

PREFACE

Due to advancement in modern technologies, various sources like network of sensors, interconnected devices, etc generate millions of data every day. This has lead to circumstances where proportion of data to the number of tools to access the same is large. Such ever expansive data is rich both in dimension and size (number of instances). Also, noise, human error in measurement, lack of proper communication, etc further lead to presence of irrelevant and redundant features, missing values in the dataset gathered. Hence, it is necessary to preprocess the datasets before applying any classification algorithm. Feature selection is a preprocessing step to remove irrelevant and/or redundant features and offers more concise and explicit descriptions of data. Feature selection has got wide applications in data mining, signal processing, bioinformatics, machine learning, etc. While instance selection removes conflicting or spurious data sample arising in the datasets. However, feature selection (FS) or instance selection (IS) alone cannot handle the ever increasing size and dimensionality of dataset. Both the aspects of data reduction must be taken into consideration for enhancing classification accuracy along with handling missing values and noise.

Rough set theory has been effectively employed as a tool for FS to solve many real-life problems without any additional parameter. However, one of the main limitations of rough set theory is discretization of data, which might lead to information loss. Fuzzy and intuitionistic fuzzy rough set comes in handy as a tool to overcome the limitations of rough set theory. Further, these tools precisely handle the vagueness and uncertainty arising in the data.

This thesis dive into the details of applying data pre-processing techniques like missing value imputation, noise removal, feature selection, data reduction or bireduct generation using fuzzy rough and intuitionistic fuzzy rough set assisted techniques.

Mathematical formulation of each concept along with underlying model construction is introduced herein. Hence, an exhaustive study is conducted covering areas such as data reduction, missing value imputation, noise removal, etc both in supervised and unsupervised domain.