Disease Prediction using Bayesian Networks

RUSHABH SHAH

(RDS2174)

Introduction

- Disease risk prediction
- Using diagnosis history
- Applications
  - Treatment and intervention
  - Health insurance risk management
  - Public Health research

Data

- Electronic Health records: Laboratory results, physician notes, medication, diagnosis codes
- ICD-9
  - 5 digit diagnosis codes
  - Primary and Secondary diagnosis
- ~9000 documented diseases

Challenges

- Limited publicly available datasets
- High Dimensional data
- 9000 variables
- Sparse data

Approach

- Clustering
  - Multiple specialized models
  - K-Means for categorical data
- Learning
  - Structure
  - Parameters
- Prediction
  - Risk for a disease
  - Most likely disease

Structure Learning

- Search Optimization
- Model selection
  - BIC criteria
  - Prefer simpler models
- Hill Climbing
  - Greedy Search
  - Adjacent space: Add, delete, reverse edges
  - Random restarts

Structure learning

- Sparse Candidate algorithm
- Each node has at most k parents
- Mutual Information

$$I(X;Y) = \sum_{x,y} \hat{P}(x,y) \log \frac{\hat{P}(x,y)}{\hat{P}(x)\hat{P}(y)}$$

Prediction

- Assign a test patient to a cluster
- Approximate inference of conditional probabilities
- Monte Carlo simulations
- Generate samples from distributions
- Calculate conditional probabilities by counting

References

- Friedman, Nir, Iftach Nachman, and Dana Peér. "Learning bayesian network structure from massive datasets: the «sparse candidate «algorithm."*Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999.
- Kleinberg, Samantha, and George Hripcsak. "A review of causal inference for biomedical informatics." *Journal of biomedical informatics* 44.6 (2011): 1102-1112.
- Ranganath, Rajesh, et al. "The Survival Filter: Joint Survival Analysis with a Latent Time Series." UAI, 2015.
- A comorbidity-based recommendation engine for disease prediction. Folino, F. ; Inst. for High Performance Comput. & Networking (ICAR), Italian Nat. Res. Council (CNR), Rende, Italy ; Pizzuti, C.