(1)

An Effective Evolutionary Clustering Algorithm: Hepatitis C case study

M. H. Marghny,Rasha M. Abd El-Aziz,Ahmed I. Taloba

Abstract:

ABSTRACT Clustering analysis plays an important role in scientific research and commercial application. K-means algorithm is a widely used partition method in clustering. However, it is known that the K-means algorithm may get stuck at suboptimal solutions, depending on the choice of the initial cluster centers. In this article, we propose a technique to handle large scale data, which can select initial clustering center purposefully using Genetic algorithms (GAs), reduce the sensitivity to isolated point, avoid dissevering big cluster, and overcome deflexion of data in some degree that caused by the disproportion in data partitioning owing to adoption of multi-sampling. We applied our method to some public datasets these show the advantages of the proposed approach for example Hepatitis C dataset that has been taken from the machine learning warehouse of University of California. Our aim is to evaluate hepatitis dataset. In order to evaluate this dataset we did some preprocessing operation, the reason to preprocessing is to summarize the data in the best and suitable way for our algorithm. Missing values of the instances are adjusted using local mean method.

Keywords:

Genetic algorithms, Clustering, K-means algorithm, Squarederror criterion, hepatitis-C Virus (HCV).

Published In:

International Journal of Computer Applications (0975 – 8887) , Volume 34– No.6 ,

(2)

Outlier Detection using Improved Genetic K-means

M. H. Marghny,Ahmed I. Taloba

Abstract:

ABSTRACT The outlier detection problem in some cases is similar to the classification problem. For example, the main concern of clustering-based outlier detection algorithms is to find clusters and outliers, which are often regarded as noise that should be removed in order to make more reliable clustering. In this article, we present an algorithm that provides outlier detection and data clustering simultaneously. The algorithmimprovesthe estimation of centroids of the generative distribution during the process of clustering and outlier discovery. The proposed algorithm consists of two stages. The first stage consists of improved genetic k-means algorithm (IGK) process, while the second stage iteratively removes the vectors which are far from their cluster centroids.

Keywords:

Outlier detection, Genetic algorithms, Clustering, K-means algorithm, Improved Genetic K-means (IGK)

Published In:

International Journal of Computer Applications , Volume 28 No.11 ,

(3)

Fast Efficient Clustering Algorithm for Balanced Data

Adel A. Sewisy , M. H. Marghny , Rasha M. Abd ElAziz , Ahmed I. Taloba

Abstract:

The Cluster analysis is a major technique for statistical analysis, machine learning, pattern recognition, data mining, image analysis and bioinformatics. K-means algorithm is one of the most important clustering algorithms. However, the k-means algorithm needs a large amount of computational time for handling large data sets. In this paper, we developed more efficient clustering algorithm to overcome this deficiency named Fast Balanced k-means (FBK-means). This algorithm is not only yields the best clustering results as in the k-means algorithm but also requires less computational time. The algorithm is working well in the case of balanced data.

Keywords:

Clustering; K-means algorithm; Bee algorithm; GA algorithm; FBK-means algorithm

Published In:

International Journal of Advanced Computer Science & Applications , Vol. 5 - No. 6 , pp 123-129

(4)

An effective evolutionary clustering algorithm: Hepatitis C case study

M. H. Marghny Rasha M. Abd El-Aziz Ahmed I. Taloba

Abstract:

Most Read Research Articles Novel Application of Multi-Layer Perceptrons (MLP) Neural Networks to Model HIV in South Africa using Seroprevalence Data from Antenatal Clinics An Effective Evolutionary Clustering Algorithm: Hepatitis C Case Study Adaptivity and Adaptability of Learning Object s Interface Enhanced TCP Westwood Congestion Avoidance Mechanism (TCP WestwoodNew) Migration of Legacy Information System based on Business Process Theory HomeArchivesVolume 34Number 6An Effective Evolutionary Clustering Algorithm: Hepatitis C Case Study Call for Paper - July 2015 Edition IJCA solicits original research papers for the July 2015 Edition. Last date of manuscript submission is June 20, 2015. Read More An Effective Evolutionary Clustering Algorithm: Hepatitis C Case Study E-mail Print inShare Download Full Text International Journal of Computer Applications © 2011 by IJCA Journal Volume 34 - Number 6 Year of Publication: 2011 Authors: M. H. Marghny Rasha M. Abd El-Aziz Ahmed I. Taloba 10.5120/4092-5420 M H Marghny, Rasha Abd M El-Aziz and Ahmed I Taloba. Article: An Effective Evolutionary Clustering Algorithm: Hepatitis C Case Study. International Journal of Computer Applications 34(6):1-6, November 2011. Full text available. BibTeX Abstract Clustering analysis plays an important role in scientific research and commercial application. K-means algorithm is a widely used partition method in clustering. However, it is known that the K-means algorithm may get stuck at suboptimal solutions, depending on the choice of the initial cluster centers. In this article, we propose a technique to handle large scale data, which can select initial clustering center purposefully using Genetic algorithms (GAs), reduce the sensitivity to isolated point, avoid dissevering big cluster, and overcome deflexion of data in some degree that caused by the disproportion in data partitioning owing to adoption of multi-sampling. We applied our method to some public datasets these show the advantages of the proposed approach for example Hepatitis C dataset that has been taken from the machine learning warehouse of University of California. Our aim is to evaluate hepatitis dataset. In order to evaluate this dataset we did some preprocessing operation, the reason to preprocessing is to summarize the data in the best and suitable way for our algorithm. Missing values of the instances are adjusted using local mean method.

Keywords:

Genetic Algorithms Clustering K-means algorithm Squared-error criterion Hepatitis-C Virus (HCV)

Published In:

International Journal of Computer Applications , 34 - 6 , 1-6

(5)

Developing an Efficient Spectral Clustering Algorithm on Large Scale Graphs in Spark

Ahmed I. Taloba Marwan R. Riad Taysir Hassan A. Soliman

Abstract:

Recently, most of the data can be represented by graph structures, such as social media, Protein-Protein Interaction, transportation system, systems biology,..., etc. Many researches have been achieved to cluster very large graphs but more efficient algorithms are required since such a process takes a long time and requires more memory. In this paper, we propose an Efficient Spectral Clustering Algorithm on Large Scale Graphs in Spark (ESCALG), using map reduce function and shuffling phases in Dijkstra's algorithm. In addition, ESCALG depends mainly on a sparse matrix as a data structure, which less time in execution. Then, GraphX is applied to deal with graph data processing and in GraphX used Pregel in computing shortest path. To test the performance of ESCALG, it is compared with Large-Scale Spectral Clustering on Graphs and Standard Spectral Clustering Algorithms using seven datasets, where ESCALG proved high efciency in terms of memory and time performance.

Keywords:

Spectral Clustering , Apache Spark, Large scale Graph Clustering

Published In:

2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS) , NULL , 292-298