193. OUTLIER DETECTION FOR MIXED DATA USING ITB-SP ALGORITHM AND LOF

C.Aruna1, ME final year CSE, CSICE, Ketti, The Nilgiris (aruna.rajesh02@gmail.com)
Mr.M.A.Gopisaran2, Asst. Professor, CSICE, Ketti, The Nilgiris (gopisaran.cse@gmail.com)

Outliers are the data or attributes that does not be-long to any group or cluster. There are basically two types of data, categorical data and numerical data. The data set that contains both these types of data are called mixed data. There are different types of algorithms to form clusters and to detect the outliers for different data set. It is easy to detect the outliers if it is either purely categorical or numerical. But it is really challenging to detect the outliers in a mixed data. In this paper, we propose the new concept to detect the outliers. For the given data set we first partition the data set into categorical and numerical data set. For the partitioned categorical data set we apply the ITB-SP algorithm to detect the outlier set. This algorithm uses the concept of holoentropy which is the summation of both entropy and total correlation. Here the outlier is detected based on the outlier factor. Then we find the outlier candidate set for the numerical data using Local Outlier Factor. This result in deriving two outlier set from the categorical and the numerical data set which will be clustered together to form perfect top n outliers.

Keywords- holoentropy, entropy, total correlation, local out-lier factor, candidate set.

ISBN 978-81-904760-6-5