World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:12, No:07, 2018

Decision Trees Constructing Based on K-Means Clustering Algorithm

Authors : Loai Abdallah, Malik Yousef

Abstract : A domain space for the data should reflect the actual similarity between objects. Since objects belonging to the same cluster usually share some common traits even though their geometric distance might be relatively large. In general, the Euclidean distance of data points that represented by large number of features is not capturing the actual relation between those points. In this study, we propose a new method to construct a different space that is based on clustering to form a new distance metric. The new distance space is based on ensemble clustering (EC). The EC distance space is defined by tracking the membership of the points over multiple runs of clustering algorithm metric. Over this distance, we train the decision trees classifier (DT-EC). The results obtained by applying DT-EC on 10 datasets confirm our hypotheses that embedding the EC space as a distance metric would improve the performance.

Keywords : ensemble clustering, decision trees, classification, K nearest neighbors
Conference Title : ICDM 2018 : International Conference on Data Mining
Conference Location : Dublin, Ireland
Conference Dates : July 23-24, 2018

Open Science Index, Computer and Systems Engineering Vol:12, No:07, 2018 publications.waset.org/abstracts/89656/pdf