ID de Contribution: **165**   Type: **Invited talk**

Learning Optimal Classification Trees Robust to Distribution Shifts

mercredi 30 juillet 2025 11:45 (30 minutes)

We consider the problem of learning classification trees that are robust to distribution shifts between training and testing/deployment data. This problem arises frequently in high stakes settings such as public health and social work where data is often collected using self-reported surveys which are highly sensitive to e.g., the framing of the questions, the time when and place where the survey is conducted, and the level of comfort the interviewee has in sharing information with the interviewer. We propose a method for learning optimal robust classification trees based on mixed-integer robust optimization technology. In particular, we demonstrate that the problem of learning an optimal robust tree can be cast as a single-stage mixed-integer robust optimization problem with a highly nonlinear and discontinuous objective. We reformulate this problem equivalently as a two-stage linear robust optimization problem for which we devise a tailored solution procedure based on constraint generation. We evaluate the performance of our approach on numerous publicly available datasets, and compare the performance to a regularized, non-robust optimal tree. We show an increase of up to 12.48% in worst-case accuracy and of up to 4.85% in average-case accuracy across several datasets and distribution shifts from using our robust solution in comparison to the non-robust one.

Authors: JUSTIN, Nathan (University of Southern California); Dr AGHAEI, Sina (Harvard University); Prof. GOMEZ, Andres (University of Southern California); Prof. VAYANOS, Phebe (University of Southern California)

Orateur: JUSTIN, Nathan (University of Southern California)

Classification de Session: ML

Classification de thématique: Stochastic Programming