R documentation

of 'predictChemPC.Rd'

December 1, 2013

---

predictChemPC *predictChemPC*

Description

This function performs a prediction method from 4 predicting potent compounds methods of this package. These methods are Random selection, EI selection, 1NN selection and GP selection.

Usage

```
predictChemPC( trainData, testData, targetData, method, loghyper = NULL)
```

Arguments

trainData	trainData specifies a data frame including an array of a data with dimension of n*m. This data is same data used in training.
testData	testData specifies a data frame including an array of training data with dimension of p*m. m is same with number of m in trainData.
targetData	targetData is a one dimensional array with n rows which is equal to number of rows in trainData.
method	method a string value to specify prediction method. Its value can be random, 1NN, EI or GP.
loghyper	loghyper is calculated in a train method of EI or GP , they are to use EI or GP prediction method.

Details

This function withholds 4 methods to predict potent compounds.

`method` is one of:

`random` One compound will be selected randomly and added to train data each time.

`1NN` The compound for which is nearest (based on Tonimito Coefficient) to the most potent compound in training data is selected and added to train data.

`EI` a compound for which maximum expected improvement is reached, is selected and then it is added to train data.

`GP` a compound holding maximum potency in test data is selected.

`Feature selection` Feature selection employed in this package is based on Spearman Rank Correlation such that before each training step those attributes in which revealed a significant Spearman rank correlation with the logarithmic potency values (q-value < 5 are computed from original p-values via the multiple testing correction method by Benjamini and Hochberg.

Value

It returns a predicted value of a selected method.

Author(s)

Mohsen Ahmadi

References

1. Predicting Potent Compounds via Model-Based Global Optimization, Journal of Chemical Information and Modeling, 2013, 53 (3), pp 553-559, M Ahmadi, M Vogt, P Iyer, J Bajorath, H Froehlich. 2. Software MOE is used to calculate the numerical descriptors in data sets. Ref: http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm 3. ChEMBL was the source of the compound data and potency annotations in data sets. Ref: https://www.ebi.ac.uk/chembl/

Examples

```
library(gpr)
library(SimuChemPC)
a =  as.data.frame(array(1:10, dim=c(2,5)))
b = as.data.frame(array(5:15, dim=c(2,5)))
c = array(1:2)
myindex = predictChemPC( a, b, c, "random", NULL)
```

Index