

HDBIG-SCCA Documentation

Release 1.0.0, 6/17/2015

© Copyright 2015, [ShenLab](#) at [Indiana University School of Medicine](#)

Acknowledgements: [NIH R01 LM011360](#) and [NSF IIS-1117335](#).

Contact: Jingwen Yan (jingyan@uemail.iu.edu) and/or Li Shen (shenli@iu.edu)

Question or bug reporting: The HDBIG team (hdbig@iu.edu)

1. Introduction

Recent advances in brain imaging and high throughput genotyping and sequencing techniques enable new approaches to study the influence of genetic variation on brain structure and function. HDBIG is a collection of software tools for high dimensional brain imaging genomics. These tools are designed to perform comprehensive joint analysis of heterogeneous imaging genomics data. HDBIG-SCCA is an HDBIG toolkit focusing on Sparse Canonical Correlation Analysis (SCCA). The current version includes matlab implementation of knowledge guided SCCA model (KG-SCCA). It can be applied to examine the association between genetic variations and imaging phenotypes. See below for the relevant paper.

- Yan J, Du L, Kim S, Risacher SL, Huang H, Moore JH, Saykin AJ, Shen L, for the ADNI (2014) Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics*, Vol. 30 ECCB 2014, pages i564–i571.

2. License

HDBIG-SCCA uses [GNU General Public License \(GPL\)](#). The license description is included in the software package. Please review and accept the license before installing HDBIG-SCCA via any source.

3. Download

Software

- Available at <http://www.iu.edu/~hdbig/SCCA/>

Documentation

- HTML: <http://www.iu.edu/~hdbig/SCCA/HDBIG-SCCA-v1.0.0.html>
- PDF: <http://www.iu.edu/~hdbig/SCCA/HDBIG-SCCA-v1.0.0.pdf>

4. Folder Structure and Demo Examples

The package “HDBIG-SCCA-v1.0.0.zip” consists of five subfolders.

- 00_data: Synthetic X, Y, and their prior structures (group and network)
- 01_example: Example functions for demonstration
- 02_data_preprocessing: Functions for data preprocessing
- 03_association_code: association functions (See “Methods” for details)

- 99_license: The license description.

All the functions described in the following “Methods” section are located in “03_association_code”. The current version only supports Matlab. For each of these functions, we have a corresponding example function for demonstration. These examples can be found under “01_example”. Within each example, we perform the following steps

- Load synthetic data
- Data quality control (such as removing empty entries)
- Data Normalization (Let mean = 0 and standard deviation = 1)
- Running the corresponding association model and return three outputs: two canonical loadings for X, Y respectively and objective function values during iteration

5. Methods

In this package, currently only one state-of-art association model is included.

- Knowledge-guided Sparse Canonical Correlation Analysis (KG-SCCA)

Knowledge-guided Sparse Canonical Correlation Analysis (KG-SCCA): Structured sparsity has received substantial attention in the past few years. Although group and neighboring structure has been proved to help improve the association power, human brain is mostly known to function as network. KG-SCCA takes this advantage and works as an advanced version of SCCA, which takes more complicated network structure as input to guide the association procedure.

Example usage:

```
[u,v,obj] = A_KG_SCCA(X, Y, group, network, para);
```

where X is n x c matrix and Y is n x d matrix. “group” is a vector of c, indicating the group belongings of each X features. “network” is a p x d matrix, indicating the network relationship among Y features. Each row in “network” indicates an edge, where only i^{th} and j^{th} element is not zero if it is connecting node i and j. “para” controls the strength of the penalty term.