Hybrid Reliability-Similarity-Based Approach for Supervised Machine Learning

Authors : Walid Cherif

Abstract : Data mining has, over recent years, seen big advances because of the spread of internet, which generates everyday a tremendous volume of data, and also the immense advances in technologies which facilitate the analysis of these data. In particular, classification techniques are a subdomain of Data Mining which determines in which group each data instance is related within a given dataset. It is used to classify data into different classes according to desired criteria. Generally, a classification technique is either statistical or machine learning. Each type of these techniques has its own limits. Nowadays, current data are becoming increasingly heterogeneous; consequently, current classification techniques are encountering many difficulties. This paper defines new measure functions to quantify the resemblance between instances and then combines them in a new approach which is different from actual algorithms by its reliability computations. Results of the proposed approach exceeded most common classification techniques with an f-measure exceeding 97% on the IRIS Dataset.