Package 'DDMarkerFAST'

August 1, 2016

Type Package

Title Diagnose and Detect Markers in Extracellular Circulating Feature Selection Method

Version 1.1

Date 2016-07-12

Author
Yu Shang, Qiong Yu, Huansheng Cao, Guoqing Liu, Xiufeng Liu, Hao Wu, Yan Wang, Ying Xu

Maintainer Yu Shang <yushang@uga.edu>

Correspond Qiong Yu <yuqiong@uga.edu>, Ying Xu <xyn@bmb.uga.edu>

Description Diagnose and Detect Markers in Extracellular Circulating is a homo sapiens deductive system solving the markers in extracellular circulating. It entails the symbols of markers, like the genes, the proteins, the micro RNAs, and the isoforms, whether can be diagnose and detect in extracellular circulating, especially the blood serum and the urine for the biological and medicine significance. With the help of a homo sapiens annotation database in DDMarkerData package, DDMarker can even diagnose and detect the sequence among the genes, the proteins, the micro RNAs, and the isoforms. There are two main function in this package, the ddmarker, and the MMC, short for Minimal Metabolize Circulation. MMC entails the markers among the minimal metabolize circulation. FAST is the feature selection method of DDMarker. It shows a way to predict your own data not only like DDMarker and cui's nucleic acids research did, but also can do it in 4 new ways: Support vector machines, the most stable method in biomarkers diagnosis; Adaboost, also known as the the most efficiency and accuracy method; Naive bayes cluster; Recursive partitioning and regression trees; Decision trees.

License GPL (>= 2)

LazyData TRUE

Depends R (>= 3.0.3), e1071, adabag, rpart, C50

Imports e1071, adabag, rpart, C50

LinkingTo e1071, adabag, rpart, C50

R topics documented:

DDMarkerFAST-package   *DDMarkerFAST*

Description

Diagnose and Detect Markers in Extracellular Circulating is a homo sapiens deductive system solving the markers in extracellular circulating. It entails the symbols of markers, like the genes, the proteins, the micro RNAs, and the isoforms, whether can be diagnose and detect in extracellular circulating, especially the blood serum and the urine for the biological and medicine significance. With the help of a homo sapiens annotation database in DDMarkerData package, DDMarker can even diagnose and detect the sequence among the genes, the proteins, the micro RNAs, and the isoforms. There are two main function in this package, the `ddmarker`, and the `DDMarkerMMC`, short for Minimal Metabolize Circulation. `DDMarkerMMC` entails the markers among the minimal metabolize circulation.

FAST package is the FeAture SelecTion method of DDMarker, the main function is DDMarkerFAST() and DDMarkerP(). The more details you can find in `DDMarkerFAST-method` and `DDMarkerP-method`

Details

Package:	DDMarkerFAST
Type:	Package
Version:	1.0
Date:	2016-07-12
Depends:	R (>= 3.0.3), e1071, adabag, rpart, C50
License:	GPL (>= 2)
LazyLoad:	yes
LazyData:	true

Author(s)

Yu Shang (JLU & UGA) <yushang@uga.edu>
Qiong Yu (JLU & UGA) <yuqiong@uga.edu> <yujoan_2001@163.com>
Huansheng Cao (UGA) <hshcao@uga.edu>
Guoqing Liu (IMUST & UGA) <gqliu@uga.edu> <gqliu1010@163.com>
Xiufeng Liu (GZUCM & UGA) <xfliu@uga.edu> <liu_xf@gzucm.edu.cn>
Hao Wu (BIT & UGA) <wuhao@uga.edu> <wuhao@bit.edu.cn>
Yan Wang (JLU & UGA) <wy6868@hotmail.com>
Ying Xu (JLU & UGA) <xyn@uga.edu> <xyn@bmb.uga.edu>

Maintainer: Yu Shang (JLU & UGA) <yushang@uga.edu>

References

citation("DDMarkerFAST");

[1] Juan Cui, et al. (2011) *An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer.* Nucleic Acids Research, 39: 1197-1207

[2] Juan Cui, et al. (2008) *Computational prediction of human proteins that can be secreted into the bloodstream.* BIOINFORMATICS, Vol.24 no. 20 2008 pages 2370-2375

