Semi-supervised learning over network structures
March 3, 2019

Semi-supervised learning over network structured datasets is an increasingly popular approach to problems where labelled data is either costly or time consuming. The era of big data adds to this challenge by aggregating datasets of prohibitive size and complexity.

In the thesis work “Semi-supervised learning for complex networks”, Life Science Technologies major Henrik Ambos from Aalto University presents scalable, convex optimization techniques to tackle the challenges posed by massive networked data (”big data over networks”). Two particular ML methods were considered: The Network Lasso (nLasso) and the Logistic Network Lasso (lnLasso) for binary classification problems, for predicting numeric labels. In these problems, a real-valued graph signal is associated with the nodes of this empirical graph, which is optimized by imposing similarity over known labels and smoothness conditions over the set of edges. This is achieved by simultaneously minimizing a loss function over a sampling set (the index set of sampled nodes) and the total variation over the edges of the graph.

The work applies three algorithms to solve the Network Lasso problems. For the nLasso, the Alternating Direction Method Of Multipliers (ADMM) leads to an iterative algorithm with closed-form updates. The lnLasso can be similarly solved, but owing to the logistic loss term, the ADMM updates are replaced by approximate updates that can be executed efficiently. Also included are implementations via inexact ADMM and the primal-dual method.

In the experimental section of the work, the proposed learning algorithms are compared on synthetic data and against known methods: Label propagation, belief propagation and max-flow. lnLasso has also been applied to image segmentation for natural images.

The thesis work resulted in novel algorithms for scalable semi-supervised learning from networked data. Due to their low complexity, these algorithms can be applied to processing massive datasets such as arise in image or video segmentation. Future work will consider vector-valued graph signals and automated optimization of model parameters.

Contact information:
Henrik Ambos  
henrik.ambos@aalto.fi  

Computer Science Department  
Aalto University