Shiraz University

Effective Graph Classification Based on Topological and Label Attributes

Hamid Shayesteh-manesh

January 5, 2016

Most Statistical Pattern Recognition algorithms assume that given samples are represented in numerical vectors. However, much real-world data is not, but as more complicated structures, such as signals, text, trees, or graphs. Examples are biological sequences (e.g., DNA and RNA), chemical compounds, voices, and semi-structured data. We try to answer the following Graph classification question: given a collection of graphs and labeled classes, how can one predict the class of a newly observed graph?

Graph kernels compute the similarity between pairs of graphs in dataset D, based on the common patterns they share. The patterns can range from the simple to the complex. Specifically the kernels are designed to exploit random walks. The similarity of two graphs can be quantified by counting labeled walks that are common to both of them. The random walk kernel proposed by B. Scholkopf et al is one of the first graph kernels, based on this idea. Shortest paths first computes the SP graph $S = (V_s, E_s)$ for each graph $G = (V, E)$. Here $V_s = V$ and a weighted edge (v_a, v_b) exists in E_s if v_a and v_b are connected by a path in G, with the edge weight representing the SP length between v_a and v_b. Given the SP graphs S_i and S_j for two input graph G_i and G_j the kernel is defined as the sum over all pairs of edges from S_i and S_j , using any suitable positive definite kernel on the edges, cyclic patterns and subtrees. The common problem among these kernels is the complexity, However, our focus in this paper is on kernels between different graphs with better computation time, which we discuss in more detail below.

While many sophisticated graph kernels have been proposed, efficiency and scalability remain as challenges, for large graph datasets. Our basic idea is to compute several topological and label attributes for each graph in the dataset, and to use the derived feature-vector attributes for classification. Like most of the graph kernel work, we use SVM. The idea is that the graphs from the same class should have similar topological and label attributes. Our method is simple, and via a detailed comparison on real benchmark datasets, we show that our topological and label feature-based approach delivers competitive classification accuracy, with better results on those datasets that have large unlabeled graph instances.

Refrences

1. Taku Kudo, Eisaku Maeda, Yuji Matsumoto,” An Application of Boosting to Graph Classification”, Nara Institute of Science and Technology, (2005)
2. oshua T. Vogelstein, William R. Gray, R. Jacob Vogelstein, and Carey E. Priebe, ”Graph Classification using Signal-Subgraphs: Applications in Statistical Connectomics”, (2015)
3. Geng Li, Murat Semerci , B ulent Yener and Mohammed J. Zaki, ”Effective Graph Classification Based on Topological and Label Attributes”, Computer 2 Science Department, Rensselaer Polytechnic Institute, Troy, Department of Computer Engineering, Bogazici University, Istanbul, Turkey, (2012)