[3] http://bioinfosrv1.bmb.uga.edu/DMarker/

[4] Breiman L. (1999) *Prediction games and arcing classifiers.* Neural Comput 11(7):1493:1517

[5] Breiman L, et al. (1984) *Classification and regression trees.* Wadsworth, Belmont

[6] CheungDW, et al. (1996) *Maintenance of discovered association rules in large databases: an incremental updating technique.* In: Proceedings of the ACM SIGMOD international conference on management of data, pp. 13:23

[7] Dietterich TG. (1997) *Machine learning: Four current directions.* AI Mag 18(4):97:136

[8] Domingos P. (1999) *MetaCost: A general method for making classifiers cost-sensitive.* In: Proceedings of the fifth international conference on knowledge discovery and data mining, pp 155:164

[9] Domingos P, et al. (1997) *On the optimality of the simple Bayesian classifier under zero-one loss.* Mach Learn 29:103:130

[10] Fix E, et al. (1951) *Discriminatory analysis, nonparametric discrimination.* USAF School of Aviation Medicine, Randolph Field, Tex., Project 21-49-004, Rept. 4, Contract AF41(128)-31, February 1951

[11] Freund Y, et al. (1997) *A decision-theoretic generalization of on-line learning and an application to boosting.* J Comput Syst Sci 55(1):119:139

[12] Friedman JH, et al. (1977) *An algorithm for finding best matches in logarithmic time.* ACM-Trans.Math. Software 3, 209. Also available as Stanford Linear Accelerator Center Rep. SIX-PUB-1549, February 1975

[13] Friedman JH, et al. (1996) *Lazy decision trees.* In: Proceedings of the thirteenth national conference on artificial intelligence, San Francisco, CA. AAAI Press/MIT Press, pp. 717:724

[14] Friedman N, et al. (1997) *Bayesian network classifiers.* Mach Learn 29:131:163

[15] Hand DJ, et al. (2001) *Idiot Bayes not so stupid after all.* Int Stat Rev 69:385:398

[16] Friedman J, et al. (2000) *Additive logistic regression: a statistical view of boosting with discussions.* Ann Stat 28(2):337:407

[17] Herbrich R, et al. (2000) *Rank boundaries for ordinal regression.* Adv Mar Classif pp 115:132

[18] Hunt EB, et al. (1966) *Experiments in induction.* Academic Press, New York

[19] Inokuchi A, et al. (2005) *General framework for mining frequent subgraphs from labeled graphs.* Fundament Inform 66(1-2):53:82

[20] Messenger RC, et al. (1972) *A model search technique for predictive nominal scale multivariate analysis.* J Am Stat Assoc 67:768:772

[21] Morishita S, et al. (2000) *Traversing lattice itemset with statistical metric pruning.* In: Proceedings of PODS 00, pp 226:236

[22] Olshen R. (2001) *A conversation with Leo Breiman.* Stat Sci 16(2):184:198

[23] Quinlan JR. (1979) *Discovering rules by induction from large collections of examples.* In: Michie D (ed), Expert systems in the micro electronic age. Edinburgh University Press, Edinburgh

[24] Quinlan R. (1989) *Unknown attribute values in induction.* In: Proceedings of the sixth international workshop on machine learning, pp. 164:168

[25] Quinlan JR. (1993) *C4.5: Programs for machine learning.* Morgan Kaufmann Publishers, San Mateo

[26] Reyzin L, et al. (2006) *How boosting the margin can also boost classifier complexity.* In: Proceedings of the 23rd international conference on machine learning.

