Performance evaluation of disease progression models

Razvan Valentin Marinescu

Center for Medical Image Computing, University College London

15 November 2016

Performance evaluation - Aims

Measure accuracy of:

- the fitted model parameters that describe the disease process
- predicted stages of subjects

The datasets analysed fall into three broad categories:

- simulated datasets: ground truth is user-defined
- well-phenotyped datasets: diagnoses are confirmed, stages and the disease process are known
  - post-mortem confirmed dementia
  - familial AD (e.g. DIAN)
  - prion disease
- less well-phenotyped datasets: little information on diagnoses, stages or the disease process
  - Rotterdam study
  - ADNI

Performance evaluation - measures

Simulated datasets:

- same model used for generating and fitting data:
  - direct comparison of fitted parameters with true parameters
- different model:
  - need to transform the true parameters then compare

Well-phenotyped datasets:

- post-mortem confirmed dementia datasets:
  - differential diagnosis
- autosomal dominant dementia datasets:
  - prognostic accuracy of later time points from early data
  - prediction of genetic groups - supervised or unsupervised

Performance evaluation - measures

Less well-phenotyped datasets:

- Goodness of fit measures: AIC, BIC
- Staging consistency: follow-up stages $>$ baseline stages
- Elapsed time prediction: predict the elapsed time between two visits
- Correlation of stages with clinical or leave-out biomarkers:
  - MMSE
  - CDR-SOB
  - hippocampal volume
- Resampling methods:
  - cross-validation
  - bootstrapping
- Reproducibility for different:
  - models
  - datasets
  - fitting procedures
  - missing data entries

Summary

Simulated datasets

- direct comparison between fitted and true parameters

Well-phenotyped datasets

- differential diagnosis
- prognosis of later time points from early data
- prediction of genetic groups

Less well-phenotyped datasets:

- Goodness of fit: AIC, BIC
- Staging consistency
- Elapsed time prediction
- Correlation of stages with clinical or leave-out biomarkers
- Resampling methods
- Reproducibility for different