Xueer.Chen[1]

Survival analysis with Network Features

***Short Abstract* — Traditional diagnosis of cancer is challenged by individual differences. Explosion of biology technology is offering great datasets for individual genetic expression as more information for diagnosing diseases. Statistic techniques are increasingly developed to use microarray data for diagnosis recently; among them, iterativeBMA algorithm is able to handle uncertainty of prediction efficiently with small numbers of genes; however, it does not consider co-expression features of genes, which is quite common in biology process. This study modifies iterativeBMA by adding co-expression features described by partial linear square (PLS) scores, to improve the prediction accuracy of the iterativeBMA process.**

***Keywords* — Survival analysis, microarray data, bayesian average, partial linear square, network**

I. Introduction

THE Microarray technology is increasingly used to identify potential biomarkers for cancer prognostics and diagnostics. The main goal in applying survival analysis to microarray data is to determine a highly predictive model of patients' time to event (such as death, relapse, or metastasis) using a small number of selected genes. In the context of survival analysis, a model refers to a set of selected genes whose regression coefficients have been calculated for use in predicting survival prognosis. In the application of survival analysis to high-dimensional microarray data, a feature selection algorithm identifies this subset of genes from the gene expression training dataset. These genes are then used to build a statistical model for the continuous time to event data.

A problem with most feature selection algorithms used to produce continuous predictors of patient survival is that they fail to account for model uncertainty. With thousands of genes and only tens to hundreds of samples, it often happens that a number of different models describe the data about equally well. Yeung et al. [1] developed an iterative BMA algorithm that takes a rank-ordered list of genes and successively applies the traditional BMA algorithm until all genes have been processed.

However, Yeung's work doesn't consider "network information" among genes, which could be common in gene regulation process. The present paper presents a way to combine expression of individual genes with network information, in the aim of improving accuracy for prediction and indicating information of genetic regulation for further signal transduction research.

II. Method and result

IterativeBMA process is applied to separate patients to high- and low-risk groups first, then search for genes with significantly different PLS scores between these two groups; treat PLS scores[2,3] of these selected genes as new items to be added in a new iterativeBMA process, therefore obtain a new model which introduces network information to original iterativeBMA model.

To compare the predicting results of two models, log-rank test was conducted and found fairly pleasant accuracy. To test the reliability of adding PLS scores, permutation test was applied and shew the uniqueness of those selected genes with much worse predicting accuracy by adding PLS scores of other genes.

III. Conclusion

Including PLS scores in the cox proportional analysis can provide equally or more accurate result to iterativeBMA process. Significant differences of PLS scores for individual genes are observed between high- and low-risk groups of patients, which indicates the reasonability of introducing network features to survival analysis. In addition, this combination may bring us more relevant network information which is closely related with survival conditions.

References

[1] Yeung K, Bumgarner R, Raftery AE. Bayesian Model Averaging: Development of an Improved Multi-Class, Gene Selection and Classification Tool for Microarray Data. Bioinformatics. 2005;21:2394–2402. doi: 10.1093/bioinformatics/bti319.

[2] Pihur V, Datta S, Datta S. Reconstruction of genetic association networks from microarray data: A partial least squares approach. Bioinformatics. 2008;24:561–568. doi: 10.1093/bioinformatics/btm640.

[3] Gill R, Datta S, Datta S. A statistical framework for differential network analysis from microarray data. BMC Bioinforma. 2010;14:95. doi: 10.1186/1471-2105-11-95.

[1]Department of Chemical Engineering, Villanova University, E-mail: xuer.chen.human@gmail.com

Nothing should be here on page 2! Please limit your abstract to a single page, and create a one-page .pdf file for submission.