Politecnico di Torino

1859

SmartData@PoliTO

SmartData

Luca Colomba

Ph.D Student - Politecnico di Torino

Bridge-Aware Clustering: clustering by means of bridge point identification

ABSTRACT

Clustering is among the most popular unsupervised data mining tasks. Within the years, density-based clustering algorithms proved to be one of the most prominent strategies to identify clusters and noisy points. Such methodology encompasses a large variety of clustering algorithms that focus on the identification of core points, i.e. points in cluster centers. The main drawback is the difficulty in handling clusters with different densities and thus several variants of the DBSCAN algorithm were proposed.
Recently, the data mining community has explored alternative approaches towards cluster identification, aimed at identifying cluster borders instead of cluster centers. In this SmartTalk, the proposed Bridge-Aware Clustering algorithm is presented: it relies on the concept of bridge point to separate different clusters, i.e. points whose neighborhood includes points of different clusters, and label propagation. To correctly identify bridge points without any prior cluster information, we leverage on a well-known unsupervised outlier detection technique. We validated our methodology with over 20 datasets with different density distributions.

Smart Talks Season 2