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

Abstract: Feature selection is a key process for supervised learning algorithms. It involves discarding irrelevant attributes from the training dataset from which the models are derived. One of the vital feature selection approaches is Filtering, which often uses mathematical models to compute the relevance for each feature in the training dataset and then sorts the features into descending order based on their computed scores. However, most Filtering methods face several challenges including, but not limited to, merely considering feature-class correlation when defining a feature's relevance; Additionally, not recommending which subset of features to retain. Leaving this decision to the end-user may be impractical for multiple reasons such as the experience required 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 which suggests the subset of features to be retained by the end-user, hence reducing time and effort. Empirical evaluation using small, medium, and large datasets that belong to various dissimilar domains reveals that CARF was able to reduce the dimensionality of the search space when contrasted with other common Filtering methods. More importantly, the classification models devised by the different machine learning algorithms against the subsets of features selected by CARF were highly competitive in terms of various performance measures. These results indeed reflect the quality of the subsets of features selected by CARF and show the impact of the new cut-off procedure proposed.

Committee:

- Professor Abdullah Tansel, Mentor, Baruch College
- Professor William Sakas, Hunter College
- Professor Xiangdong Li, NYCCT

Outside Member:
- Professor Reda Alhajj, University Of Calgary