ABSTRACT

Graph-based semi-supervised learning is a popular machine learning paradigm for making predictions using datasets where only a small portion of the observations are labeled. In a world where increasingly large amounts of information are collected, but labeling of data can be expensive and time-consuming, the scalability of semi-supervised learning methods is critical. Obstacles to the scalability of existing algorithms center around matrix inversion, determinant computation, and matrix exponentiation. Multiresolution Matrix Factorization (MMF) was recently introduced as a method for characterizing multiscale structure and defining wavelets on graphs and matrices. In this paper, I demonstrate the utility of MMF for network analysis and for graph-based learning. Using MMF as a compression scheme also permits fast matrix operations. Promising results are presented for existing benchmark semi-supervised learning datasets, as well as new datasets much larger than those used in the past. Future work will focus on improving MMF as an approximation method for large-scale learning problems.