Package ‘FesSelectionCUDA’

September 26, 2017

Type Package

Title Incremental Feature Selection Algorithm accelerated by GPU

Version 1.0

Date 2017-09-15

Depends R (>= 3.0) ,ROCR, parallel, doParallel, foreach

SystemRequirements Nvidia's CUDA toolkit (>= release 5.5)

Author Ruiquan Ge, Qinghan Meng, Goku

Maintainer goku<jym775096422@gmail.com>

Description There is a features selection algorithm to solve gene data which has a ``big p small n'' feature. Particularly, it selects subset which has both low feature numbers and high accuracies .Logistic regression classification algorithm is used for measuring chosen feature subsets.

License GPL

NeedsCompilation yes

Repository CRAN

R topics documented:

FesSelectionCUDA-package 2
f.ratio 2
getGPUIds 3
LRCUDA 3
rm.bias 4

Index **5**

FesSelectionCUDA-package *A gene feature selection algorithm.*

Description

There is a features selection algorithm to solve gene data which has a "big p small n" feature. Particularly, it selects subset which has both low feature numbers and high accuracies .Logistic regression classification algorithm is used for measuring chosen feature subsets.

Details

Package:	FesSelectionCUDA
Type:	Package
Version:	0.1
Date:	2017-09-25
License:	GPL-2

Author(s)

Ruiquan Ge,Qinghan Meng ,Goku Maintainer:Goku<jym775096422@gmail.com>

f.ratio *Screen features by fisher's ratio(statistic).*

Description

Fisher's ratio is a measure for discriminating power of some variable. We use it to detect useless features.

Usage

```
f.ratio(x, y, k)
```

Arguments

x	Gene expression profile matrix. The columns represent different prob sets. The rows represent different samples. The values in the matrix represent gene expression levels.
y	The response vector, positive(1) or negative(0).
k	Top k features are returned which are mostly correlated with response(y).

Value

A vector containing selected feature index.

`getGPUIds` *Get GPU device IDs from localhost or a cluster.*

Description

Query GPU device number and ID from the localhost or a cluster.

Usage

```
getGPUIds(cl = NULL)
```

Arguments

`cl` The cluster of computers which you created. If it is NULL, the program will execute in localhost.

Value

`Ids` A vector contains devices' IDs.

Examples

```
gpu.ids <- getGPUIds()
print(gpu.ids)
```

`LRCUDA` *An incremental feature selection algorithm for gene expression profile. We implement it using logistic regression which is accelerated by NVIDIA GPU technology.*

Description

A gene feature selection algorithm acclerating by NIVIDIA CUDA technology.

Usage

```
LRCUDA(x, y, n.comb = 2, error.threshhold = 0 , fold = 10, device.id = 0, cl = NULL)
```

Arguments

`x` Gene expression profile matrix. The columns represent different prob sets. The rows represent different samples. The values in the matrix represent gene expression levels.

`y` The label vector, positive(1) or negative(0)

`n.comb` The number of features which are selected.

`fold` The number of fold of cross validation used in the program.

`device.id` The GPU device ID.

`cl` The cluster of computers which you created. If it is NULL, the program will create clusters based on the number of devices automatically.

Value

`result` The result of feature selection . Each row in result represent the one combination of selected features and its log loss.

`rm.bias` *Remove the bias of selected features by iFes.*

Description

The features which selected by iFes have bias, we conduct multiple cross validation to screen the features again to make sure the generality of features.

Usage

```
rm.bias(result, x, y, fold, times, parallel = 1)
```

Arguments

`result` The data frame returned by iFes function.

`x` Gene expression profile matrix. The columns represent different prob sets. The rows represent different samples. The values in the matrix represent gene expression levels.

`y` The response vector, positive(1) or negative(0).

`fold` The number of fold to do cross validation.

`times` The repeated times of cross validation.

`parallel` The number of parallel progress to do the task.

Value

A data frame containing the average acc and auc for selected features by multiple cross validation.

Index

∗Topic **LRCUDA**
LRCUDA, 3
∗Topic **bias**
rm.bias, 4
∗Topic **gene, features, selection, CUDA**
FesSelectionCUDA-package, 2
∗Topic **gpu**
getGPUIds, 3
∗Topic **id**
getGPUIds, 3
∗Topic **ratio**
f.ratio, 2

f.ratio, 2
FesSelectionCUDA
*(*FesSelectionCUDA-package*)*, 2
FesSelectionCUDA-package, 2

getGPU *(*getGPUIds*)*, 3
getGPUIds, 3

LRCUDA, 3

rm.bias, 4