we can learn about the true values of these parameters. We consider the basic scheduling problem of non-preemptively scheduling jobs on a single machine so as to minimize the expected total completion time. We consider a setting in which there are m classes of (a finite number of) jobs and the processing

times of jobs of the same class are independent and identically exponentially distributed with an unknown parameter. The initial beliefs on this parameter are modelled as a prior distribution and after processing a job of the class, the posterior distribution models the updated beliefs on the parameter. For this setting, optimal policies based on Gittins-indices exist (see, e.g., Hamada and Glazebrook, 1993). However, computing these indices may be computationally challenging. In this talk, we consider two simple policies, based on SEPT (Shortest Expected Processing Time), the optimal policy for the variant in which the parameters of the exponential distribution are known. We consider one version of SEPT, where the expected processing time are based on the initial belief and one version that updates the expected value each time a job has been processed. We can show that the first version of SEPT is at most a factor of m away from the optimal policy and that this is tight. Moreover, for the second variant we conjecture that the bound is also tight.

3.27 Minimizing Completion Times for Stochastic Jobs via Batched Free Times

Rudy Zhou (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
© Rudy Zhou

Joint work of Anupam Gupta, Benjamin Moseley, Rudy Zhou

Main reference Anupam Gupta, Benjamin Moseley, Rudy Zhou: “Minimizing Completion Times for Stochastic Jobs via Batched Free Times”, CoRR, Vol. abs/2208.13696, 2022.

URL <https://doi.org/10.48550/arXiv.2208.13696>

In this talk, we consider the classic problem of minimizing the expected total completion time of jobs on m identical machines in the setting where the sizes of the jobs are stochastic. Specifically, the size of each job is a random variable whose distribution is known to the algorithm, but whose realization is revealed only after the job is scheduled. We give a $O(m^{1/2}poly(\log n))$ -approximation for stochastic jobs which have Bernoulli processing times. This is the first approximation for this problem that is both independent of the variance in the job sizes, and is sublinear in the number of machines m . Our algorithm is based on a novel reduction from minimizing the total completion time to a natural makespan-like objective, which we call the weighted free time. We hope this free time objective will be useful in further improvements to this problem, as well as other stochastic scheduling problems.

Participants

- Antonios Antoniadis
University of Twente –
Enschede, NL
- Yossi Azar
Tel Aviv University, IL
- Etienne Bamas
EPFL – Lausanne, CH
- Siddhartha Banerjee
Cornell University – Ithaca, US
- Sanjoy Baruah
Washington University –
St. Louis, US
- Sami Davies
Northwestern University –
Evanston, US
- Christoph Dürr
Sorbonne University– Paris, FR
- Franziska Eberle
London School of Economics, GB
- Daniel Freund
MIT – Cambridge, US
- Naveen Garg
Indian Institute of Technology –
New Delhi, IN
- Vineet Goyal
Columbia University –
New York, US
- Thomas Kesselheim
Universität Bonn, DE
- Samir Khuller
Northwestern University –
Evanston, US
- Amit Kumar
Indian Institute of Technology –
New Delhi, IN
- Silvio Lattanzi
Google – Barcelona, ES
- Thomas Lavastida
University of Texas – Dallas, US
- Alexander Lindermayr
Universität Bremen, DE
- Alberto Marchetti-Spaccamela
Sapienza University of Rome, IT
- Claire Mathieu
CNRS – Paris, FR
- Nicole Megow
Universität Bremen, DE
- Benjamin J. Moseley
Carnegie Mellon University –
Pittsburgh, US
- Seffi Naor
Technion – Haifa, IL
- Debmalya Panigrahi
Duke University – Durham, US
- Kirk Pruhs
University of Pittsburgh, US
- Lars Rohwedder
Maastricht University, NL
- Thomas Rothvoss
University of Washington –
Seattle, US
- Kevin Schewior
University of Southern Denmark –
Odense, DK
- Jens Schlöter
Universität Bremen, DE
- Jiri Sgall
Charles University – Prague, CZ
- David Shmoys
Cornell University – Ithaca, US
- Martin Skutella
TU Berlin, DE
- Frits C. R. Spijksma
TU Eindhoven, NL
- Clifford Stein
Columbia University –
New York, US
- Leen Stougie
CWI – Amsterdam, NL
- Ola Svensson
EPFL – Lausanne, CH
- Chaitanya Swamy
University of Waterloo, CA
- Marc Uetz
University of Twente –
Enschede, NL
- Ali Vakilian
TTIC – Chicago, US
- Sergei Vassilvitskii
Google – New York, US
- Jose Verschae
PUC – Santiago de Chile, CL
- Tjark Vredeveld
Maastricht Univ. School of
Business & Economics, NL
- Andreas Wiese
TU München, DE
- Rudy Zhou
Carnegie Mellon University –
Pittsburgh, US

