Bayesian Generalized Biclustering Analysis via Adaptive Structured Shrinkage[1]

Qi Long, Ph.D.

Department of Biostatistics, Epidemiology and Informatics

Perelman School of Medicine

University of Pennsylvania

November 7, 2018

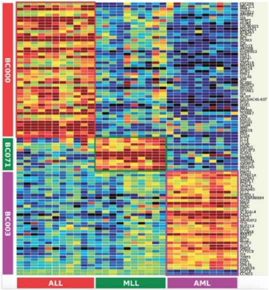
[1] Joint work with Ziyi Li, Changgee Chang, Suprateek Kundu

Biclustering

- Biclustering, also called block clustering, co-clustering, or two-mode clustering, is a data mining technique which cluster the rows and columns of a data matrix simultaneously.

- The first biclustering method dates to 1972 by J.A.Hartigan. The first application of biclustering method in gene expression data was by Y. Cheng and G. M. Church in 2000.

    - Biclustering identifys the clusters of features in different conditions, which is useful for visualization, pattern recognition, clustering, and etc..



Biclustering: Existing Methods

- A large number of methods develdoped (Padilha et al. 2017).
- Loosely, the current biclustering methods can be grouped to four groups.
    - Distribution parameter identification algorithm: Plaid, Factor analysis for bicluster acquisition (FABIA) etc.
    - Greedy algorithm: CC, xMotifs, ISA, etc.
    - Divide and conquer algorithm: Binary Inclusion-Maximal Biclustering Algorithm (Bimax)
    - Exhaustive enumeration algorithm: Statistical-Algorithmic Method for Biclustering Analysis (SAMBA)

Gaps

- Although many biclustering approaches have been developed, few of them can utilize the existing biological information such as gene regulatory networks for identifying biclustering patterns.
- Most existing methods focus on analyzing gene expression microarray data which are of continuous data type.
- Our simulation results have shown the current methods cannot identify biclusters with good accuracy on inputs of discrete data types or mixed data types, for example, continuous and binary data.

Our Goals

- To develop biclustering algorithm that can handle data of multiple types, continuous and discrete.
- To enable feature selection guided by existing biological information such as gene regulatory networks that can be represented by a graph.

Notation

n number of subjects

H -omic platforms, such as microarray and next-generation sequencing

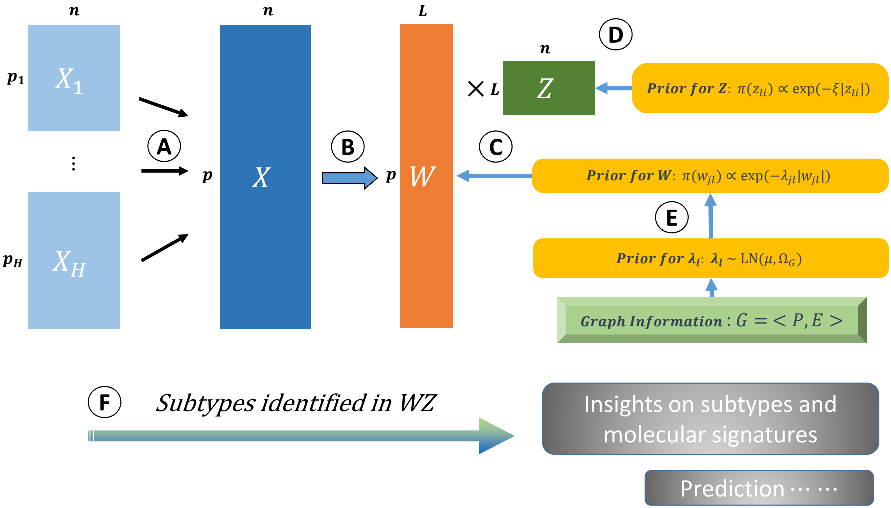
$\mathbf{X}_h$ $h = 1, \cdots, H$, observed data from H -omic platforms, each matrix has size $p_h \times n$

