

Chapter 5

Swarm–Based Clustering for Gene Expression Data

P. K. Nizar Banu

B. S. Abdur Rahman University, India

S. Andrews Samraj

Mahendra Engineering College, India

ABSTRACT

Clustering is one of the most important techniques, which group genes of similar expression pattern into a small number of meaningful homogeneous groups or clusters. Gene expression data has certain special characteristics and is a challenging research problem. There are many applications for clustering gene expression data. Clustering can be applied for genes called gene clustering. Hard clustering allows a gene to get placed in exactly one cluster and converges in local optima. Soft clustering approach allows gene to get placed in all the clusters with some membership values. As the hard clustering approach converges in local optimum, an evolutionary computation technique like swarm clustering is required to find the global optimum solution. This chapter studies swarm clustering techniques such as Particle Swarm Clustering K-Means, Cuckoo Search Clustering, Cuckoo Search Clustering with levy flight, harmony search, Fuzzy PSO and Ant Colony Optimization based Clustering for clustering gene expression data. Evaluation measures for clustering gene expression data are also discussed.

DOI: 10.4018/978-1-5225-2375-8.ch005

Copyright ©2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

1. INTRODUCTION

The revolution in the development of DNA microarray technology for examining gene expression has created a new era for further exploration of living systems, source of disease and drug development (He & Hui, 2009). Clustering is concerned with representing a new cancer or disease as a new class. It involves analyzing a given set of gene expression profiles with the goal of discovering subgroups that share common features. It involves grouping together specimens that are based on the similarity of their expression profiles with regard to the genes represented on the array (Tarca, Romero, & Draghici, 2006). Clustering of microarray gene expression data helps to understand the gene functions, gene regulation and cellular processes (Daxin, Chaun, & Aidong, 2004). Genes in the same cluster exhibit similar expression patterns and are likely to be co-regulated. Clustering gene expression data emphasizes on finding new biological classes or refining the existing ones (Gregory & Pablo, 2003). Gene groups enable researchers to predict the functional role or regulatory control of a novel gene, based on the similarity in expression patterns of tissue samples collected from various people including healthy persons and people affected by cancer helps in effective classification of unknown samples which in turn can lead in the early diagnosis of diseases (Marcilio, Ivan, Daniel, Teresa, & Alaxander, 2008). According to Jiang *et al.*, (2004), elucidating the patterns hidden in gene expression data offers a tremendous opportunity for enhanced understanding of functional genomics. In cancer studies, (Golub *et al.*, 1999; Alon *et al.*, 1999; Spellman *et al.*, 1998; Eisen, Spellman, Brown, & Botstein, 1998; Wen *et al.*, 1998) both gene expression, signatures for cell types and signatures for biological processes have been successfully identified by clustering (Alizadeh *et al.*, 2000). GenClust is a gene based clustering approach which is capable of identifying clusters and sub-clusters of arbitrary shapes of any gene expression dataset is proposed (Sauravjyoti & Dhruva, 2010). A novel harmony search K-Means hybrid algorithm for clustering gene expression dataset is proposed by Abdul, Sebastian, & Madhu (2013). Fuzzy C-Means (Bezdek, 1981) and Genetic Algorithms (Bandyopadhyay, Mukhopadhyay, & Maulik, 2007; Maulik, Mukhopadhyay, & Bandyopadhyay, 2009) have been used effectively in clustering gene expression data. Lu, Lu, Fotouhi, Deng, & Brown, (2004) has applied Fast Genetic K-means Algorithm (FGKA) for clustering genes.

Swarm based clustering algorithms for gene expression datasets adopts the similarity of gene expression patterns and gives good clusters to predict benign and malignant tumors. Though Particle Swarm Optimization is extensively used for many practical applications, it fails in finding the initial seeds. Harmony search PSO introduced by (Nizar & Andrews, 2015a) solves the issue of finding initial seed. The most important issue in using microarray technology is the huge amount of data produced. Finding meaningful information pattern and dependencies among

25 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/swarm-based-clustering-for-gene-expression-data/177985

Related Content

Source Location Privacy Using Ant Colony Optimization in Wireless Sensor Networks

Jyoti Prakash Singhand Paramartha Dutta (2016). *Handbook of Research on Natural Computing for Optimization Problems* (pp. 227-245).

www.irma-international.org/chapter/source-location-privacy-using-ant-colony-optimization-in-wireless-sensor-networks/153815

Methods and Graphical Tools for Exploratory Data Analysis of Artificial Olfaction Experiments

Matteo Falasconi, Matteo Pardoand Giorgio Sberveglieri (2013). *Human Olfactory Displays and Interfaces: Odor Sensing and Presentation* (pp. 317-339).

www.irma-international.org/chapter/methods-graphical-tools-exploratory-data/71931

Granger Causality: Its Foundation and Applications in Systems Biology

Tian Geand Jianfeng Feng (2011). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications* (pp. 511-532).

www.irma-international.org/chapter/granger-causality-its-foundation-applications/52330

Center Symmetric Local Descriptors for Image Classification

Vaasudev Narayananand Bhargav Parsi (2018). *International Journal of Natural Computing Research* (pp. 56-70).

www.irma-international.org/article/center-symmetric-local-descriptors-for-image-classification/217023

On the Ordering Property and Law of Importation in Fuzzy Logic

Huiwen Dengand Huan Jiang (2010). *International Journal of Artificial Life Research* (pp. 17-30).

www.irma-international.org/article/ordering-property-law-importation-fuzzy/46025