Ruprecht-Karls-Universität Heidelberg
Medizinische Fakultät Mannheim
Dissertations-Kurzfassung

Dimensional reconstruction of psychotic disorders through multi-task learning

Autor: Han Cao
Institut / Klinik: Zentralinstitut für Seelische Gesundheit Mannheim (ZI)
Doktorvater: Dr. E. Schwarz

Schizophrenia is a severe and heritable disorder affecting approximately 1% of the population. It has become clear that an improved mechanistic understanding of its underlying biology is a critical factor for improving the clinical management of schizophrenia, for refining the current diagnostic system, and for advancing psychiatry closer to precision medicine. Advanced sequencing technology has led to a fast accumulation of molecular data. Combined with the availability of extensive computing resources and sophisticated ML methods, data science is playing an increasingly important role in schizophrenia research. However, dimensionality of data, as well as the availability of different modalities, may increase faster than the number of individuals for whom such data is available, which may lead to a loss of predictive value and interpretability of algorithms derived from molecular studies. To address this, the present thesis conducted methodological developments in two areas.

First, a significant effort at the algorithmic and computational level has been made to provide the MTL algorithms as a useful tool for both individual researchers and large-scale collaboration projects. RMTL (standalone MTL package) supports a “simultaneous approach” for signature identification in heterogeneous, multi-modal datasets, e.g., for comorbidity analysis and the prediction of multiple clinical outcomes. We showed that such heterogeneity could be captured by cross-task regularization. dsMTL (federated MTL package) supports a secure MTL analysis for geographically distributed datasets. Due to the requirement of privacy protection, the institution-level heterogeneity is challenging to remove when each dataset is analyzed individually. To address this, dsMTL provides a distributed learning system resilient to such heterogeneity. We also showed that dsMTL was computationally efficient for the typical scale of molecular studies.

Second, focusing on gene expression studies of schizophrenia, this thesis explored computational approaches to extract meaningful and biologically reproducible signatures. We found an expression signature associated with schizophrenia as well as T2D, which implied mitochondrial dysfunction and oxidative stress as a unifying theme underlying the comorbidity of these conditions. We also identified a highly accurate, consistent and robust signature in heterogeneous expression cohorts of schizophrenia and controls using MTL.