

ÉLÉMENT DE PORTFOLIO 04



Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Scheduling with Untrusted Predictions, par Evripidis Bampis, Kostantinos Dogeas, Alexander V. Kononov, Giorgio Lucarelli et Fanny Pascual, présenté à la conférence IJCAI 2022 : 4581-4587.

URL de l'élément : <https://www.ijcai.org/proceedings/2022/636>

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Au cours des dernières décennies, la conception et l'analyse des algorithmes se sont basées sur l'analyse du pire cas, où la performance d'un algorithme est caractérisée par sa pire performance sur toutes les instances d'une taille donnée. Cette approche a donné lieu au développement d'algorithmes élégants et mathématiquement cohérents pour de nombreux problèmes de calcul et a permis le développement de la théorie de la complexité. L'analyse du pire des cas a été utilisée pour mesurer le temps de calcul ou la qualité de la solution des algorithmes.

Cependant, pour de nombreux problèmes, l'analyse du pire cas ne permet pas de prédire les performances des algorithmes en pratique, les exemples les plus célèbres étant la méthode du simplexe pour la programmation linéaire qui prend un temps exponentiel dans le pire des cas, mais qui est très efficace en pratique, ou le problème du remplacement du cache où l'analyse du pire cas n'est pas capable de différencier les algorithmes. Cet échec est dû à des hypothèses trop pessimistes concernant l'incertitude sur l'entrée. En effet, dans de nombreux scénarios, l'entrée est loin d'être le pire cas et présente certaines caractéristiques prévisibles. La disponibilité et l'accessibilité des données pour de nombreuses applications, ainsi que les progrès considérables de l'apprentissage automatique et de l'intelligence artificielle au cours des dernières décennies, permettent d'espérer de bonnes prédictions de l'entrée. Cependant, il n'existe aucune garantie formelle sur la qualité d'un prédicteur (d'apprentissage automatique).

Depuis 2018, une nouvelle thématique a émergé dans le domaine de l'algorithmique, l'algorithmique avec prédictions. Elle vise à exploiter les données pour traiter l'incertitude (d'entrée) afin de concevoir des algorithmes qui, lorsque la précision des prédictions est bonne, offrent une performance (garantie) proche de celle d'un algorithme qui a un accès complet à l'entrée, et en même temps, lorsque les prédictions sont fausses, offrent une performance qui n'est pas beaucoup plus mauvaise que celle d'un algorithme sans accès aux prédictions. Cette nouvelle thématique a déjà donné lieu à des nombreux travaux (voir le site [3] et l'état de l'art récent [6]). C'est dans cette thématique que ce travail s'inscrit.

3 PRÉSENTATION DE CET ÉLÉMENT

Dans ce travail, nous considérons un problème d'ordonnancement non-clairvoyant sur une ou plusieurs machines. On nous donne un ensemble de tâches dont les durées réelles ne sont pas connues à l'avance. Cependant, une valeur prédite de la durée est donnée pour chaque tâche. Chaque tâche est également caractérisée par une date d'arrivée. Dans cet article, la préemption et la migration sont autorisées sans coût supplémentaire.

En d'autres termes, les tâches peuvent être interrompues et poursuivre leur exécution à un moment ultérieur (préemption) et éventuellement sur une autre machine (migration). Chaque machine peut exécuter au maximum une tâche à la fois. Dans la première version, nous considérons l'ordonnancement sur une seule machine. Les tâches arrivent au fil du temps, et l'algorithme n'a aucune connaissance préalable de l'existence d'une tâche. Dans la deuxième version, nous considérons l'ordonnancement sur plusieurs machines identiques. Dans ce cas, nous supposons que tous les temps d'arrivées de tâches sont nuls et que l'algorithme connaît le nombre total de tâches ainsi que la prédiction de leurs durées.

L'objectif dans les deux problèmes est de minimiser la somme des temps de complétude. Nous proposons des algorithmes qui sont à la fois cohérents (optimaux quand les prédictions sont correctes) et robustes (de "bonne qualité" lorsque les prédictions sont mauvaises). Ce travail a été motivé par deux questions ouvertes par Purohit

et al. (NeurIPS 2018) dans leur papier fondateur de l'étude de l'ordonnancement non-clairvoyant dans le cadre de l'algorithmique avec prédictions.

L'article dont il est question dans cet élément de portfolio [2] est référencé dans le site qui rassemble les papiers du domaine (voir [3]) et, bien que très récent, il commence à être cité (voir par exemple [1, 4, 5]).

4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Eric Balkanski, Vasilis Gkatzelis, and Xizhi Tan. Strategyproof Scheduling with Predictions. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11 :1–11 :22, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [2] Evripidis Bampis, Konstantinos Dogeas, Alexander V. Kononov, Giorgio Lucarelli, and Fanny Pascual. Scheduling with untrusted predictions. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4581–4587. ijcai.org, 2022.
- [3] ALPS contributors. Algorithms with predictions paper-tracker. 2023.
- [4] Alexandra Lassota, Alexander Lindermayr, Nicole Megow, and Jens Schlöter. Minimalistic predictions to schedule jobs with online precedence constraints. *CoRR*, abs/2301.12863, 2023.
- [5] Russell Lee, Bo Sun, John C. S. Lui, and Mohammad H. Hajiesmaili. Pareto-optimal learning-augmented algorithms for online k-search problems. *CoRR*, abs/2211.06567, 2022.
- [6] Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. In *Beyond Worst Case Analysis*, pages 646–662. Cambridge University Press, 2021.