Question
Joerg Meier · Nov 2, 2016

iKnow - Text categorization "Category 1 covers the whole dataset"

Hi,

I am using iknow text categorization to classify texts. I have 11 medical articles as my training set. Here is part of the source code:

```
ListerAndLoader
  SET domId=domoref.Id
  SET flister=##class(%iKnow.Source.File.Lister).%New(domId)
  SET myloader=##class(%iKnow.Source.Loader).%New(domId)
UseLister
  SET dirpath = "D:\iKnowTestCase\SmallDataBase\Medical"
  SET stat = myloader.SetLister(flister)
  SET stat = myloader.ProcessList(dirpath,$LB("txt"),0,"")
  IF stat '= 1 {WRITE "The lister failed: ",$System.Status.DisplayError(stat) QUIT }
TrainingSet
  SET tTrainingSet = ##class(%iKnow.Filters.RandomFilter).%New(domId, 0.7)
  SET tTestSet = ##class(%iKnow.Filters.GroupFilter).%New(domId, "AND", 1) // NOT filter
  DO tTestSet.AddSubFilter(tTrainingSet)
  SET numSrcFD=##class(%iKnow.Queries.SourceAPI).GetCountByDomain(domId,tTrainingSet)
  WRITE "The training set includes ",numSrcFD," sources",!
  SET numSrcTD=##class(%iKnow.Queries.SourceAPI).GetCountByDomain(domId,tTestSet)
  WRITE "The test set includes ",numSrcTD," sources",!
TextClassifier
  SET tBuilder = ##class(%iKnow.Classification.IKnowBuilder).%New("MIT",tTrainingSet)
  SET tBuilder.ClassificationMethod = "naiveBayes"

  SET mstat = tBuilder.%AddCategory("Medical","",5)
  SET stat = tBuilder.%GetCategoryInfo(.pcategories)
  WRITE $LISTTOSTRING(pcategories(1)),!
  SET stat = tBuilder.%CreateClassifierClass("MIT.Classifier",1,1,1,1)

Optimizer
  WRITE "In the Optimizer",!
  SET tOpt = ##class(%iKnow.Classification.Optimizer).%New(domId,tBuilder)
  WRITE "optimizer is ",tOpt,!
  SET tOpt.ScoreMetric="MicroPrecision"
  WRITE "ScoreMetric is ",tOpt.ScoreMetric,!
  DO ##class(%iKnow.Queries.EntityAPI).GetTop(.result,domId,1,50)
  WRITE "load terms ",tOpt.LoadTermsArray(.result),!
  SET optstat=tOpt.Optimize(5)
  WRITE "Optimize status: ",optstat,!
  WRITE "End of Optimizer",!!
  DO ##class(%iKnow.Classification.Utils).%RunModelFromDomain(.r,
    "MIT.Classifier",domId,"helicopter",tTestSet)
  SET i=1
  WHILE $DATA(r(i)) {
      WRITE $LISTTOSTRING(r(i),",")
      SET i=i+1 }
TestClassifier
```

```
WRITE tBuilder.%TestClassifier(tTestSet,.testresult,.accuracy),!
WRITE "model accuracy: ",$FNUMBER(accuracy*100,"L",2)," percent",!
SET n=1
SET wrongcnt=0
WHILE $DATA(testresult(n)) {
 IF $LISTGET(testresult(n),2) '= $LISTGET(testresult(n),3) {
  SET wrongcnt=wrongcnt+1
  WRITE "WRONG: ",$LISTGET(testresult(n),1)
  WRITE " actual ",$LISTGET(testresult(n),2)
  WRITE " pred. ",$LISTGET(testresult(n),3),! }
SET n=n+1 }
WRITE wrongcnt," out of ",n-1,!
```

When I run the function: tBuilder.%CreateClassFierClass("MIT.Classfier", 1,1,1,1)

I got the error: category 1 covers the whole data. I can not find this in the documentation.

Thank you.

#Analytics #ObjectScript #iKnow

Source URL:https://community.intersystems.com/post/iknow-text-categorization-category-1-covers-whole-dataset