$\mathbf{X}$ the vertical concatenation of observed data matrices with size $p \times n$ and $p = \sum_{h=1}^{H} p_h$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_H \end{bmatrix}.$$

- $\mathcal{G} = \langle P, E \rangle$: biological/network information from say KEGG, where P denotes th set of p variables and $E = \{(\iota(h,j), \iota(h,k)) : (j,k) \in E_h, 1 \leq h \leq H\}$

Bayesian Sparse Generalized Bi-Clustering (GBC)



Prior for Z: $\pi(z_{li}) \propto \exp(-\xi|z_{li}|)$

Prior for W: $\pi(w_{jl}) \propto \exp(-\lambda_{jl}|w_{jl}|)$

Prior for $\boldsymbol{\lambda_l}$: $\boldsymbol{\lambda_l} \sim \mathrm{LN}(\mu, \Omega_G)$

Graph Information : $G = < P, E >$

Subtypes identified in WZ

Insights on subtypes and molecular signatures

Prediction ⋯ ⋯

Mean Model

μ : mean of $\mathbf{X}$ is related to latent components through

$$\mu = \mathbf{m} + \mathbf{WZ}$$

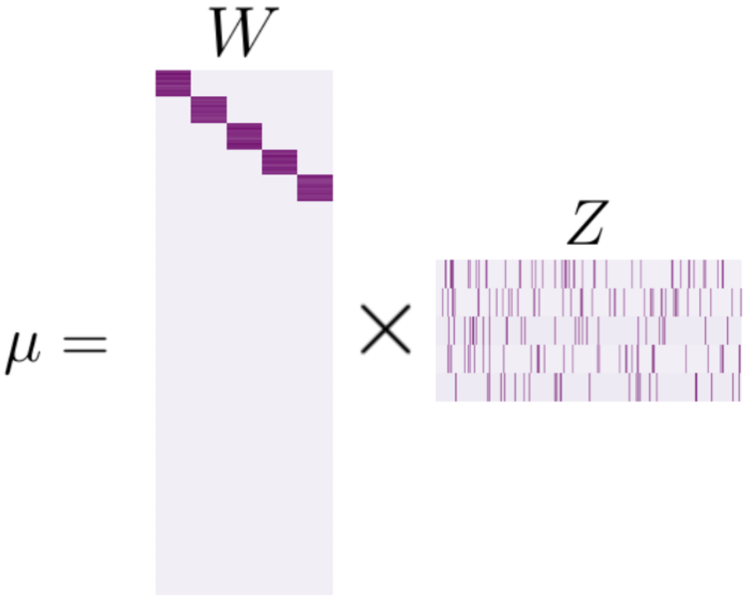
where $\mathbf{m}$ is the location vector.

$\mathbf{W}$: $p \times L$ factor loading matrix

$\mathbf{Z}$: $L \times n$ latent factor matrix

- Define biclusters: $\mathbf{w}_k \times \mathbf{z}_k$ forms the k-th bicluster, where $\mathbf{w}_k$ is column k of $\mathbf{W}$ and $\mathbf{z}_k$ is row k of $\mathbf{Z}$, $k = 1, \ldots, L$.
- Our model induces sparsity in $\mathbf{w}_k$ and $\mathbf{z}_k$, so the non-zero elements in $\mathbf{w}_k$ ($\mathbf{z}_k$) represent the subset of features (subjects) belonging to the k-th bicluster.
- Allow overlapping biclusters.
- L is the maximum number of biclusters, noting that $\mathbf{w}_k$ or $\mathbf{z}_k$ can be 0.

