Coordinative Component Analysis Toolbox for MATLAB

1. Introduction:

Coordinative component analysis (CCA) is a knowledge-guided gene ranking approach to rank genes for pathway member identification. The detailed description of the CCA approach can be found in the manuscript, entitled "knowledge-guided gene ranking by coordinative component analysis", submitted to *Bioinformatics*, 2009.

2. Functions

2.1 Basic CCA (without bootstrapping)

Gene Ranking by Coordination Component Analysis

Syntax:

```
[sorted_index, y, convg,w] = CCA(E, Knowledge_indicator);

[sorted_index, y, convg,w] = CCA(E, Knowledge_indicator, mju, c,
                                 max_iters);
```

Parameter	**value**
'E'	gene expression matrix of gene by sample
'K_indices'	index vector of the knowledge gene
'mju' (optional)	gradient step size of CCA, default 5e-2
'c' (optional)	perturbation parameter of CCA, default 1e-2
'max_iters' (optional)	maximum iteration number, default 1e4

Output	**value**
'sorted_index'	ranked gene index, from top to down
'y'	estimated coordinative component
'convg'	convergence curve of CCA
'w'	estimated CCA filter

2.2 Bootstrapped CCA

Bootstrapped Coordination Component Analysis

Syntax:

```
[Y_all, W_all, ConvCurv_all] = Bstrp_CCA(E, K_indices )

[Y_all, W_all, ConvCurv_all] = Bstrp_CCA(E, K_indices, Bootstrap_Num,
                                         Sampled_Gene_Num, Replace_flag,
                                         mju, c, max_iters, verbose)
```

Parameter	**value**
'E'	gene expression matrix of gene by sample
'K_indices'	index vector of the knowledge gene
'Bootstrap_Num'(optional)	number of bootstraps
'Sampled_Gene_Num'(optional)	sampled gene number for each bootstrap step
'Replace_flag'(optional)	replacement or not when bootstrapping, default true
'mju' (optional)	gradient step size of CCA, default 5e-2
'c' (optional)	perturbation parameter of CCA, default 1e-2

'max_iters' (optional)	maximum iteration number, default 1e4
'verbose' (optional)	displaying the progress of bootstrapping or not, default true

Output	**value**
'Y_all'	estimated bootstrapped coordinative component
'W_all'	estimated bootstrapped CCA filter
'ConvCurv_all'	convergence curve of bootstrapped CCA

2.3 Assembling the bootstrapped CCA results

Gene ranking according to the results from bootstrapped CCA

Syntax:

```
[sorted_index, CCA_comp] = CCA_Bstrp_asmb(E, W_all)
```

Parameter	**value**
'E'	Gene expression matrix of gene by sample
'W_all'	estimated bootstrapped CCA filter

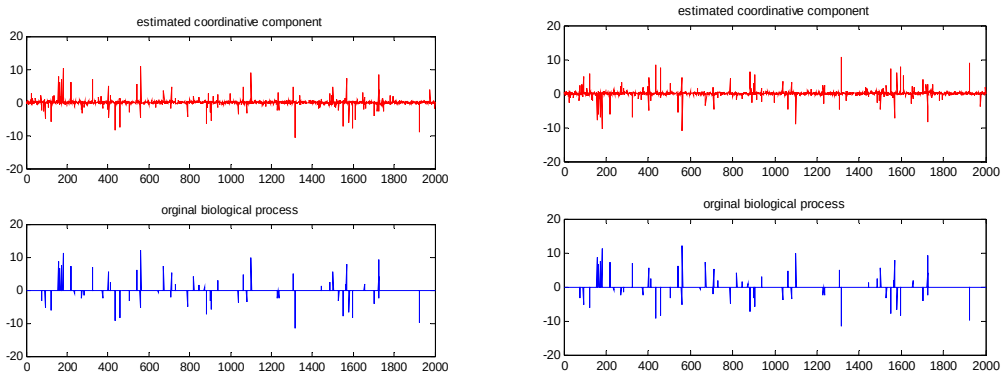
Output	**value**
'sorted_index'	ranked gene index, from top to down
'CCA_comp'	combined coordinative component estimation

3. Demo

CCA_demo.m gives an example of the CCA ranking when partial prior knowledge is available (i.e., 25%). One can adjust various parameters (gene/sample/biological process number, signal-to-noise ratio and CCA parameters, etc.) to get familiar with CCA. Graphical results are generated to better help understand CCA visually.

3.1 Comparison of the estimated coordinative component and the original biological process

One should be able to observe that the estimated coordinative component is very similar to the original biological process, in a positively or negatively correlated way, shown in Fig. 1(a) and Fig. 1(b), respectively. That's due to the sign ambiguity in the latent variable model, and such sign ambiguity does not affect the gene ranking result, which only accounts for the absolute amplitude.



(a) (b)

Fig. 1. Comparison of the estimated coordinative component and the original biological process: (a) positive correlation and (b) negative correlation.

3.2 Convergence curve of CCA

One can check the convergence curve of CCA, to make sure that with current parameter setting, the algorithm converges to a stable result. One can modify the step size or maximum iteration number to observe its impact on the CCA convergence. Practically, it is sufficient to call "being converged" when the curve slows down dramatically at the end of iterations.

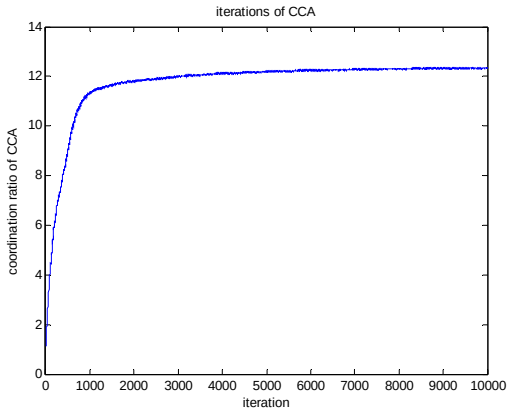


Fig. 2. Convergence curve of CCA.

3.3 Receiver operating characteristic (ROC) curve of CCA

To check the overall gene ranking performance of CCA, one can use the ROC curve to evaluate the sensitivity and specificity of gene ranking (see Fig. 3 for an example).

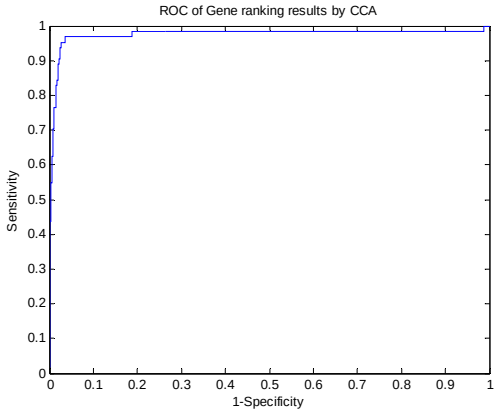


Fig. 3. ROC curve of the gene ranking results by CCA.