Classification with unstructured predictors with an application to sentiment analysis

Xiaotong Shen

University of Minnesota

Abstract

Unstructured data refers to information that lacks certain structures and cannot be organized in a predefined fashion. Unstructured data often involve words, texts, graphs, objects or multimedia types of files that are difficult to process and analyze with traditional computational tools and statistical methods. In this presentation, I will focus on ordinal classification for unstructured predictors with ordered class categories, where imprecise information concerning strengths between predictors is available for predicting class labels. However, imprecise information here is expressed in terms of a directed graph, with each node representing a predictor and a directed edge containing pairwise strengths between two nodes. Statistically, we integrate the imprecise predictor relations into linear relational constraints over classification function coefficients, where large margin ordinal classifiers are introduced, subject to many quadratically linear constraints. Moreover, the proposed classifiers are applied in sentiment analysis using binary word predictors. Finally, I will discuss some computational and theoretical aspects, in addition to an application to opinion survey.

(This work is joint with Junhui Wang, Yiwen Sun and Peiyong Qu.)