Abstract:

Now a day's extracting useful information from a huge data collection is an important and challenging issue. Feature selection (FS) refers to the problem of selecting minimal relevant features which produce the most predictive outcome and retaining the original meaning of the features after reduction. Some of the successful techniques for feature selection from datasets are the rough set theory (RST) and extension of rough set. This dissertation starts with an outline of the fundamental concepts behind Rough set and Fuzzy-rough set related to FS methods. Feature selection methods like Quickreduct (QR), Relative Reduct (RR), Discernibility matrix , Fuzzy- rough Quickreduct (FRQR), Fuzzy-rough feature grouping are discussed here. Two new modified methods based on Lower approximation based feature selection i.e. Improved Quickreduct and Improved Fuzzy-rough feature grouping algorithms are proposed, which selects minimal subset of features and it is observed that execution time is better than the base methods. This is performed by using the concept of significance of features, a threshold based stopping criteria and a new scaling function. A comparative experimental study of the methods has been done and the experiments are carried out on twelve public domain datasets available in UCI machine learning repository to analyze the performance study.

Keywords: Feature selection, Rough set, Fuzzy set, Lower approximation, Significance, Scaling, Threshold, UCI.