

Probabilistic Inference for Graph Classification

Koji Tsuda

Computational Biology Research Center, National Institute of Advanced Industrial
Science and Technology, Japan

Abstract. Graph data is getting increasingly popular in, e.g., bioinformatics and text processing. A main difficulty of graph data processing lies in the intrinsic high dimensionality of graphs, namely, when a graph is represented as a binary feature vector of indicators of all possible sub-graphs, the dimensionality gets too large for usual statistical methods. We propose an efficient method for learning a binomial mixture model in this feature space. Combining the L1 regularizer and the data structure called DFS code tree, the MAP estimate of non-zero parameters are computed efficiently by means of the EM algorithm. Our method is applied to the clustering of RNA graphs, and is compared favorably with graph kernels and the spectral graph distance.