See Also

DDMarkerFAST-method DDMarkerP-method DDMarker NAR

DDMarkerFAST	*DDMarker FeAture SelecTion*

Description

The Feature Selection Method of Diagnose and Detect Markers in Extracellular Circulating

results = DDMarkerFAST(data, . . .);

Arguments

data   Data matrix, the first column must contain the feature types, eg. BLOOD, URINE,
               . . . , and the following ones should be the features just like the references [1]
               and [2] did.

method   -method should be ("svm", "adaboost", "bayes", "cart", "c5"), "svm" denotes
               Support Vector Machines, the most stable method in biomarkers diagnosis,
               "adaboost" denotes AdaBoost, also known as the the most efficiency and accu-
               racy method, "bayes" denotes Naive Bayes Cluster, "cart" denotes Recursive Partitioning and
               "c5" denotes Decision Trees.
               default: "svm"

Details

models = DDMarkerFAST(data = data, method = "svm");

Value

The R function, DDMarkerFAST returns an object of list:

class	shows the method used.
names	shows the model attributes, which can be used in the prediction method, the more deta
vardep.summary, xlevels, ylevels	adaboost and cart will return the elements, the more details will be found in refere

Author(s)

Yu Shang (JLU & UGA) <yushang@uga.edu>
Qiong Yu (JLU & UGA) <yuqiong@uga.edu> <yujoan_2001@163.com>
Huansheng Cao (UGA) <hshcao@uga.edu>
Guoqing Liu (IMUST & UGA) <gqliu@uga.edu> <gqliu1010@163.com>
Xiufeng Liu (GZUCM & UGA) <xfliu@uga.edu> <liu_xf@gzucm.edu.cn>
Hao Wu (BIT & UGA) <wuhao@uga.edu> <wuhao@bit.edu.cn>
Yan Wang (JLU & UGA) <wy6868@hotmail.com>
Ying Xu (JLU & UGA) <xyn@uga.edu> <xyn@bmb.uga.edu>

Maintainer: Yu Shang (JLU & UGA) <yushang@uga.edu>

References

citation("DDMarkerFAST");

[1] Juan Cui, et al. (2011) *An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer.* Nucleic Acids Research, 39: 1197-1207

[2] Juan Cui, et al. (2008) *Computational prediction of human proteins that can be secreted into the bloodstream.* BIOINFORMATICS, Vol.24 no. 20 2008 pages 2370-2375

[3] http://bioinfosrv1.bmb.uga.edu/DMarker/

[4] Breiman L. (1999) *Prediction games and arcing classifiers.* Neural Comput 11(7):1493:1517

[5] Breiman L, et al. (1984) *Classification and regression trees.* Wadsworth, Belmont

[6] CheungDW, et al. (1996) *Maintenance of discovered association rules in large databases: an incremental updating technique.* In: Proceedings of the ACM SIGMOD international conference on management of data, pp. 13:23

[7] Dietterich TG. (1997) *Machine learning: Four current directions.* AI Mag 18(4):97:136

[8] Domingos P. (1999) *MetaCost: A general method for making classifiers cost-sensitive.* In: Proceedings of the fifth international conference on knowledge discovery and data mining, pp 155:164

[9] Domingos P, et al. (1997) *On the optimality of the simple Bayesian classifier under zero-one loss.* Mach Learn 29:103:130

[10] Fix E, et al. (1951) *Discriminatory analysis, nonparametric discrimination.* USAF School of Aviation Medicine, Randolph Field, Tex., Project 21-49-004, Rept. 4, Contract AF41(128)-31, February 1951

[11] Freund Y, et al. (1997) *A decision-theoretic generalization of on-line learning and an application to boosting.* J Comput Syst Sci 55(1):119:139

[12] Friedman JH, et al. (1977) *An algorithm for finding best matches in logarithmic time.* ACM-Trans.Math. Software 3, 209. Also available as Stanford Linear Accelerator Center Rep. SIX-PUB-1549, February 1975

[13] Friedman JH, et al. (1996) *Lazy decision trees.* In: Proceedings of the thirteenth national conference on artificial intelligence, San Francisco, CA. AAAI Press/MIT Press, pp. 717:724

[14] Friedman N, et al. (1997) *Bayesian network classifiers.* Mach Learn 29:131:163

[15] Hand DJ, et al. (2001) *Idiot Bayes not so stupid after all.* Int Stat Rev 69:385:398

[16] Friedman J, et al. (2000) *Additive logistic regression: a statistical view of boosting with discussions.* Ann Stat 28(2):337:407

[17] Herbrich R, et al. (2000) *Rank boundaries for ordinal regression.* Adv Mar Classif pp 115:132

