

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 5 |
| 1.1 | Objective | 5 |
| 1.2 | Motivation | 6 |
| 1.3 | Contribution | 6 |
| 1.4 | Report Organization | 6 |
| 2 | Background | 7 |
| 2.1 | Microarray Technology | 7 |
| 2.2 | Gene Expression Data Analysis | 8 |
| 2.3 | Computational Biological Networks | 9 |
| 3 | Related Work and Approaches | 10 |
| 3.1 | Use of Proximity Measures in Co-expression Network | 10 |
| 3.1.1 | Advantage of MI over other Proximity Measures | 10 |
| 3.1.2 | Mutual Information | 11 |
| 3.2 | Related Work | 12 |
| 3.2.1 | Comparison of some MI-based Co-expression Networks | 13 |
| 3.3 | Subspace Network Modules | 13 |
| 3.4 | Discussion | 14 |
| 4 | Methods | 15 |
| 4.1 | Proposed Architecture | 16 |
| 4.2 | Method for extraction of network modules | 16 |
| 4.3 | Method for detecting sub-space modules | 18 |
| 4.4 | Selection of Input Parameters | 19 |
| 4.5 | Complexity Analysis | 19 |
| 5 | Experimental Results | 21 |
| 5.1 | Visualisation of the coexpression network and the network modules | 21 |
| 5.2 | Validation | 21 |
| 5.2.1 | p and Q values | 22 |
| | Comparison of p and Q values at different Input Parameters | 22 |
| 5.2.2 | Topological Validation | 32 |
| 5.2.3 | Internal validation | 35 |
| 5.3 | Comparison with Existing Method | 36 |
| 5.4 | Discussion | 37 |

5.5 Conclusion and Future Work 37

List of Figures

| | | |
|-----|--|----|
| 2.1 | Microarray Technology | 8 |
| 2.2 | Gene Expression Data Matrix | 8 |
| 3.1 | Two gene expressions correlated over samples {1, 3, 5, 7, 8} | 14 |
| 4.1 | Block diagram of the whole procedure | 17 |
| 5.1 | (a) Co-expression Network representation of Network Modules from dataset 1 and (b) from dataset 4 | 22 |
| 5.2 | Some of the co-expressed sub-space modules with their selected conditions obtained from different datasets | 23 |
| 5.3 | p values of one of the modules from datasets 1-3 and dataset 5 and Q values of one of the modules from dataset 4 at different k values | 24 |
| 5.4 | (a),(b),(c) shows the p values of the subspace modules on datasets 1,2 and 3 at different sub-module thresholds | 25 |
| 5.5 | (a) shows the Q values of the subspace modules on dataset 4 and (b) the p values of the subspace modules on dataset 5 at different sub-module thresholds | 26 |
| 5.6 | Interactive 3D topological properties plots displaying results of (a) Gene set topological properties and (b) comparison with existing database (KEGG) for dataset 5 | 34 |

List of Tables

| | | |
|------|---|----|
| 3.1 | Comparison of some of the important existing methods | 13 |
| 4.1 | Symbolic Representations | 16 |
| 5.1 | Datasets Used | 21 |
| 5.2 | p values of sub-space modules from dataset 1 | 27 |
| 5.3 | p values of sub-space modules from dataset 2 | 28 |
| 5.4 | p values of sub-space modules from dataset 2 | 29 |
| 5.5 | p values of sub-space modules from dataset 3 | 30 |
| 5.6 | Q values of sub-space modules from dataset 4 | 31 |
| 5.7 | p values of sub-space modules from dataset 5 | 32 |
| 5.8 | General topological statistics of dataset 5, a random model and the entire network | 33 |
| 5.9 | General topological statistics of dataset 6, a random model and the entire network | 34 |
| 5.10 | Ranking of KEGG gene sets based on topological similarity to uploaded module for dataset 5 | 35 |
| 5.11 | Ranking of KEGG gene sets based on topological similarity to uploaded module for dataset 6 | 35 |
| 5.12 | Analysis of the internal validity indices from different algorithms on a subspace module of Dataset 1 with k=4 number of clusters | 36 |
| 5.13 | Comparison with existing method for Dataset 1 in terms of p values | 37 |
| 5.14 | Comparison with existing method for Dataset 1 in terms of Q values | 37 |