SEMINAR ANNOUNCEMENT

Wednesday June 12, 2024
at 03:00 pm
Room "Sala Seminari" - Abacus Building (U14)

Resource Allocation and Scheduling Problems in Computing Continua for AI

Speaker
Federica Filippini
post-doc researcher DISCo

Abstract
The market interest in edge and cloud computing reached skyrocketing marks in the last years, with a particular interest on Artificial Intelligence (AI), and the computing landscape nowadays includes a plethora of largely heterogeneous resources, spanning from IoT devices to large datacenters. The pervasiveness of AI applications and Deep Learning (DL) applied to diverse domains and the existence of this computing continuum dictate the need of efficient algorithms to tackle the component placement and resource selection problems, minimizing the execution costs and satisfying user-imposed quality of service constraints. In this seminar, I will present heuristic and Reinfocement Learning (RL)-based techniques for the optimization and management of AI and DL applications both at design-time and at runtime, and then give some insights on the ongoing work related to RL applied to the management of both AI and Function-as-a-Service workflows.

Short Bio
Federica Filippini received her M.Sc. degree in Mathematical Engineering and her Ph.D. degree in Information Technology at Politecnico di Milano, and she is currently a post-doc researcher at the University of Milano-Bicocca, at the Deparment of Informatics, Systems and Communication. Her research interests include optimization problems applied to resource selection and scheduling in Cloud and distributed environments, with a specific focus on the design-time and runtime management of application workflows in the Edge-to-Cloud Computing Continuum.

contact person for this seminar: michele.ciavotta@unimib.it