ssl-cotraining-example

February 21, 2018

1 Package 'SSL'

Type Package Title Semi-Supervised Learning Version 0.1 Date 2016-05-01 Author Junxiang Wang Maintainer Junxiang Wang xianggebenben@163.com Description Semi-supervised learning has attracted the attention of machine learning community because of its high accuracy with less annotating effort compared with supervised learning.The question that semi-supervised learning wants to address is: given a relatively small labeled dataset and a large unlabeled dataset, how to design classification algorithms learning from both ? This package is a collection of some classical semi-supervised learning algorithms in the last few decades. License GPL (>= 3) LazyData TRUE RoxygenNote 5.0.1 Depends R (>= 3.2) Imports NetPreProc (>= 1.1), Rcpp (>= 0.12.2), caret (>= 6.0-52), proxy (>= 0.4-15), xgboost (>= 0.4), klaR (>= 0.6-12), e1071 (>= 1.6-7), stats (>= 3.2) LinkingTo Rcpp NeedsCompilation yes Repository CRAN Date/Publication 2016-05-14 23:12:09

1.1 Algoritmos usados en este notebook

1. sslCoTrain

```
In [20]: library(SSL)
         library(caret)
```

```
In [21]: conf_matrix_statistics <- function(yl, yu, ytrue, known.label) {
             ytrue_l<-ytrue[known.label]
             ytrue_u<-ytrue[-known.label]

             xtab <- table(c(yl, yu), c(ytrue_l, ytrue_u))
             results <- confusionMatrix(xtab)
             print(prop.table(as.matrix(results), margin = 1))
             print(as.matrix(results, what = "overall"))
             print(as.matrix(results, what = "classes"))
             return(results)
         }
```

2 Cargando datos

```
In [22]: data(iris)
         x<-iris[,1:4]
         ytrue<-as.numeric(iris[,5])
         #Suppose we know the first "num_labelled" observations of each class
         #and we want to predict the remaining with co-training
         # 1 setosa, 2 versicolor, 3 virginica
         num_labelled = 10
         yl<-rep(1:3,each=num_labelled)
         known.label <-c(1:num_labelled,51:(50+num_labelled),101:(100+num_labelled))
         xu<-x[-known.label,]
         xl<-x[known.label,]

In [23]: # Co-training
         yu<-sslCoTrain(xl,yl,xu,method1="xgb",nrounds1 = 100,method2="xgb",nrounds2 = 100,n=60

         stats <- conf_matrix_statistics(yl, yu, ytrue, known.label)
```

```
  1    2    3
1 1 0.00 0.00
2 0 0.92 0.08
3 0 0.08 0.92
                       [,1]
Accuracy       9.466667e-01
Kappa          9.200000e-01
AccuracyLower  8.976181e-01
AccuracyUpper  9.766962e-01
AccuracyNull   3.333333e-01
AccuracyPValue 3.741810e-57
McnemarPValue           NaN
                             1         2         3
Sensitivity          1.0000000 0.9200000 0.9200000
Specificity          1.0000000 0.9600000 0.9600000
Pos Pred Value       1.0000000 0.9200000 0.9200000
Neg Pred Value       1.0000000 0.9600000 0.9600000
Precision            1.0000000 0.9200000 0.9200000
Recall               1.0000000 0.9200000 0.9200000
F1                   1.0000000 0.9200000 0.9200000
Prevalence           0.3333333 0.3333333 0.3333333
Detection Rate       0.3333333 0.3066667 0.3066667
Detection Prevalence 0.3333333 0.3333333 0.3333333
Balanced Accuracy    1.0000000 0.9400000 0.9400000
```