

# Some Applications of Collective Learning

**Balaraman Ravindran**

**Abstract** Much of the real-world data have complex dependencies between the individual tuples. For example, the chance that a patient has a particular disease depends on the prevalence of the disease in the immediate neighborhood. One approach to handling such linked data is “collective learning.” In collective learning, one deals with a set of data points taken at a time. The dependencies between the data points are modeled as a graph, with the nodes representing the tuples and the edges between them representing the influence of the tuples on one another. A variety of domains lend themselves naturally to such graph-based modeling. There have been a variety of collective learning and inferencing approaches that have been proposed in the literature. In this talk, I will give a brief introduction to collective learning and describe two applications.

**keywords** Sentiment analysis • Functional site prediction • Interaction network • Label misclassification

The first of these is a sentiment analysis task. Sentiment analysis is the task of identifying the sentiment expressed in the given piece of text about the target entity under discussion. In this work, we look at the problem of analyzing sentiments at different granularities. For example, we want to analyze sentiment about a movie as whole as well as about the acting and directing. Models built for such multigrain sentiment analysis assume fully labeled corpus at fine-grained level or coarse-grained level or both. Huge amount of online reviews are not fully labeled at any of the levels, but are partially labeled at both the levels. We propose a multigrain collective classification framework to not only exploit the information available at all the levels but also use intra dependencies at each level and interdependencies between the levels. We demonstrate empirically that the proposed framework enables better performance at both the levels compared to

---

B. Ravindran (✉)  
Computer Science and Engineering, Indian Institute of Technology Madras,  
Chennai 600036, India  
e-mail: ravi@cse.iitm.ac.in

baseline approaches. Part of this work was reported in ECAI 2010, and it is a joint work with S. Shivashankar and Shamsu Dharwez.

The second task is that of functional site prediction in proteins. Functional site prediction is an important problem in the structural genomics era where we have a large number of experimentally determined protein structures with unknown function. The functional sites provide useful insights into protein function. In this paper, we propose a method for prediction of functional residues in a given protein from its three-dimensional (3D) structure. Our method exploits correlation between labels of interacting residues to obtain significant performance improvements over the existing methods on the benchmark dataset. We represent each protein as a weighted undirected residue interaction network, where spatially proximal residues in terms of their van der Waal's radii are connected by an edge. The edge weight captures correlation between the labels of interacting residues. The correlation is estimated based on the features of interacting residues. We then obtain a label assignment by minimizing combined cost of residue-wise label misclassification and violation of label correlation constraints. We solve this problem in two stages, where the first stage minimizes residue-wise label misclassification cost followed by an iterative collective inference scheme that adjusts the labels predicted in the first stage so as to minimize the correlation constraint violations. Our approach significantly outperforms state-of-the-art methods on standard benchmark dataset. This work was reported in ACM BCB 2012, and it is a joint work with Ashish V. Tendulkar, Saradindu Kar, and Deepak Vijayakeerthi.