
Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
National Doctoral Program in Artificial Intelligence ($37^{th}$ cycle)

# Leveraging Structured Feedback in Online Learning

By

**Khaled Eldowa**
******

**Supervisor(s):**
Prof. Nicolò Cesa-Bianchi, Supervisor
Prof. Marcello Restelli, Co-Supervisor

Politecnico di Torino
2025

# Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Khaled Eldowa
2025

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

# Leveraging Structured Feedback in Online Learning

Khaled Eldowa

The field of online learning provides a general framework for modeling sequential decision-making problems. In many of its applications, the learner only receives partial feedback upon making a decision. One extreme of this phenomenon, modeled by the multi-armed bandit problem, is when the learner only gets feedback concerning the chosen decision, which provides no information about the resulting outcome had they chosen a different decision. However, it is often the case that realistic problems are endowed with more nuanced structures, emerging from the properties of the decision space and the form of the loss function, that allow leakage of information between the elements of the decision space. In this dissertation, four such structures are studied; namely, bandits with mediator feedback, online learning with graph feedback, bandits with expert advice, and online convex Markov decision processes. New algorithms are presented with proven regret guarantees, complemented in most cases with impossibility results. The aim of these findings is to characterize to what extent the structure of the problem (and the resulting information leakage) can be leveraged by the learner towards optimizing their performance. In particular, for bandits with mediator feedback, we introduce a new information-theoretic measure for the complexity of the decision set and achieve new regret bounds featuring this quantity. Additionally, we prove nearly matching lower bounds for certain classes of instances. For online learning with graph feedback and bandits with expert advice, we provide improved upper and lower regret bounds that contribute towards a refined characterization of the minimax regret rate in both settings. Finally, we study online episodic Markov decision processes with convex objectives in the adversarial setting, and present new algorithms enjoying sub-linear regret guarantees in this problem.