Defining Biclusters



Likelihood Functions in a Unified Form

Observed data likelihood: $\pi(\mathbf{X}|\boldsymbol{\mu}) = \prod_j \prod_i \pi_j(x_{ji}|\mu_{ji})$

- For Gaussian data:

$$\pi_j(x_{ji}|\mu_{ji}, \rho_j) = \frac{\rho_j^{1/2}}{\sqrt{2\pi}} e^{-\rho_j (x_{ji}-\mu_{ji})^2/2}.$$

- For Binomial data with logit link:

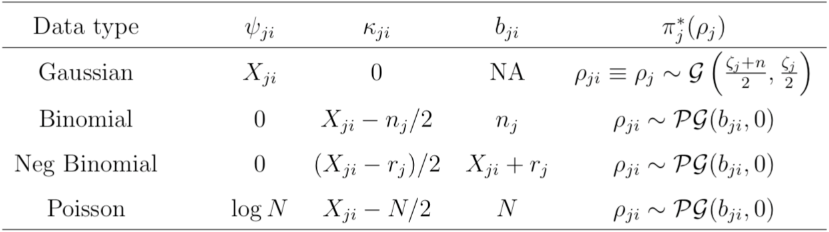
$$\pi_j(x_{ji}|\mu_{ji}, n_j) = \binom{n_j}{x_{ji}} \frac{e^{\mu_{ji} x_{ji}}}{(1+e^{\mu_{ji}})^{n_j}}, x_{ji} = 0, 1, \ldots, n_j.$$

- For Negative Binomial data with logit link:

$$\pi_j(x_{ji}|\mu_{ji}, r_j) = \binom{r_j + x_{ji} - 1}{x_{ji}} \frac{e^{\mu_{ji} x_{ji}}}{(1+e^{\mu_{ji}})^{r_j + x_{ji}}}, x_{ji} = 0, 1, 2, \ldots.$$

Likelihood Functions in a Unified Form

All likelihood functions can be written in a unified form (Polson et al. 2013): $\pi_j(\mathbf{x}_j|\mu_j) \propto e^{-\frac{1}{2}\sum_i \rho_{ji}(\mu_{ji}-\psi_{ji})^2+\sum_i \kappa_{ji}\mu_{ji}}\pi_j^*(\rho_j)$



Data type	ψ_{ji}	κ_{ji}	b_{ji}	$\pi_j^*(\rho_j)$
Gaussian	X_{ji}	0	NA	$\rho_{ji} \equiv \rho_j \sim \mathcal{G}\left(\frac{\zeta_j+n}{2}, \frac{\zeta_j}{2}\right)$
Binomial	0	$X_{ji} - n_j/2$	n_j	$\rho_{ji} \sim \mathcal{PG}(b_{ji}, 0)$
Neg Binomial	0	$(X_{ji} - r_j)/2$	$X_{ji} + r_j$	$\rho_{ji} \sim \mathcal{PG}(b_{ji}, 0)$
Poisson	$\log N$	$X_{ji} - N/2$	N	$\rho_{ji} \sim \mathcal{PG}(b_{ji}, 0)$

Advantages: closed-form M steps in EM algorithm; enable the use of Gibbs sampling instead of Metropolis-Hasting in MCMC.

Prior Specification

$\boldsymbol{\mu} = \mathbf{m} + \mathbf{WZ}$

- **Prior for Z**

$$\log \pi(\mathbf{Z}|\boldsymbol{\xi}) = C + \sum_{l,i} \log \xi_{li} - \sum_{l,i} \xi_{li}|z_{li}|,$$

Gamma prior on ξ:
$\log \pi(\boldsymbol{\xi}) = C_{\nu_3,\nu_4} + (\nu_3 - 1)\sum_{l,i} \log \xi_{il} - \frac{1}{\nu_4}\sum_{l,i} \xi_{li}$
where ν_3 and ν_4 are tuning parameters.

- **Prior for W**

$$\log \pi(\mathbf{W}|\boldsymbol{\lambda}) = C + \sum_{j,l} \log \lambda_{jl} - \sum_{j,l} \lambda_{jl}|w_{jl}|$$

Prior for $\boldsymbol{\lambda}$: Graph-Laplacian prior incorporating biological information

Prior for λ Incorporating Biological Information

Adaptive Structured Shrinkage (Chang et al. 2018):

