Bag Constrained Structure Pattern Mining for Multi-Graph Classification

Abstract:

This paper formulates a multi-graph learning task. In our problem setting, a number of graphs form a bag, with each bag being labeled as either positive or negative. A bag is labeled positive if at least one graph in the bag is positive, and negative otherwise. In addition, the genuine label of each graph in a positive bag is unknown, and all graphs in a negative bag are negative. The aim of multi-graph learning is to build a learning model from a number of labeled training bags to predict previously unseen test bags with maximum accuracy. This problem setting is essentially different from existing multiinstance learning (MIL), where instances in MIL share welldefined feature values, but no features are available to represent graphs in multi-graph bags. To solve the problem, we propose a Multi-Graph Feature based Learning (gMGFL) algorithm that explores and selects a set of discriminative subgraphs as features to transfer each bag into a single instance, with the bag label being propagated to the transferred instance. As a result, the multi-graph bags form a labeled training instance set, so generic learning algorithms, such as decision trees, can be used to derive learning models for multi-graph classification. Experiments and comparisons on real-world multi-graph tasks demonstrate the algorithm performance.