

Learning Phenotype Structure Using Sequence Model

Abstract:

Advanced microarray technologies have enabled to simultaneously monitor the expression levels of all genes. An important problem in microarray **data** analysis is to discover phenotype structures. The goal is to 1) find groups of samples corresponding to different phenotypes (such as disease or normal), and 2) for each group of samples, find the representative expression pattern or signature that distinguishes this group from others. Some methods have been proposed for this issue, however, a common drawback is that the identified signatures often include a large number of genes but with low discriminative power. In this paper, we propose a g^* -sequence model to address this limitation, where the ordered expression values among genes are profitably utilized. Compared with the existing methods, the proposed sequence model is more robust to noise and allows to discover the signatures with more discriminative power using fewer genes. This is important for the subsequent analysis by the biologists. We prove that the problem of phenotype structure discovery is NP-complete. An efficient algorithm, FINDER, is developed, which includes three steps: 1) trivial g^* -sequences identifying, 2) phenotype structure discovery, and 3) refinement. Effective pruning strategies are developed to further improve the efficiency. We evaluate the performance of FINDER and the existing methods using both synthetic and real gene expression **data** sets. Extensive experimental results show that FINDER dramatically improves the accuracy of the phenotype structures discovered (in terms of both statistical and biological significance) and detects signatures with high discriminative power. Moreover, it is orders of magnitude faster than other alternatives.

SMART REACH
(S/W CREATORS & TRAINERS)



Ph: 9585554590, 9585554599

Email: support@salemsmartreach.com

URL: www.salemsmartreach.com

**450/526, Trichy Main Road, Near Sri Sakthi Kaliamman Temple, Dadhagapatti Gate,
Salem-636 006, Tamil Nadu, India.**