BATS: A user friendly Software for analyzing time series microarray experiments.

ID - 108

Angelini Claudia[1], Cutillo Luisa[1,2], De Canditiis Daniela[1], Mutarelli Margherita[3], Pensky Marianna[4]
[1]Istituto per le Applicazioni del Calcolo, CNR
[2]Telethon Institute of Genetics and Medicine
[3]Istituto di Scienze dell'Alimentazione, CNR
[4]Department of Mathematics, University of Central Florida

Motivation

Expression level of genes in a given cell can be influenced by a pharmacological or medical treatment. The response to a given stimulus is usually different for different genes and may depend on the time. Microarray experiments allow one to simultaneously monitor the level of expression of thousands of genes. Suitable statistical methods are required to automatically detect those genes that can be associated with the biological condition under investigation. In what follows we consider microarray experiments involving comparisons between two biological conditions like control and treatment made in the course of time. Special statistical algorithms are necessary for efficient analysis of this type of experiments. BATS is a user-friendly software for the Bayesian Analysis of Time Series microarray experiments based on the novel statistical approach proposed in Angelini et al.(2006). BATS can carry out analysis with both simulated and real experimental data, also it handles data from different platforms.

Methods

BATS implements a truly functional fully Bayesian method which allows the user to automatically identify and rank differentially expressed genes and estimate their profiles on the basis of time series microarray data. The arrays are taken at n different not necessarily uniformly spaced time points on the interval [0,T] with possible replicates at some or all time points. For each gene, we assume that evolution in time of its expression level is governed by a regular function, true gene expression profile, which is observed with some additive noise. According to Angelini et al. (2006) each gene expression profile is modeled as an expansion over some orthonormal basis, with unknown number of terms and coefficients. Then a fully Bayesian model for the data is developed by eliciting prior distributions on the number of terms, the coefficients and the level of the noise. In particular, all parameters in the model are treated either as random variables or as nuisance parameters which are recovered from the data. Three different Bayesian models, which vary by the way the noise is treated, are considered. All evaluations are based on analytic expressions. The proposed procedure manages successfully various technical difficulties which arise in microarray time-course experiments such as a small number of observations available, non-uniform sampling intervals, presence of missing or multiple data as well as temporal dependence between observations for each gene.

Results

BATS is a user-friendly software written in Matlab which is freely available upon request. It allows a user to analyze time series microarray experiments using three different models to account for various types of errors thus offering a good compromise between nonparametric and normality assumption based techniques. It allows a user to specify hyper-parameters of the model or estimate them from the data. The method accounts for multiplicity, selects and ranks differentially expressed genes and estimates their expression profiles. Since all evaluations are performed using analytic expressions, the entire procedure requires very small computational effort.

In the talk, we describe statistical model used in BATS, the main features of the software interface and an application of BATS to a case study of a human breast cancer cell stimulated with estrogen. The latter led to the discovery of some new differentially expressed genes which were not marked earlier due to the high variability in the raw data.

Although originally designed for cDNA time series microarray experiments, BATS supports different platforms, such as Affimetrix and Illumina. Future work will focus on the development of a method for clustering time series genes expression profiles which is designed for the data described above.

References F. Abramovich and C. Angelini (2006), Bayesian Maximum a Posteriori Multiple Testing Procedure, Sankhya, 68, 436--460. (2006) C. Angelini, D. De Canditiis, M. Mutarelli, M. Pensky. (2006) Bayesian approach to estimation and testing in time course microarray experiments, Tech. Rep. IAC-CNR n.317/06. Available http:// www.na.iac.cnr.it/rapporti/anno2006.htm Cicatiello, L., Scarfoglio, C., Altucci, L., Cancemi, M., Natoli, G., Facchiano, A., Iazzetti G., Calogero, R., Biglia, N., De Bortoli, M., Sfiligol, C., Sismondi, P., Bresciani, F. and Weisz, A., (2004). A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive trascriptome.
Journal of Molecular Endocrinology, 32, 719--775.

Email: c.angelini@iac.cnr.it