Abstract

Data mining is a process of finding a data group pattern. There are 3 tasks in data mining, which are classification, association, and clusterization. Classification is a task of finding the data group pattern. The aim of finding the data group pattern is to get the data value which has not known. The classification task of data mining is started with data pre-processing, model-building, and model-testing.

The pre-processing task is a very important step in classification task of data mining. It is because this is the phase where the data to used in the pattern-building step is prepared. A kind of pre-processing step in data mining is the variable selection or which are commonly called as variable subset selection, feature selection, feature reduction, or attribute selection.

The aim of variable selection is to find the optimal data, so that the final model is more optimal, and also to find the model which is more briefed and to accelerate the learning process. In this thesis, the variable selection implemented by the rough set. Rough set is used to find the brief rule approximation of a data, in this case the variable or column subtraction.

Keywords :   Data mining, Rough Set, Classification, Pre-processing,
                    Variable selection, Feature selection