Summary: We introduce a novel class of factor analysis methodologies for the joint analysis of multiple studies. The goal is to separately identify and estimate (1) common factors shared across multiple studies, and (2) study-specific factors. We develop an Expectation Conditional-Maximization algorithm for parameter estimates and we provide a procedure for choosing the numbers of common and specific factors. We present simulations for evaluating the performance of the method and we illustrate it by applying it to gene expression data in ovarian cancer. In both, we clarify the benefits of a joint analysis compared to the standard factor analysis. We have provided a tool to accelerate the pace at which we can combine unsupervised analysis across multiple studies, and understand the cross-study reproducibility of signal in multivariate data. An R package (MSFA), is implemented and is available on GitHub.

MSC:

62P10 Applications of statistics to biology and medical sciences; meta analysis
62H20 Measures of association (correlation, canonical correlation, etc.)
62H25 Factor analysis and principal components; correspondence analysis
62J15 Paired and multiple comparisons; multiple testing

Keywords:
cross-study analysis; dimension reduction; ECM algorithm; gene expression; meta-analysis; reproducibility

Software:
MSFA; R

Full Text: DOI arXiv