Kernelized Bayesian Matrix Factorization (KBMF)

Mehmet Gönen, Muhammad Ammad-ud-din, Suleiman A. Khan, Samuel Kaski

1 Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, Finland
2 Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland
http://research.ics.aalto.fi/mi/

Abstract

1. We extend kernelized matrix factorization
   - with a fully Bayesian treatment,
   - with an ability to work with multiple side information sources.
2. Side information is necessary for making out-of-matrix predictions (e.g., cold-start predictions in recommender systems).
3. We mainly discuss bipartite graph inference, where the output matrix is binary.
4. We show the performance of our method
   - by predicting drug—protein interactions on two data sets.

Proposed Method

- Kernel based non-linear dimensionality reduction
- Multiple kernel learning
- Matrix factorization
- Binary classification (if data is binary)

Probabilistic Model

- A drug–protein network by Yamanishi et al. (2008)
  - 445 drugs, 664 proteins, and 2926 validated interactions
  - C: chemical similarity for drugs
  - G: genomic similarity for proteins
  - N: network similarity for proteins
- 2. 5 replication of 5-fold CV over drugs
  - Another drug–protein interaction network by Khan et al. (2012)
    - 855 drugs, 800 proteins, and 4659 validated interactions
    - 2. Two standard 3D chemical structure descriptors for drugs:
      - Amanda (Duran et al., 2008) and VolSurf (Cruciani et al., 2000)
      - A Gaussian kernel whose width is selected as √D

Drug–Protein Interaction Data Sets

KBMF is statistically significantly better than KPMF of Zhou et al. (2012) according to paired t-test (p < 0.01) on both data sets.


Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research in information technology.