[18] Hunt EB, et al. (1966) *Experiments in induction.* Academic Press, New York

[19] Inokuchi A, et al. (2005) *General framework for mining frequent subgraphs from labeled graphs.* Fundament Inform 66(1,2):53:82

[20] Messenger RC, et al. (1972) *A model search technique for predictive nominal scale multivariate analysis.* J Am Stat Assoc 67:768:772

[21] Morishita S, et al. (2000) *Traversing lattice itemset with statistical metric pruning.* In: Proceedings of PODS 00, pp 226:236

[22] Olshen R. (2001) *A conversation with Leo Breiman.* Stat Sci 16(2):184:198

[23] Quinlan JR. (1979) *Discovering rules by induction from large collections of examples.* In: Michie D (ed), Expert systems in the micro electronic age. Edinburgh University Press, Edinburgh

[24] Quinlan R. (1989) *Unknown attribute values in induction.* In: Proceedings of the sixth international workshop on machine learning, pp. 164:168

[25] Quinlan JR. (1993) *C4.5: Programs for machine learning.* Morgan Kaufmann Publishers, San Mateo

[26] Reyzin L, et al. (2006) *How boosting the margin can also boost classifier complexity.* In: Proceedings of the 23rd international conference on machine learning.

See Also

`DDMarkerFAST-package` `DDMarkerP-method` `DDMarker` `NAR`

Examples

```
models = DDMarkerFAST();
```

DDMarkerP   *DDMarker Prediction*

Description

The Prediction Method of Diagnose and Detect Markers in Extracellular Circulating

results = DDMarkerP(model, . . .);

Arguments

model   FAST class, should be the result computing by `DDMarkerFAST()`

data   Data matrix, which will be entailed whether can be diagnosed and detected in
                    extracellular circulating, just like the `references` did

Details

results = DDMarkerP(model = model, data = data);

Value

The R function, `DDMarkerP` returns an object of `vector`, and entails the results.

Author(s)

Yu Shang (JLU & UGA) <yushang@uga.edu>
Qiong Yu (JLU & UGA) <yuqiong@uga.edu> <yujoan_2001@163.com>
Huansheng Cao (UGA) <hshcao@uga.edu>
Guoqing Liu (IMUST & UGA) <gqliu@uga.edu> <gqliu1010@163.com>
Xiufeng Liu (GZUCM & UGA) <xfliu@uga.edu> <liu_xf@gzucm.edu.cn>
Hao Wu (BIT & UGA) <wuhao@uga.edu> <wuhao@bit.edu.cn>
Yan Wang (JLU & UGA) <wy6868@hotmail.com>
Ying Xu (JLU & UGA) <xyn@uga.edu> <xyn@bmb.uga.edu>

Maintainer: Yu Shang (JLU & UGA) <yushang@uga.edu>

References

citation("DDMarkerFAST");
[1] Juan Cui, et al. (2011) *An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer.* Nucleic Acids Research, 39: 1197-1207
[2] Juan Cui, et al. (2008) *Computational prediction of human proteins that can be secreted into the bloodstream.* BIOINFORMATICS, Vol.24 no. 20 2008 pages 2370-2375
[3] http://bioinfosrv1.bmb.uga.edu/DMarker/

See Also

DDMarkerFAST-package DDMarkerFAST-method DDMarker NAR

Examples

```
models = DDMarkerFAST();
results = DDMarkerP(models);
```

LOG	*LOG*

Description

The logs of Diagnose and Detect Markers in Extracellular Circulating

logs();

Details

logs();

Author(s)

Yu Shang (JLU & UGA) <yushang@uga.edu>
Qiong Yu (JLU & UGA) <yuqiong@uga.edu> <yujoan_2001@163.com>
Huansheng Cao (UGA) <hshcao@uga.edu>
Guoqing Liu (IMUST & UGA) <gqliu@uga.edu> <gqliu1010@163.com>
Xiufeng Liu (GZUCM & UGA) <xfliu@uga.edu> <liu_xf@gzucm.edu.cn>
Hao Wu (BIT & UGA) <wuhao@uga.edu> <wuhao@bit.edu.cn>
Yan Wang (JLU & UGA) <wy6868@hotmail.com>
Ying Xu (JLU & UGA) <xyn@uga.edu> <xyn@bmb.uga.edu>