- Let $\alpha_{jl} = \log \lambda_{jl}$
- Graph-Laplacian prior for $\boldsymbol{\alpha}_l = (\alpha_{1l}, \ldots, \alpha_{pl})'$ $(1 \leq l \leq L)$

  $$\log \pi(\boldsymbol{\alpha}|\boldsymbol{\Omega}) = C_{\nu_2} + \frac{L}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2\nu_2} \sum_{l} (\boldsymbol{\alpha}_l - \nu_1 \underline{1}) \boldsymbol{\Omega} (\boldsymbol{\alpha}_l - \nu_1 \underline{1}),$$

  where ν_1 and ν_2 are tuning parameters.
- The precision matrix $\boldsymbol{\Omega}$ imposes dependency among α_{jl}'s, allowing us to incorporate the network information $\mathcal{G}$.
  - H graphs $\mathcal{G}_h = \langle P_h, E_h \rangle$;
  - $\mathcal{G} = \langle P, E \rangle$ where P denotes th set of p variables and $E = \{(\iota(h,j), \iota(h,k)) : (j,k) \in E_h, 1 \leq h \leq H\}$

Prior for λ Incorporating Biological Information

$$\boldsymbol{\Omega} = \begin{bmatrix} 1+\sum_{j\neq 1}\omega_{1j} & -\omega_{12} & \cdots & -\omega_{1p} \\ -\omega_{21} & 1+\sum_{j\neq 2}\omega_{2j} & \ddots & -\omega_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ -\omega_{p1} & -\omega_{p2} & \cdots & 1+\sum_{j\neq p}\omega_{pj} \end{bmatrix}$$

- If $G_{jk} = 0, \omega_{jk} = 0$ and nodes j and k receive (partially) independent shrinkage
- If $G_{jk} = 1, \omega_{jk} > 0$ and they tend to receive similar levels of shrinkage
- $\boldsymbol{\Omega}$ is symmetric and is diagonally dominant and thus positive definite

Prior for λ Incorporating Biological Information

- Prior on $\boldsymbol{\omega} = \{\omega_{jk} : j < k\}$

  $$\pi(\boldsymbol{\omega}) \propto |\boldsymbol{\Omega}|^{-L/2} \prod_{(j,k)\in E} \omega_{jk}^{a_\omega - 1} \exp(-b_\omega \omega_{jk}) 1(\omega_{jk} > 0) \prod_{(j,k)\neq E} \delta_0(\omega_{jk}).$$

  $\delta_0(\cdot)$ is the Dirac delta function concentrated at 0 and $1(\cdot)$ is the indicator function.

- $|\boldsymbol{\Omega}|^{-L/2}$ induces correlation among ω and ensures a closed-form posterior density for ω.

- a_ω takes the role of the shape parameter and b_ω determines the scale of ω_{jk}.

- It has been shown that this prior is proper (Chang et al. 2018).

MAP Estimator

- MCMC is computationally expensive for high-dimensional data.
- Consider the Maximum-A-Posteriori (MAP) estimator $(\hat{\boldsymbol{W}}, \hat{\boldsymbol{Z}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\xi}})$ with $\boldsymbol{\rho}, \boldsymbol{\Omega}$ marginalized out.

$$(\hat{\boldsymbol{W}}, \hat{\boldsymbol{Z}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\xi}}) = \underset{W,Z,\alpha,\xi}{\arg\max} \int\int \pi(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\rho}, \boldsymbol{\Omega}|\boldsymbol{X}) d\boldsymbol{\rho} d\boldsymbol{\Omega}.$$

- We develop an EM algorithm for obtaining MAP

$$(\boldsymbol{W}^{(t)}, \boldsymbol{Z}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\xi}^{(t)}) = \underset{W,Z,\alpha,\xi}{\arg\max} \tilde{\mathbb{E}}_t \log \pi(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\rho}, \boldsymbol{\Omega}, \boldsymbol{X}),$$

where the expectation $\tilde{\mathbb{E}}_t$ is taken with respect to $\tilde{\pi}_t(\boldsymbol{\rho}, \boldsymbol{\Omega}) = \pi(\boldsymbol{\rho}, \boldsymbol{\Omega}|\boldsymbol{W}^{(t-1)}, \boldsymbol{Z}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\xi}^{(t-1)}, \boldsymbol{X})$.

