

**Title:** Redefining Class Definitions Using Constraint-Based Clustering

**Presented by:** Carla Brodley

**Authors:** Jingjing Liu, Roni Khardon, Department of Computer Science, School of Engineering, Tufts University; Kelly Moran, Byron Wallace, School of Public Health, Brown University; Mark Friedl, Department of Earth and Environment, Boston University; Damien Sulla-Menashe, Department of Earth and Environment, Boston University

**Abstract:**

Two aspects are crucial when constructing any real world supervised classification task: the set of classes whose distinction might be useful for the domain expert, and the set of classifications that can actually be distinguished by the data. Often a set of labels is defined with some initial intuition but these are not the best match for the task. For example, labels have been assigned for land cover classification of the Earth, but it has been suspected that these labels are not ideal and some classes may be best split into subclasses whereas others should be merged. We present an approach that formalizes this problem using three ingredients: the existing class labels, the underlying separability in the data, and input from the domain expert specifying an  $L \times L$  matrix of pairwise probabilistic constraints expressing their beliefs as to whether the  $L$  classes should be kept separate, merged, or split. We describe how the problem can be solved by casting it as an instance of constraint-based clustering. We present results demonstrating its application to the task of redefining a class taxonomy for land cover classification of the Earth.