

BENIGN 利用マニュアル

1. 概要

BENIGN は遺伝子ネットワーク解析の並列実行のためのプログラムであり、種々の条件下での遺伝子の発現データから、各条件下での遺伝子間の発現の依存関係を表すネットワークをベイジアンネットワークにより推定し、相互の比較を可能にします。

BENIGN の開発では、遺伝子ネットワークの推定部分は SiGN-BN をベースにしています。また、複数の遺伝子ネットワーク推定プロセスの並列実行機能の実現には MPIDP (GHOST-MP のパッケージに含まれています) をベースにしています。

SiGN-BN: <http://www.scls.riken.jp/scruise/software/sign-bn.html>

GHOST-MP: <http://www.scls.riken.jp/scruise/software/GHOST-MP.html>

2. benign コマンド

benign コマンドは MPI で並列実行されるプログラムです。このため実行には MPI の並列実行環境が必要です。Web で公開されているバイナリファイルを使用するときは、実行に先立って、`chmod +x benign` で実行可能権限のパーミッションを設定しておく必要があります。

benign コマンドの書式は以下の通りです。

```
$ mpiexec benign [-lg log] -tb input [user options]
```

指定可能なオプションについては以下の通りです。

`-lg log`

- 任意
- BENIGN の実行ログのファイル名を指定します。
- 指定がない場合、`mpidp.log` という名前でログを出力します。

`-tb input`

- 必須
- Job リストファイルを指定します。

`user options`

- 任意
- `-lg`, `-tb` 以外のオプションは解析処理へのパラメータとして渡されます。

3. Job リストファイル

Job リストファイルのフォーマットは以下の通りです。

```
[TITLE=タイトル]
パラメータ行 1
パラメータ行 2
...
```

TITLE 行の指定は任意であり、指定した場合、実行ログに記録されます。

パラメータ行には、空白区切りでパラメータを指定します。

以下は Job リストファイルの例です。

```
TITLE=example
benign -y --blocks 20 -o ex1 -N 1000 -L 1 -T 0.1 -S linear gene_exp1.edf
benign --blocks 20 -o ex2 -N 1000 -L 1 -T 0.1 -S linear gene_exp2.edf
benign --blocks 20 -o ex3 -N 1000 -L 1 -T 0.1 -S linear gene_exp3.edf
```

TITLE 行の内容（上の例では example）は実行ログファイルに記録されます。

以降の benign で始まる行では遺伝子ネットワーク推定処理の引数を指定します（行ごとに並列に実行されます）。

上の例の引数の意味は次の通りです。

-y ダイナミックベイジアンネットワークモデルを指定します（時系列データの時に使用できます）。省略したときはベイジアンネットワークモデルが選択されます。

--blocks n 時系列データの時点数を疑似的に n 倍に増やす（-y を指定した時は推奨）

-o name 出力結果のファイル名の接頭辞を name にセット

-N num ブートストラップサンプリングを num 回実行

-L 1 実行履歴を個別のファイルに出力（出力ファイル名は、name.log.XXXXXX となります。Name は -o で指定した接頭辞、XXXXXX は通し番号です。-L 1 を省略すると標準エラー出力にまとめられます）

-T 0.1 ブートストラップ確率が 0.1 以上の制御辺のみを出力

-S linear 遺伝子発現量を線形モデルで近似

詳細は SiGN-BN のマニュアル(<http://sign.hgc.jp/signbn/manual.html>)を参照して下さい。

4. EDF ファイルフォーマット

EDF (Expression Data Format)は遺伝子発現データを表現するために設計されたフォーマットです。タブまたはカンマ区切りのテキストファイルで、次の3つのセクションから構成されます。

1. メタデータセクション
2. 属性セクション
3. データセクション

以下は EDF ファイルフォーマットの例です。タブ区切りのテキストファイルとなっています。

```
# Meta Data Section
$Version 1.0
# Attribute Section
@PrimaryKeyGroupID 1 1 2 2 3 3
@SecondaryKeyGroupID 1 2 1 2 1 2
# Data Section
gene1 1.1 2.2 3.3 4.4 5.5 6.6
gene2 7.7 8.8 9.9 10.1 11.11 12.12
gene3 13.13 14.14 15.15 16.16 17.17 18.18
```

メタデータセクションでは、メタデータは"\$"から始まり、データセット、属性セクションにある属性についてのグローバルな情報を定義します。

属性セクションでは、属性は"@"から始まり、発現サンプルの属性を定義します。最初のカラムは属性のキーを表します。上の例では、2つの属性キーPrimaryKeyGroupID と SecondaryKeyGroupID が発現サンプルのために指定されています。SecondaryKeyGroupID は反復実験の回数を表すのに使います。上の例では発現データで n=2 の反復(replicate)で取得した場合を表しています。

データセクションにおいて、各行は遺伝子の発現データを表します。最初のカラムに遺伝子の名前を指定し、以降のカラムに発現データを設定します。

詳細は SiGN-BN の EDF ファイルの説明(<http://sign.hgc.jp/formats.html#edf>)を参照して下さい。

5. 実行結果

BENIGN の実行結果は、標準では SGN3 形式で出力されます。SGN3 形式は次の3つで構成されます。

1. ヘッダ情報 ([Information]セクション)
2. ノードデータ ([Nodes]セクション)
3. エッジデータ ([Edges]セクション)

以下は SGN3 形式の例です。

```
SiGN SGN3 FORMAT
[Information]
Edge Attr    BS.Prob:double    edgeScore:double    BS.Gain:double
             EdgeType:int BS.up/down:string    BS.EdgeTypeRatio:string
Node Attr    hubness:int    BS.HubIndex:double
X:BS.total:int    1000
[Nodes]
Cebpa Cebpa 0      1      0.110000
Cebpb Cebpb 1      8      4.853000
...
[Edges]
1  0  0.927000  0.927000  40.873586  0  up  1.00/0.00/0.00
3  0  0.199000  0.199000  21.401210  0  up  1.00/0.00/0.00
...
```

[Information]セクションでは、次の[Nodes]セクションと[Edges]セクションが持つ属性の一覧が表示されています。

[Nodes]セクションでは、ノードに関する次のような情報が表示されています。

- ノード名
- ノード番号
- 子ノードの数
- ブートストラップでの **HubIndex**

[Edges]セクションでは、エッジに関する次のような情報が表示されています。

- 親ノード番号
- 子ノード番号
- ブートストラップ確率
- 制御の符号 (up: 促進、down: 抑制)