

Chapter 49

Optimized Clustering Techniques with Special Focus to Biomedical Datasets

Anusuya S. Venkatesan
Saveetha University, India

ABSTRACT

The clinical data including clinical test results, MRI images and drug responses of patients are documented and analyzed with machine learning and data mining tools. The scale and complexity of these datasets is a big challenge to machine learning and data mining community as the data is of mixed type. The extraction of meaningful or desired information from these datasets provides knowledge in decision making process which in turn helps for the diagnosis and treatment of the diseases. Biomedical datasets are a collection of data with diverse types as it involves images, clinical studies, statistical reports etc. The recent researches have focused on different clustering and classification methods to manage and analyze the biomedical datasets. The objective of this chapter is to cluster or classify the patterns of interest from Brain MRI images, Liver disorder and Breast cancer datasets using efficient clustering methodologies. Among the different algorithms in data mining for clustering, classification, visualization and interpretation, K Means, Fuzzy C Means and Neural Networks (NN) are frequently used for clustering and classification of biomedical datasets. The performance of these methods are greatly influenced by the initialization of K value and its convergence speed. This chapter discusses about FCM and K Means clustering methods and its optimization with meta heuristics such as Particle Swarm Optimization (PSO) and Quantum Particle Swarm Optimization (QPSO). The experimental section of this paper exhibits analysis in terms of Intra cluster distances, elapsed time and Davis Bouldin Index (DBI).

INTRODUCTION

Data mining is an interdisciplinary area involves artificial intelligence, soft computing, database system etc. The tools of data mining infer information from the databases and this information converted to knowledge of patterns and relationships. The relationships among data are referred as Classes, Clusters,

DOI: 10.4018/978-1-5225-3158-6.ch049

Associations and Sequential Patterns. Data clustering have been applied in the area of data mining and machine learning. The specific applications include statistics (McLachlan et al., 1997), bioinformatics (He et al. 2006), machine learning (Ethem Alpaydin., 2004) exploratory data analysis, image segmentation, security, medical image analysis, web handling and mathematical programming (Pyle 1999; Panov et al. 2008). The process of clustering split the data into homogeneous and inhomogeneous classes with respect to similarity between data.

Clusters are formed by finding the distance between data points. The existing tools explore the data and help to visualize it in different models. The cluster representation is one of the widely accepted exploratory models to analyze the data with the different levels of observation. Clustering is applied to biomedical datasets to understand the characteristics of bio information and to find interesting patterns associated to prior information. Most of the bio medical datasets have inherent noise and inconsistency, sometimes mixed with semantic information and experimental results. Hence, generating quality clusters on biomedical datasets is a challenging task. The role of clustering in biomedical datasets is to derive meaningful information which assists pathologists on decision making.

In case of medical image segmentation, the system works by segmenting the whole image into multiple segments and extracts only the specific region for investigation. In image segmentation, the intensities of all pixels within a homogenous cluster are similar but the intensities of inhomogeneous clusters are different from homogeneous one. Medical image analysis is mainly dependent on effective image segmentation to extract suspicious regions from complex medical images (Neeraj et al., 2010).

Clustering is categorized as an optimization problem to satisfy the criteria of minimizing the similarity within a cluster and maximizing the dissimilarity between clusters. Table 1 shows the different distance measures used to find the distance between data points. Some clustering techniques use heuristic algorithms (Bandyopadhyay S, 2002)(Das S et al.,2008)(S. Ouadfel et al.,2010) to obtain centres for clusters. The objectives of using Optimization techniques with clustering techniques are 1) to find global optima 2) to enforce robustness against initialization 3) to improve the partitioning quality 4) to deal with unknown and known number of clusters 5) to speed up convergence etc. Clusters are represented in different forms such as Connectivity models, Centroid models, Distribution models, Density models and Graph-based models.

- **Hierarchical Clustering (Ward et al, 1963):** Is an example for connectivity models which fall into two types, they are Agglomerative and Divisive. Agglomerative is bottom up approach while Divisive is top down approach.
- **Centroid Models:** Clusters are formed based on the centroids found on the original data, the number of centroids is fixed to K, the clusters are formed by assigning the data items to the nearest cluster centroid. Example: K Means (Hartigan et al, 2009).
- **Density Based Models:** Martin Ester et al. (1996) defines clusters as connected dense regions in the data space. Example: Density Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points to Identify the Clustering Structure (OPTICS).
- **Distribution Models (Bishop, 2006):** Clusters are modeled using statistical distributions, such as multivariate normal distributions using Expectation-Maximization algorithm.

29 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/optimized-clustering-techniques-with-special-focus-to-biomedical-datasets/186721

Related Content

Recordings of Impedance and Communication Between Defibrillator and Pacemaker Electrodes

Anders Jarløv, Anne Elisabeth Jarløvand Tim Toftgaard Jensen (2019). *International Journal of Biomedical and Clinical Engineering* (pp. 45-68).

www.irma-international.org/article/recordings-of-impedance-and-communication-between-defibrillator-and-pacemaker-electrodes/233542

Evaluation of Olfactory Impairment in Parkinson's Disease Using Near-Infrared Spectroscopy

Karaki Masayuki, Kobayashi Eiji, Touge Tetsuoand Mori Nozomu (2013). *Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications* (pp. 293-302).

www.irma-international.org/chapter/evaluation-olfactory-impairment-parkinson-disease/69929

EEG Based Detection of Alcoholics: A Selective Review

T. K. Padma Shriand N. Sriraam (2012). *International Journal of Biomedical and Clinical Engineering* (pp. 59-76).

www.irma-international.org/article/eeg-based-detection-alcoholics/73694

What Influences the Growth of Canadian Biotechnology Firms?

Catherine Beaudryand Joël Levasseur (2017). *Comparative Approaches to Biotechnology Development and Use in Developed and Emerging Nations* (pp. 282-319).

www.irma-international.org/chapter/what-influences-the-growth-of-canadian-biotechnology-firms/169522

Determinants of Time-to-Under-Five Mortality in Ethiopia: Comparison of Parametric Shared Frailty Models

Abebe Argaw Wogi, Shibru Temesgen Wakweyaand Yohannes Yebabe Tesfay (2018). *International Journal of Biomedical and Clinical Engineering* (pp. 1-24).

www.irma-international.org/article/determinants-of-time-to-under-five-mortality-in-ethiopia/199093