

Master thesis project "Subgroup identification methods in classification scenario with multiple treatments"

Background

Methods from personalized medicine aim to identify and propose the best treatments to a patient with given characteristics. Subgroup identification methods aim at finding subgroups of patients which benefit from the same treatments, and this identification can be based on the characteristics of various natures (binary, categorical, real-valued). Recently, a decision tree method for probabilistic subgroup identification with multiple treatments (PSICA) was proposed. However, this method works only in regression setting.

Data

In the Cuatro Santos area in Nicaragua, a Health and Demographic Surveillance System was established in 2004 with the latest update in 2014. To reduce the poverty, some local interventions were implemented such as microcredit programs, technical training and gardening programs. In total, 5966 households participated in interventions and more than 50 characteristics (variables) were recorded.

Aims

- Generalize PSICA method to classification settings and modify the current implementation appropriately.
- Investigate the efficiency of classification PSICA in terms of CPU and in terms of true discovery rates and compare the results to alternative decision tree methods.
- Apply classification PSICA to Nicaragua data and analyse which interventions are most efficient in reducing poverty for different population groups.

Prerequisites

- Good knowledge of Machine learning and Statistics
- Good programming skills and willingness to learn new software tools as a part of the project

Research Team

- Oleg Sysoev, STIMA, Linköping University
- Carina Källestål, Dept of Women's and Children's Health, Uppsala University

Contact and application

- Oleg Sysoev, oleg.sysoev@liu.se