

3rd International Conference on
Big Data Analysis & Data Mining
September 26-27, 2016 London, UK

Automatic identification of irrelevant features for clustering with artificial neural network map on synthetic datasets

Aliyu Usman Ahmad and Andrew Starkey
University of Aberdeen, UK

The effective modeling of high-dimensional data with hundreds to thousands of input features remains a challenging task in the field of machine learning. One of the major challenges is the implementation of effective methods for identifying a set of relevant features, buried in high-dimensional irrelevant noises by choosing a subset x_n of the complete set of input features $x = \{x_1, x_2, \dots, x_m\}$ such that the subset x_n predicts the output y with accuracy comparable to the performance of the complete input set x , to tackle the curse of dimensionality. The problem of feature selection is very popular and has been studied by statistic and machine learning communities for a very long time, with no fully automated solution to date. In this work, we introduced a method of measuring the relevance of each individual input feature value in the competition phase of the neural network self organizing map (SOM) training using the quantization error with an automated method that uses the relevance information to prune the irrelevant inputs and guide the training of the SOM with the relevant inputs for a higher performance. A number of synthetic datasets were created with different properties to test this method and to compare against a number of current existing feature weighting methods; we demonstrated the effect of irrelevant features on the self organizing training and the performance of these methods, with proposed method having a higher performance.

Biography

Aliyu Usman Ahmad is currently a 2nd year PhD student at the University of Aberdeen, UK. He is working on Automated Big Data Analysis Methods. He is a beneficiary of the University's Elphinstone Scholarship of Excellence with an MSc in Software Development from Coventry University, UK and a BSc in Software Engineering from University of East London.

r01aua14@abdn.ac.uk

Notes: