Package 'optSelect'

November 9, 2020

Type Package

Title optSelect: an R package for ensemble feature selection and stabil-ity assessment

Version 0.0.1

Date 2020-11-01

Author Karan Uppal

Maintainer Karan Uppal <kuppal3gt@gmail.com>

Description optSelect, a multi agent-based stochastic optimization approach for ensemble feature selection. Stage one involves function perturbation, where ranked list of features are generated using different feature selection methods and stage two involves data perturbation, where feature selection is performed within randomly selected learning sets of the training data. The agents are assigned to different behavior states and move according to a binary PSO algorithm. A multi-objective fitness function is used to evaluate the classifica-tion accuracy of the agents.

License GPL2.0

LazyLoad yes

Depends R (>= 3.0), snow,e1071,yaImpute,pROC,bioDist,RankAggreg,CMA,expm

R topics documented:

`optSelect-package` *optSelect*

Description

optSelect, a multi agent-based stochastic optimization approach for ensemble feature selection. Stage one involves function perturbation, where ranked list of features are generated using different feature selection methods and stage two involves data perturbation, where feature selection is performed within randomly selected learning sets of the training data. The agents are assigned to different behavior states and move according to a binary PSO algorithm. A multi-objective fitness function is used to evaluate the classifica-tion accuracy of the agents.

Details

Package:	optSelect
Type:	Package
Version:	0.0.1
Date:	2020-11-01
License:	gpl2.0
LazyLoad:	yes

Author(s)

Karan Uppal Maintainer: <kuppal3gt@gmail.com>

`run_pso` *run_pso*

Description

This function uses a multi agent-based stochastic optimization approach for ensemble feature selection. Stage one involves function perturbation, where ranked list of features are generated using different feature selection methods and stage two involves data perturbation, where feature selection is performed within randomly selected learning sets of the training data. The agents are assigned to different behavior states and move according to a binary PSO algorithm. A multi-objective fitness function is used to evaluate the classifica-tion accuracy of the agents.

Usage

```
run_pso(trainm, trainclass, testm, testclass, outloc, transition_matrix, c1 = 2.05, c2 = 2.05, itr =
```

Arguments

`trainm`	A n x p data matrix with training data, where n is the number of samples in the training set and p is the number of variables
`trainclass`	A n x 1 vector with class labels for instances in the training set
`testm`	A m x p data matrix with training data, where m is the number of samples in the test set and p is the number of variables
`testclass`	A m x 1 vector with class labels for instances in the test set
`outloc`	Output folder location
`maxnum`	Maximum number of features to select (e.g. maxnum=5)

Value

Returns a list

`scoringmatrix`	Binary matrix with 0 (not selected) or 1 (selected) assignment in each iteration
`bestfeatlist`	Indices of selected features

`bestfeatnames`	Names of selected features
`trainm.new`	Updated training set with only selected features
`testm.new`	Updated test set with only selected features
`trainacc`	Classification accuracy using the optimal set of features in the training set
`testacc`	Classification accuracy using the optimal set of features in the test set
`testauc`	AUC using the optimal set of features in the test set

Author(s)

Karan Uppal

Index