Chapter 95 Performance Analysis of DE over K-Means Proposed Model of Soft Computing

Kapil Patidar Amity School of Engineering and Technology, India

Manoj Kumar Amity School of Engineering and Technology, India

Sushil Kumar Amity School of Engineering and Technology, India

ABSTRACT

In real world data increased periodically, huge amount of data is called Big data. It is a well-known term used to define the exponential growth of data, both in structured and unstructured format. Data analysis is a method of cleaning, altering, learning valuable statistics, decision making and advising assumption with the help of many algorithm and procedure such as classification and clustering. In this chapter we discuss about big data analysis using soft computing technique and propose how to pair two different approaches like evolutionary algorithm and machine learning approach and try to find better cause.

1. INTRODUCTION

Day by day amount of data generation is accumulative in drastic manner. Where in to describe the data, for zetta byte, popular term used is "Big data". The marvelous volume and mixture of real world data surrounded in massive databases clearly overcome old-fashioned manual method of data analysis, such as worksheets and ad-hoc inquiries. A new generation of tools and techniques with the capabilities of perceives and repeatedly, promotes users in investigating elevations of data in warehouse in bits for useful knowledge. These procedures and tools are the issue of the field of Knowledge Discovery on Database (KDD), which is mining fascinating information or design from data in large databases (Andy, Burak, Cetin, & Harun, 2012).

DOI: 10.4018/978-1-5225-1759-7.ch095

Performance Analysis of DE over K-Means Proposed Model of Soft Computing

As in the current situation Data Mining tools are very expensive only few companies have enough money to afford them. The techniques being use for data analysis is spontaneous cluster recognition. As doing on line analysis the algorithm is used should be fast that is in scientific duration it should not be calculate exhaustive but quiet probable provide a good result, so the technique expresses worldwide finest discrete cluster.

Clustering is a corporate data mining task it has been examined for use in a number of different areas of data mining and statistics recovery. It is an important unsupervised classification technique where set of design and frequently vector in a multi-dimensional space, are grouped into a cluster if pattern is same then it belongs to the same cluster but, if pattern is different, then the cluster is dissimilar. The aim of clustering techniques is to partition a heterogeneous multi-dimensional data set into group of more homogenous characteristics (Sarafis, Trinder, & Zalzala, 2002).Unsupervised clustering may be generally classified into two types- 'hierarchical' and 'partitional'. Hierarchical clustering produces a nested sequence of panels, with a single, all-encompassing cluster at the highest and singleton clusters of separable points at the lowest. The result of hierarchical clustering can be represented by graphically displaying it as a tree, called a dendogram. Partitional technique produces an unnested, just opposite to hierarchical clustering. If K is the anticipated numeral of clusters, then partitional methodologies typically find all K clusters at once (Abraham, 2006; Park, 2015).The first suggested clustering algorithm based on Euclidean distance as similarity measure and used clustering and evaluates fitness function of DE.

DE (Differential evolution algorithm) stochastic, population based optimization algorithm, it is introduced by Storn and Price in 1996. Why we use DE? It is use for global optimization and is compulsory in turfs such as engineering, measurements and investment but many real-world problematic scenarios have unbiased function that are non-differentiable, non-constant, non-linear, loud, horizontal, multi-dimensional or have many local minima, constraints or stochasticity such problem are difficult (Hatamlou, 2013). DE can be used to find estimate explanation to such problem. It is self-organizing pattern takes the modification vector of two casually selected population vector to disturb a prevailing vector (Coletta, Acharya, Ghosh, & Hruschka, 2015).

2. RELATED WORK

Clustering algorithm is broadly classified into four ways partitioning algorithm, hierarchical algorithm, density-based algorithm and grid based algorithm. Partitioning algorithm build a partition of N object into a set of k clusters. The problem of partition clustering has been approached from miscellaneous fields of knowledge, such as artificial neural network, graph theory, expectation maximization algorithms, evolutionary computing, and so on. For the evolutionary computation between them, Differential Evolution (DE) algorithm developed. Is has been shared by many features of classical Genetic Algorithms (GA).

In hierarchical algorithm, it creates a classified decomposition of the database that can be represented by dendrogram. It can be agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms initiate per element as a distinct cluster and merge them in sequentially large cluster. Disruptive algorithms initiate the entire set and continue to divide it into one after another minor clusters. Hierarchical clustering algorithms have two basic advantages. First, the numbers of classes requirement not to be quantified a priori, and second, they are independent of the initial conditions. The main drawback of hierarchical clustering techniques is that it is static. Density based clustering algorithm played important 12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/performance-analysis-of-de-over-k-means-

proposed-model-of-soft-computing/173425

Related Content

Fuzzy Logic in Healthcare

Güney Gürsel (2015). Handbook of Research on Artificial Intelligence Techniques and Algorithms (pp. 679-707).

www.irma-international.org/chapter/fuzzy-logic-in-healthcare/123096

Quantitative Research on Contribution Degree of Vibration Affecting Factors of Vehicle Suspension System Based on Robustness Analysis

Jianqiang Xiongand Le Yuan (2020). International Journal of Ambient Computing and Intelligence (pp. 71-86).

www.irma-international.org/article/quantitative-research-on-contribution-degree-of-vibration-affecting-factors-of-vehiclesuspension-system-based-on-robustness-analysis/243448

Devnagari Script Recognition: Techniques and Challenges

P. Mukherjiand P.P. Rege (2012). Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies (pp. 249-269). www.irma-international.org/chapter/devnagari-script-recognition/62694

Ambulatory EEG Data Management System for Home Care Epileptic Patients: A Design Approach

Amol Pardhiand Suchita Varade (2022). International Journal of Ambient Computing and Intelligence (pp. 1-15).

www.irma-international.org/article/ambulatory-eeg-data-management-system-for-home-care-epileptic-patients/311500

Generative AI Methodology for Producing Assisted Art: Representation of the Historical-Cultural Identity of Southern Brazil

William Alberto Cruz Castañeda, Marcellus Amadeus, André Felipe Zanellaand Felipe Rodrigues Perche Mahlow (2024). *Making Art With Generative AI Tools (pp. 179-196).*

www.irma-international.org/chapter/generative-ai-methodology-for-producing-assisted-art/343427