Alternating Expectation-Maximization Algorithm for a Bilinear Model in Isoform Quantification from RNA-Seq Data

Authors : Wenjiang Deng, Tian Mou, Yudi Pawitan, Trung Nghia Vu

Abstract : Estimation of isoform-level gene expression from RNA-seq data depends on simplifying assumptions, such as uniform reads distribution, that are easily violated in real data. Such violations typically lead to biased estimates. Most existing methods provide a bias correction step(s), which is based on biological considerations, such as GC content-and applied in single samples separately. The main problem is that not all biases are known. For example, new technologies such as single-cell RNA-seq (scRNA-seq) may introduce new sources of bias not seen in bulk-cell data. This study introduces a method called XAEM based on a more flexible and robust statistical model. Existing methods are essentially based on a linear model X β , where the design matrix X is known and derived based on the simplifying assumptions. In contrast, XAEM considers X β as a bilinear model with both X and β unknown. Joint estimation of X and β is made possible by simultaneous analysis of multi-sample RNA-seq data. Compared to existing methods, XAEM automatically performs empirical correction of potentially unknown biases. XAEM implements an alternating expectation-maximization (AEM) algorithm, alternating between estimation of X and β . For speed XAEM utilizes quasi-mapping for read alignment, thus leading to a fast algorithm. Overall XAEM performs favorably compared to other recent advanced methods. For simulated datasets, XAEM obtains higher accuracy for multiple-isoform genes, particularly for paralogs. In a differential-expression analysis of a real scRNA-seq dataset, XAEM achieves substantially greater rediscovery rates in an independent validation set.

Keywords : alternating EM algorithm, bias correction, bilinear model, gene expression, RNA-seq

 $\textbf{Conference Title:} ICTB \ 2019: International \ Conference \ on \ Translational \ Bioinformatics$

Conference Location : Barcelona, Spain

Conference Dates : June 11-12, 2019