Feature Selection to Classify Healthcare Data using Wrapper Method with PSO Search

Thinzar Saw
thinzarsaw@gmail.com
University of Computer Studies, Mandalay, UCSM

Phyu Hnin Myint
phyuhninmyint@gmail.com
University of Computer Studies, Mandalay, UCSM

As a result of the rapid development of technology, data that contain a large number of features are produced from various applications such as biomedical, social media, face recognition, etc. Processing of these data is a challenging task to existing data mining and machine learning algorithms to make the decision. To reduce the size of the data for processing, a feature selection technique is needed. The feature selection is a well-known attribute selection or variable selection. The objective of the feature selection is to minimize the number of attributes contains in the dataset by eliminating the unwanted and repeated attributes to improve the classification accuracy and reduce the computation cost. Although various feature selection methods are proposed, in literature, to classify the healthcare data especially cancer diagnosis, finding an informative feature for medical datasets has still remained a challenging issue in the data mining and machine learning domain. Therefore, this paper presents a feature selection approach with the wrapper method (WFS) using particle swarm optimization (PSO) search to improve the accuracy of healthcare data classification. This work is evaluated on five benchmark medical datasets publicly available from the UCI machine learning repository. The experimental results showed that the WFS-PSO approach produces higher classification accuracy applied to different classification algorithms.