World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:11, No:06, 2017

A Neural Network Based Clustering Approach for Imputing Multivariate Values in Big Data

Authors : S. Nickolas, Shobha K.

Abstract : The treatment of incomplete data is an important step in the data pre-processing. Missing values creates a noisy environment in all applications and it is an unavoidable problem in big data management and analysis. Numerous techniques likes discarding rows with missing values, mean imputation, expectation maximization, neural networks with evolutionary algorithms or optimized techniques and hot deck imputation have been introduced by researchers for handling missing data. Among these, imputation techniques plays a positive role in filling missing values when it is necessary to use all records in the data and not to discard records with missing values. In this paper we propose a novel artificial neural network based clustering algorithm, Adaptive Resonance Theory-2(ART2) for imputation of missing values in mixed attribute data sets. The process of ART2 can recognize learned models fast and be adapted to new objects rapidly. It carries out model-based clustering by using competitive learning and self-steady mechanism in dynamic environment without supervision. The proposed approach not only imputes the missing values but also provides information about handling the outliers.

Keywords : ART2, data imputation, clustering, missing data, neural network, pre-processing
Conference Title : ICCPSDV 2017 : International Conference on Cyber-Physical Systems Design and Verification
Conference Location : Venice, Italy
Conference Dates : June 21-22, 2017