

A new GRASP metaheuristic for biclustering of gene expression data

D. Ferone¹, A. Facchiano², A. Marabotti³, and P. Festa¹

¹ University of Napoli FEDERICO II, Dept. of Mathematics and Applications, 80126 Napoli, Italy
{daniele.ferone,paola.festa}@unina.it

² CNR, Institute of Food Science, 83100 Avellino, Italy - angelo.facchiano@isa.cnr.it

³ University of Salerno, Dept. of Chemistry and Biology, 84084 Fisciano (SA), Italy – amarabotti@unisa.it

The term biclustering stands for simultaneous clustering of both genes and conditions. This task has generated considerable interest over the past few decades, particularly related to the analysis of high-dimensional gene expression data in information retrieval, knowledge discovery, and data mining [1]. Since the problem has been shown to be NP-complete, we have recently designed and implemented a GRASP metaheuristic [2,3,4]. The greedy criterion used in the construction phase uses the Euclidean distance to build spanning trees of the graph representing the input data matrix. Once obtained a complete solution, the local search procedure tries to both enlarge the current solution and to improve its H-score exchanging rows and columns. The proposed approach has been tested on 5 synthetic datasets [5]: 1) constant biclusters; 2) constant, upregulated biclusters; 3) shift-scale biclusters; 4) shift biclusters, and 5) scale biclusters.

Compared with state-of-the-art competitors, its behaviour is excellent on shift datasets and is very good on all other datasets except for scaled ones. In order to improve its behaviour on scaled data as well and to reduce running times, we have designed and preliminarily tested a variant of the existing GRASP, whose local search phase returns an approximate local optimal solution. The resulting algorithm promises to be a more efficient, general, and robust method for the biclustering of all kinds of possible biological data.

References

- [1] Y. Cheng and G. Church. Biclustering of Expression Data, Proc. Int. Conf. Intell. Syst. Mod. Biol., 8, 93-103 (2000).
- [2] T.A. Feo and M. G.C. Resende. Greedy Randomized Adaptive Search Procedures, J. Global Optim., 6, 109-134 (1995).
- [3] P. Festa and M. G.C. Resende. An annotated bibliography of GRASP - Part I: Algorithms, Int.Trans. Oper. Res., 16(1), 1–24, (2009).
- [4] P. Festa and M. G.C. Resende. An annotated bibliography of GRASP - Part II: Applications, Int.Trans. Oper. Res., 16(2), 131–172, (2009).
- [5] K. Eren, M. Deveci, O. Küçüktunç, and Ü.V. Çatalyürek. A comparative analysis of biclustering algorithms for gene expression data, Brief. Bioinform., 1–14, (2012).