Thesis: A New Feature Selection Method Based On Class Association Rule.

Abstract: Feature selection is a key process for supervised learning algorithms which involves discarding irrelevant attributes from the training dataset, from which the models are derived. One of the key feature selection approaches is Filtering. This approach often uses mathematical models to compute the relevance for each feature in the training dataset and then sort the features into descending order based on their computed scores. However, most Filtering methods face a number of challenges including, but not limited to, merely considering feature-class correlation when defining features' relevance, and not recommending which subset of features should be retained. Leaving this decision to the end-user may be impractical for multiple reasons such as the experience needed in the application domain, care, accuracy, and time. In this research, we propose a new hybrid Filtering method called Class Association Rule Filter (CARF) that deals with the aforementioned issues by identifying relevant features through the Class Association Rule Mining approach and then using these rules to define weights for the available features in the training dataset. More crucially, we propose a new procedure based on mutual information within the CARF method that suggests the subset of features to be retained by the end-user, hence reducing time and effort.

Committee:

- Professor Abdullah Tansel, Mentor, Baruch College
- Professor William Sakas, Hunter College
- Professor Xingdong Li, Nycct

Outside member:

- Professor Reda Alhajj, University Of Calgary