Maintainer: Yu Shang (JLU & UGA) <yushang@uga.edu>

References

citation("DDMarkerFAST");

See Also

[DDMarkerFAST-package](#) [DDMarkerFAST-method](#) [DDMarkerP-method](#) [NAR](#)

Examples

```
logs();
# print the logs of DDMarker

## LOG
# DDMarker 1.0, 05/27/2015, Recomputed the blood and urine predicted biomarkers from the latest datebase;
# DDMarker 1.1, 08/27/2015, DDMarker can entails the sequence of proteins;
# DDMarker 1.2, 12/28/2015, adds a visualization function of biomarkers; isoform proteins can be diagnosed; st
# DDMarker 1.3, 01/12/2016, more illustrates are upload as the draft;
# DDMarker 1.4, 03/12/2016, adds the medical guidance in the results;
# DDMarker 1.4.1, 04/01/2016, bug fixed;
# DDMarker 1.4.2, 05/02/2016, bug fixed;
# DDMarker 1.4.3, 05/23/2016, bug fixed;
# DDMarker 1.5, 05/27/2016, bug fixed;
# DDMarker 2.0, 07/07/2016, DDMarker2 published;
# DDMarker 2.1, 07/12/2016, adds 4 methods in DDMarker FAST method;
# DDMarker 2.1.1, 07/26/2016 more illustrates are uploaded as the draft;
# DDMarker 2.1.2, 07/27/2016 Push a poster.
```

NAR   *NAR*

Description

The famous features, which were used in nucleic acids research and centers for disease control and prevention

nar();

Details

nar();

Author(s)

Yu Shang (JLU & UGA) <yushang@uga.edu>
Qiong Yu (JLU & UGA) <yuqiong@uga.edu> <yujoan_2001@163.com>
Huansheng Cao (UGA) <hshcao@uga.edu>
Guoqing Liu (IMUST & UGA) <gqliu@uga.edu> <gqliu1010@163.com>
Xiufeng Liu (GZUCM & UGA) <xfliu@uga.edu> <liu_xf@gzucm.edu.cn>
Hao Wu (BIT & UGA) <wuhao@uga.edu> <wuhao@bit.edu.cn>
Yan Wang (JLU & UGA) <wy6868@hotmail.com>

Ying Xu (JLU & UGA) <xyn@uga.edu> <xyn@bmb.uga.edu>

Maintainer: Yu Shang (JLU & UGA) <yushang@uga.edu>

References

citation("DDMarkerFAST");
[1] Juan Cui, et al. (2011) *An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer* Nucleic Acids Research, 39: 1197-1207
[2] Juan Cui, et al. (2008) *Computational prediction of human proteins that can be secreted into the bloodstream* BIOINFORMATICS, Vol.24 no. 20 2008 pages 2370-2375
[3] http://bioinfosrv1.bmb.uga.edu/DMarker/

See Also

DDMarkerFAST-package DDMarkerFAST-method DDMarkerP-method DDMarker

Examples

```
nar();
# print the famous features, which were used in Nucleic Acids Research and Centers for Disease Control and Pre
# See the references

## Features used in DDMarker:
##
# http://bidd.cz3.nus.edu.sg/cgi-bin/prof/protein/profnew.cgi
# http://www.expasy.org/proteomics/protein_structure
# http://molbiol-tools.ca/Protein_secondary_structure.htm
# http://www.cbs.dtu.dk/services/
# http://coot.embl.de/cgi/sscp_serv.pl
# http://phobius.sbc.su.se/cgi-bin/predict.pl
# http://bip.weizmann.ac.il/fldbin/findex
# http://www.cbs.dtu.dk/services/TatP/
# http://bmbpcu36.leeds.ac.uk/~andy/betaBarrel/AACompPred/aaTMB_Hunt.cgi
# http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi
# http://www.cbs.dtu.dk/services/NetNGlyc/
# http://www.cbs.dtu.dk/services/NetOGlyc/
# http://www.scfbio-iitd.res.in/software/proteomics/rg.jsp
# http://web.expasy.org/cgi-bin/compute_pi/pi_tool
```

Index