EM Algorithm: Objective Function

The objective function to be optimized at the t-th EM iteration is:

$$
\begin{aligned}
\mathbf{Q}_t(\mathbf{Z}, \mathbf{W}, \boldsymbol{m}, \alpha, \xi) = & -\frac{1}{2} \sum_{i,j} \rho_{ji}^{(t)} (\mu_{ji} - \psi_{ji})^2 + \sum_{i,j} \kappa_{ji} \mu_{ji} + \sum_{j,l} \alpha_{jl} - \sum_{j,l} \lambda_{jl} |w_{jl}| \\
& + \nu_3 \sum_{l,i} \log \xi_{i,l} - \sum_{i,l} \xi_{l,i} (|z_{li}| + \frac{1}{\nu_4}) \\
& - \frac{1}{2\nu_2} \sum_{l} (\boldsymbol{\alpha_l} - \nu_1 \mathbf{1})^T \boldsymbol{\Omega}^{(t)} (\boldsymbol{\alpha_l} - \nu_1 \mathbf{1})
\end{aligned}
$$

where $\boldsymbol{\mu} = \boldsymbol{m} + \boldsymbol{WZ}$,

$$
\begin{aligned}
\boldsymbol{\rho}^{(t)} &= \mathbb{E}(\boldsymbol{\rho}_{ij} | \boldsymbol{X}, \boldsymbol{W}^{(t-1)}, \boldsymbol{Z}^{(t-1)}, \boldsymbol{m}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\xi}^{(t-1)}, \boldsymbol{\Omega}^{(t-1)}), \\
\text{and} \quad \boldsymbol{\Omega}^{(t)} &= \mathbb{E}(\boldsymbol{\Omega} | \boldsymbol{X}, \boldsymbol{W}^{(t-1)}, \boldsymbol{Z}^{(t-1)}, \boldsymbol{m}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\xi}^{(t-1)}, \boldsymbol{\rho}^{(t)}).
\end{aligned}
$$

Tuning Parameters

- Fix $a_\omega = 4$ and $b_\omega = 1$: large prior correlation and at the same time relatively uninformative;
- Fix $\nu_2 = \ln 2$ and $\nu_3 = 1$: the corresponding priors for α and ξ have a unit coefficient of variation;
- ν_1 and ν_4 control sparsity of $\mathbf{W}$ and $\mathbf{Z}$ and are chosen by BIC:

$$BIC = -2\ln(L(\mathbf{X}, \hat{\boldsymbol{\mu}})) + (||\hat{\mathbf{W}}||_0 + ||\hat{\mathbf{Z}}||_0)\ln(np)$$

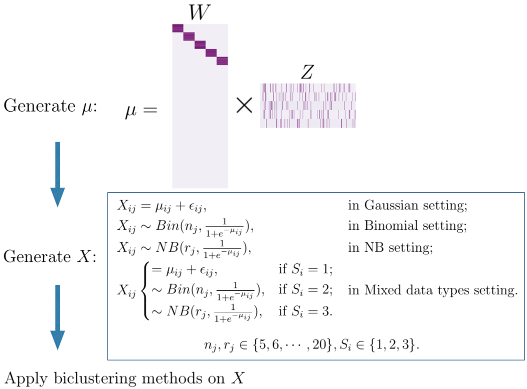
where $L(\mathbf{X}, \hat{\boldsymbol{\mu}})$ is the observed likelihood of $\boldsymbol{\mu}$, $||\hat{\mathbf{W}}||_0$ and $||\hat{\mathbf{Z}}||_0$ are the cardinalities of $\hat{\mathbf{W}}$ and $\hat{\mathbf{Z}}$;

Simulation: Methods

- Existing methods:
  - **CC** (Cheng and Church's Biclustering Algorithm)
  - **xMotifs** (Conserved gene expression motifs)
  - **ISA** (Iterative Signature Algorithm)
  - **Plaid**
  - **FABIA** (Factor Analysis for Biclustering Acquisition)
- **GBC** (Generalized Biclustering): specify $\boldsymbol{\Omega}$ as identity matrix
- **sGBC** (Generalized Biclustering with incorporation of biological information)

Simulation: Settings

- Four simulation settings: gaussian, binomial, negative binomial, and mixed datatypes.
- 100 simulation datasets with $p = 1000$, $n = 300$, $L = 5$ underlying true biclusters.
- presence or absence of overlapping clusters.



