

Integrative analysis of high dimensional data under privacy constraints

Molei Liu

*Harvard School of Public Health
E-mail: molei_liu@g.harvard.edu*

Abstract: Integrative analysis of high-dimensional data from multiple heterogeneous studies is known to be challenging. The challenge is even more pronounced under the DataSHIELD constraint, under which the individual level data cannot be transferred from the distributed data computers (DC) to the central analysis computer (AC), due to privacy concerns and the summary statistics are shared instead. To overcome this difficulty, we propose a novel framework for high dimensional integrative analysis with generalized linear model that relies solely on the summary data. Performance of our proposed method in estimation, variable selection and simultaneous inference is studied and compared with the ideal individual level data analysis theoretically and numerically. Our approach is shown to be equivalent with the individual level data analysis and dominates existing one-shot approach. Practically, our method facilitates meta-analysis and collaboration in fields like EHR data analysis and GWAS, to improve the statistical power.