ABSTRACT

Data clustering is an unsupervised technique that segregates data into multiple groups based on the features of the dataset. Soft clustering techniques allow an object to belong to various clusters with different membership values. However, there are some impediments in deciding whether or not an object belongs to a cluster. To solve these issues, an intuitionistic fuzzy set introduces a new parameter called hesitancy factor that contributes to the lack of domain knowledge. Unfortunately, selecting the initial centroids in a random manner by any clustering algorithm delays the convergence and restrains from getting a global solution to the problem. To come across these barriers, this work presents a novel clustering algorithm that utilizes crow search optimization to select the optimal initial seeds for the Intuitionistic fuzzy clustering algorithm. Experimental analysis is carried out on several benchmark datasets and artificial datasets. The results demonstrate that the proposed method provides optimal results in terms of objective function and error rate.
INTRODUCTION

Data Mining pertains to the task of discovering hidden knowledge from a huge volume of data. The role of data mining has become inevitable because of the large volumes of data available in various fields. The world has become a village connected by global data. Due to their voluminous nature, these data cannot be dealt with manually. It is tedious to analyze these data manually and also difficult to identify the patterns associated with them.

Data Mining recognizes the patterns that are available in data with the help of several techniques like Classification, Clustering, Association rule mining, Prediction, etc. Classification is a supervised technique that categorizes data as belonging to which class. Prediction tries to guess the relationship between the variables in data objects and Association rule mining correlates the behavior of data with the outcome of events. Data Mining finds its applications in various fields like Biomedical research, Behavioral and social sciences, Earth sciences, Market Analysis, web search, Decision Support Systems, Buying pattern prediction, etc.

Need for Clustering

Clustering is an exploratory and descriptive data analysis technique that divides objects into several homogeneous groups based on their traits. Due to the increase in large multidimensional datasets, the need for summarizing, analyzing the qualitative and quantitative aspects of data has become unavoidable. Objects with similar features are put into a single cluster. Clustering algorithms should show the same performance irrespective of the number of instances in the dataset. There may be different types of attributes in the dataset. Many real-world problems have several constraints to be satisfied while clustering data. Application areas of clustering include but are not limited to Medical image processing, Pattern recognition, Spatial database technology, Information retrieval, Computer vision, etc.

Types of Clustering Algorithms

Clustering algorithms can be categorized into partitional, hierarchical, density-based and grid-based methods. (Jain, Murty & Flynn, 1999) Partitional algorithms tend to find spherical clusters based on distance measures and generally use mean or medoid to represent cluster center. Hierarchical methods perform multiple levels of decomposition either in top-down or bottom-up fashion which is termed as divisive or agglomerative respectively. They are distance-based or density- and continuity based methods. Density-based algorithms continuously form a cluster until the density in the neighborhood exceeds some threshold and are good in finding arbitrarily shaped clusters. Grid-based methods use a multi-resolution grid structure and are very fast in nature.

Clustering algorithms can be classified into hard and soft based on the allotment of objects. A hard clustering algorithm like K-Means allows an object to be assigned to exactly one cluster. In case of soft clustering methods like Fuzzy C-Means (FCM), (Bezdek, Ehrlich & Full, 1984) an object is allocated to multiple clusters based on the membership value of the object to each of those clusters. The non-membership value is obtained by subtracting the membership value from one.
Related Content

Multifractal Singularity Spectrum for Cognitive Cyber Defence in Internet Time Series

A Proposal of SDN Based Disaster-Aware Smart Routing for Highly-Available Information Storage Systems and Its Evaluation

Face Recognition Based on Manifold Learning and SVM Classification of 2D and 3D Geodesic Curves
[www.igi-global.com/chapter/face-recognition-based-manifold-learning/49104?camid=4v1a](www.igi-global.com/chapter/face-recognition-based-manifold-learning/49104?camid=4v1a)