Generate μ: $\mu =$ ×

Generate X:

$$
\begin{array}{ll}
X_{ij} = \mu_{ij} + \epsilon_{ij}, & \text{in Gaussian setting;} \\
X_{ij} \sim Bin(n_j, \frac{1}{1+e^{-\mu_{ij}}}), & \text{in Binomial setting;} \\
X_{ij} \sim NB(r_j, \frac{1}{1+e^{-\mu_{ij}}}), & \text{in NB setting;} \\
X_{ij} \begin{cases} = \mu_{ij} + \epsilon_{ij}, & \text{if } S_i = 1; \\ \sim Bin(n_j, \frac{1}{1+e^{-\mu_{ij}}}), & \text{if } S_i = 2; \\ \sim NB(r_j, \frac{1}{1+e^{-\mu_{ij}}}), & \text{if } S_i = 3. \end{cases} & \text{in Mixed data types setting.}
\end{array}
$$

$$n_j, r_j \in \{5, 6, \cdots, 20\}, S_i \in \{1, 2, 3\}.$$

Apply biclustering methods on X

Figure: Work flow of the simualtion study.

Simulation: Evaluation Criteria

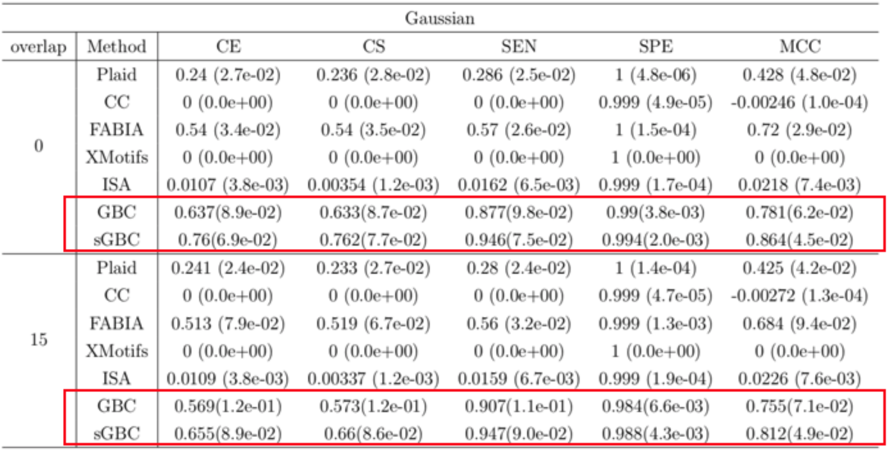
- **Clustering error (CE)** (Patrikainen and Meila, 2006) finds the maximum overlapping proportions of two biclusters after an optimal matching of clusters. CE considers the size of biclusters.

- **Consensus scores (CS)** (Hochreiter et al., 2010). CS gives the same weight to all biclsuters.

- **Sensitivity**, **Specificity**, and **Matthews correlation coefficient (MCC)**.

- All these metrics take values between 0 and 1 with higher values indicating better performance.

