

Contents

1. Introduction	6
1.1. Gene Expression Data	6
1.1.1. Phases of Gene Expression Data analysis	6
1.2. Microarray Technology	7
2. Clustering	9
2.1. Proximity Measures used in clustering	10
2.2. Overview of some Clustering approaches	11
2.2.1. Partitional Clustering	11
2.2.2. Hierarchical Clustering	12
2.2.3. Density Based Clustering	13
2.2.4. Graph Theoretic Approach to clustering	14
2.2.5. Model Based Approach	16
2.3. Cluster Evaluation Measures	17
3. Divisive Correlation Clustering Algorithm	20
3.1. Basic concepts of DCCA and steps	20
4. Enhanced Divisive Correlation Clustering Algorithm	24
4.1. The Pearson correlation similarity measure	24
4.2. The new correlation similarity measure	26
4.3. The Algorithm	27
4.4. Observations	28
5. Performance Evaluation	30
5.1. Environment	30
5.2. Dataset description	30
5.3. Comparative analysis of performance over existing algorithms	31
5.4. Expression Graphs	32
5.5. Functional Enrichment	38
5.5.1. Analysis of EDCCA with other algorithms	38
5.6. Comparison	41
6. Conclusions and Future Work	43
References	44