

Luka Matijević

Matematički Institut SANU, Beograd

13. travnja 2022.

Variable Neighborhood Search for Multi-label Feature Selection

With the growing dimensionality of the data in many real-world applications, feature selection is becoming an increasingly important preprocessing step in multi-label classification. Finding a smaller subset of the most relevant features can significantly reduce resource consumption of model training, and in some cases, it can even result in a model with higher accuracy. Traditionally, feature selection has been done by employing some statistical measure to determine the most influential features, but in recent years, more and more metaheuristics have been proposed to tackle this problem more effectively. In this lecture, we will present a brief introduction to machine learning and data mining algorithms, with a focus on feature selection. We will present different approaches to feature selection, concentrating on metaheuristic wrapper methods. We will also present the results of our recent work on this topic, where we proposed the *Basic Variable Neighborhood Search* (BVNS) algorithm to search for the optimal subset of features, combined with a local search method based on mutual information. The algorithm can be considered a hybrid between the wrapper and filter methods, as it uses statistical knowledge about features to reduce the number of examined solutions during the local search. We compared our approach against *Ant Colony Optimization* (ACO) and *Memetic Algorithm* (MA), using the *K-nearest neighbors* classifier to evaluate solutions. The experiments conducted using three different metrics on a total of four benchmark datasets suggest that our approach outperforms ACO and MA.