Simulation Results: Gaussian



		Gaussian				
overlap	Method	CE	CS	SEN	SPE	MCC
0	Plaid	0.24 (2.7e-02)	0.236 (2.8e-02)	0.286 (2.5e-02)	1 (4.8e-06)	0.428 (4.8e-02)
	CC	0 (0.0e+00)	0 (0.0e+00)	0 (0.0e+00)	0.999 (4.9e-05)	-0.00246 (1.0e-04)
	FABIA	0.54 (3.4e-02)	0.54 (3.5e-02)	0.57 (2.6e-02)	1 (1.5e-04)	0.72 (2.9e-02)
	XMotifs	0 (0.0e+00)	0 (0.0e+00)	0 (0.0e+00)	1 (0.0e+00)	0 (0.0e+00)
	ISA	0.0107 (3.8e-03)	0.00354 (1.2e-03)	0.0162 (6.5e-03)	0.999 (1.7e-04)	0.0218 (7.4e-03)
	GBC	0.637(8.9e-02)	0.633(8.7e-02)	0.877(9.8e-02)	0.99(3.8e-03)	0.781(6.2e-02)
	sGBC	0.76(6.9e-02)	0.762(7.7e-02)	0.946(7.5e-02)	0.994(2.0e-03)	0.864(4.5e-02)
15	Plaid	0.241 (2.4e-02)	0.233 (2.7e-02)	0.28 (2.4e-02)	1 (1.4e-04)	0.425 (4.2e-02)
	CC	0 (0.0e+00)	0 (0.0e+00)	0 (0.0e+00)	0.999 (4.7e-05)	-0.00272 (1.3e-04)
	FABIA	0.513 (7.9e-02)	0.519 (6.7e-02)	0.56 (3.2e-02)	0.999 (1.3e-03)	0.684 (9.4e-02)
	XMotifs	0 (0.0e+00)	0 (0.0e+00)	0 (0.0e+00)	1 (0.0e+00)	0 (0.0e+00)
	ISA	0.0109 (3.8e-03)	0.00337 (1.2e-03)	0.0159 (6.7e-03)	0.999 (1.9e-04)	0.0226 (7.6e-03)
	GBC	0.569(1.2e-01)	0.573(1.2e-01)	0.907(1.1e-01)	0.984(6.6e-03)	0.755(7.1e-02)
	sGBC	0.655(8.9e-02)	0.66(8.6e-02)	0.947(9.0e-02)	0.988(4.3e-03)	0.812(4.9e-02)

Simulation Results: Mixed Data Types

Mixed data types						
overlap	Method	CE	CS	SEN	SPE	MCC
0	Plaid	0.0105 (1.5e-03)	0.0728 (1.2e-02)	0.227 (2.6e-02)	0.997 (1.3e-02)	0.0268 (5.7e-03)
	CC	6.31e-05 (9.0e-05)	5.99e-05 (8.6e-05)	6.77e-05 (9.7e-05)	1 (2.2e-05)	-0.0011 (3.7e-04)
	FABIA	0.104 (1.8e-02)	0.104 (1.7e-02)	0.106 (1.7e-02)	1 (4.7e-04)	0.299 (4.9e-02)
	XMotifs	1.19e-06 (1.2e-05)	1.05e-06 (1.1e-05)	1.2e-06 (1.2e-05)	1 (4.5e-05)	-0.000119 (2.8e-04)
	ISA	0.00217 (2.2e-03)	0.00193 (1.9e-03)	0.00221 (2.2e-03)	1 (3.8e-05)	0.0158 (1.5e-02)
	GBC	0.476(1.6e-01)	0.506(1.3e-01)	0.847(7.3e-02)	0.983(1.1e-02)	0.693(8.7e-02)
	sGBC	0.696(1.2e-01)	0.714(1.0e-01)	0.993(9.9e-03)	0.989(6.4e-03)	0.838(6.0e-02)
15	Plaid	0.019 (1.3e-02)	0.0429 (1.4e-02)	0.163 (3.3e-02)	0.997 (1.2e-02)	0.0417 (3.3e-02)
	CC	4.12e-05 (7.1e-05)	4.01e-05 (7.0e-05)	4.36e-05 (7.5e-05)	1 (2.6e-05)	-0.0013 (3.2e-04)
	FABIA	0.101 (1.9e-02)	0.1 (1.8e-02)	0.104 (1.8e-02)	1 (7.4e-04)	0.286 (5.9e-02)
	XMotifs	5.09e-06 (3.2e-05)	4.71e-06 (3.0e-05)	5.17e-06 (3.3e-05)	1 (5.2e-05)	-0.000144 (3.7e-04)
	ISA	0.00235 (2.1e-03)	0.00204 (1.8e-03)	0.00239 (2.2e-03)	1 (4.7e-05)	0.0159 (1.4e-02)
	GBC	0.506(1.4e-01)	0.528(1.1e-01)	0.886(6.0e-02)	0.98(1.1e-02)	0.719(7.4e-02)
	sGBC	0.645(1.0e-01)	0.663(8.3e-02)	0.972(2.7e-02)	0.985(6.2e-03)	0.808(4.7e-02)

