Abstract

Progress in digital data acquisition and storage technology has resulted in exponential growth in high dimensional data. Removing redundant and irrelevant features from this high-dimensional data helps in improving mining performance and comprehensibility and increasing learning accuracy. Feature selection and feature extraction techniques as a preprocessing step are used for reducing data dimensionality. This paper analyses some existing popular feature selection and feature extraction techniques and addresses benefits and challenges of these algorithms which would be beneficial for beginners.

References


17. Qinbao Song, Jingjie Ni and Guangtao Wang, “A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data”, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No.1, pp. 1-14, 2013


Index Terms

Computer Science

Information Sciences

Keywords

Feature Selection, Feature Extraction, Principal Component Analysis (PCA), Filter methods, Wrapper Methods.