Real Data: AD proteomics dataset (continuous)

- The AMP-AD knowledge portal of the Synapse website (www.synapse.org) with ID syn3607470.
- Proteomics dataset include 6533 protein levels from 20 Alzheimer's Disease (AD) patients, 13 Asymptomatic Alzheimer's DIsease (AsymAD) patients, 14 controls.
- **Ground truth:** the status of each subject: AD/AsymAD/control.
- Biological information extracted from KEGG Pathway using Bioconductor package "KEGGgraph" and "KEGGREST".

Real Data: AD RNAseq dataset (count)

- The AMP-AD knowledge portal of the Synapse website (www.synapse.org) with ID syn5223705.
- Proteomics dataset include 64253 features from 82 AD patients, 84 progressive supranuclear palsy(PSP) patients, 28 pathologic aging(PA) subjects, and 77 elder controls.
- These measurements are from cerebellum RNA samples collected by the Mayo Clinic Brain Bank and Banner Sun Health Research Institute.
- **Ground truth:** the status of each subject: AD/PSP/PA/control.

Real Data: TCGA GBM Data (mixed)

- From the TCGA data portal, microarray gene expression data, DNA methylation data, and DNA copy number data for 233 Glioblastoma multiforme patients.
- DNA copy number data are dichotomized to 0 (normal) and 1 (abnomal).
- 48 genes from three critical signaling pathways - RPK/PI3K, p53, and Rb (migration, survival and apoptosis progression of cell cycles).
- The total number of features is $48 \times 3 = 144$.
- **Ground truth:** Kaplan-Meier imputed survival time, divided into four groups.

Analyses of Real Data: Results

Method	ASD: proteomics data		ASD: RNAseq data		GBM: mixed data	
	CE	CS	CE	CS	CE	CS
PLAID	0	0	0	0	0.263	0.175
CC	0.238	0.200	0.147	0.125	0.004	0.004
FABIA	0.254	0.140	0.147	0.103	0.260	0.186
xMotif	0.106	0.081	0	0	0	0
ISA	0.045	0.010	0.113	0.096	0.045	0.015
GBC	**0.313**	**0.167**	**0.239**	**0.211**	0.265	**0.263**
sGBC	**0.313**	0.160	**0.239**	**0.211**	**0.281**	0.221

Discussions

- Bayesian Generalized Biclustering Method: 1, applicable to data of multiple types; 2, incorporate existing biological information represented by a graph $\mathcal{G}$.
- Robust to mis-specification of biological information, $\mathcal{G}$
- Choice of L
- Li, Ziyi, Changgee Chang, Suprateek Kundu, and Qi Long. "Bayesian Generalized Biclustering Analysis via Adaptive Structured Shrinkage." in revision for Biostatistics.

  R code available at https://github.com/ziyili20/GBC.

Acknowledgments

This work is partly supported by NIH grants R01GM124111, R21NS091630 and P30CA016520.

Thank you!

References

Chang et al. (2018)
Scalable bayesian variable selection for structured high-dimensional data.
Biometrics, in press.

Hartigan. (1972)
Direct clustering of a data matrix.
Journal of the american statistical association **67**(337), 123–129.

Hochreiter et al. (2010)
FABIA: factor analysis for bicluster acquisition.
Bioinformatics **26**(12), 1520–1527.

Padilha et al. (2017)
A systematic comparative evaluation of biclustering techniques.
BMC bioinformatics **18**(1), 55.

Polson et al. (2013)
Bayesian inference for logistic models using pólya–gamma latent variables.
Journal of the American statistical Association **108**(504), 1339–1349.