

run_analysis.R

Jie

Fri Feb 15 14:19:52 2019

```
## load necessary packages
library(data.table)      ## load library for data.table(), fread()
library(dplyr)           ## load library for select()

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(reshape2)        ## load library for melt()

##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##   dcast, melt
library(tidyr)           ## load library for separate()

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:reshape2':
##
##   smiths

## download data file
url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
download.file(url, "dataFiles.zip")
## unzip and list files
unzip(zipfile = "dataFiles.zip")
list.files('./UCI HAR Dataset',recursive=TRUE)

## [1] "activity_labels.txt"
## [2] "features.txt"
## [3] "features_info.txt"
## [4] "README.txt"
## [5] "test/Inertial Signals/body_acc_x_test.txt"
## [6] "test/Inertial Signals/body_acc_y_test.txt"
## [7] "test/Inertial Signals/body_acc_z_test.txt"
## [8] "test/Inertial Signals/body_gyro_x_test.txt"
```

```

## [9] "test/Inertial Signals/body_gyro_y_test.txt"
## [10] "test/Inertial Signals/body_gyro_z_test.txt"
## [11] "test/Inertial Signals/total_acc_x_test.txt"
## [12] "test/Inertial Signals/total_acc_y_test.txt"
## [13] "test/Inertial Signals/total_acc_z_test.txt"
## [14] "test/subject_test.txt"
## [15] "test/X_test.txt"
## [16] "test/y_test.txt"
## [17] "train/Inertial Signals/body_acc_x_train.txt"
## [18] "train/Inertial Signals/body_acc_y_train.txt"
## [19] "train/Inertial Signals/body_acc_z_train.txt"
## [20] "train/Inertial Signals/body_gyro_x_train.txt"
## [21] "train/Inertial Signals/body_gyro_y_train.txt"
## [22] "train/Inertial Signals/body_gyro_z_train.txt"
## [23] "train/Inertial Signals/total_acc_x_train.txt"
## [24] "train/Inertial Signals/total_acc_y_train.txt"
## [25] "train/Inertial Signals/total_acc_z_train.txt"
## [26] "train/subject_train.txt"
## [27] "train/X_train.txt"
## [28] "train/y_train.txt"

#####
# 1. merge training and test set
#####
## Load train datasets
subjectTrain <- fread("./UCI HAR Dataset/train/subject_train.txt")
subjectTest <- fread("./UCI HAR Dataset/test/subject_test.txt")
activityTrain <- fread("./UCI HAR Dataset/train/Y_train.txt")
activityTest <- fread("./UCI HAR Dataset/test/Y_test.txt")
XTrain <- fread("./UCI HAR Dataset/train/X_train.txt")
XTest <- fread("./UCI HAR Dataset/test/X_test.txt")
## combine test and train subject
subjectDT <- rbind(subjectTrain, subjectTest)
setnames(subjectDT, "V1", "subjectCode")
## combine test and train activity
activityDT <- rbind(activityTrain, activityTest)
setnames(activityDT, "V1", "activityCode")
## combine X feature codes
DT <- rbind(XTrain, XTest)
## merge columns
DT <- cbind(subjectDT, DT)
DT <- cbind(DT, activityDT)

#####
# 2. extract only the mean and standard deviation for each mesure
#####
## get the feature code table
featureDT <- fread("./UCI HAR Dataset/features.txt",
                    col.names=c("indx", "feature"))
## extract the mean and standard deviation measures
featureDT <- featureDT[grepl('mean\\(\\(\\)|std\\(\\(\\)', feature)]
## add a column of feature code
featureDT$featureCode <- featureDT[,paste0('V',indx)]

```

```
DTmeanstd <- select(DT,c('subjectCode',featureDT$featureCode,'activityCode'))
```

```
#####
# 3. uses descriptive activity names to name the activities
#####
```

```
activitylabel <- fread("./UCI HAR Dataset/activity_labels.txt")
setnames(activitylabel, names(activitylabel), c("indx", "activity"))
```

```
#####
# 4. label data set using descriptive activity names
#####
```

```
## label activity
DTmeanstdlabeled <- merge(DTmeanstd,activitylabel,by.x='activityCode',
                          by.y='indx',all.x=TRUE)
setkey(DTmeanstdlabeled,subjectCode,activityCode,activity)
## melt to add a column of feature code
DTtbl <- melt(DTmeanstdlabeled,key(DTmeanstdlabeled),
              variable.name='featureCode')
```

```
## label feature
DTl <- merge(DTtbl,featureDT,by='featureCode',all.x=TRUE)
```

```
#####
# 5. find the average of each variable for each activity and each subject
#####
```

```
## remove brackets
DTl$feature <- gsub("\\(\\)", "",DTl$feature)
## check the following command for separate result
## separate(featureDT,feature,c('measure','type','axis'),sep='-')
DTtidy <- separate(DTl,feature,c('measure','type','axis'),sep='-')
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 185382 rows
## [144187, 144188, 144189, 144190, 144191, 144192, 144193, 144194, 144195,
## 144196, 144197, 144198, 144199, 144200, 144201, 144202, 144203, 144204,
## 144205, 144206, ...].
```

```
DTtidy <- DTtidy[,c('subjectCode','activity','measure','type','axis','value')]
## check result
## head(DT[,c('subjectCode','activityCode','V1')])
## head(DTtidy[DTtidy$activity=='STANDING',])
DTstats <- DTtidy[, list(count=.N, average=mean(value)), by=c('subjectCode',
                      'activity','measure','type','axis')]
head(DTstats)
```

```
##   subjectCode      activity measure type axis count  average
## 1:          1      WALKING tBodyAcc mean    X    95 0.2773308
## 2:          1 WALKING_UPSTAIRS tBodyAcc mean    X    53 0.2554617
## 3:          1 WALKING_DOWNSTAIRS tBodyAcc mean    X    49 0.2891883
## 4:          1      SITTING tBodyAcc mean    X    47 0.2612376
## 5:          1      STANDING tBodyAcc mean    X    53 0.2789176
## 6:          1      LAYING tBodyAcc mean    X    50 0.2215982
```

```
## output data table  
write.table(DTstats,file='./datatable.txt',row.name=FALSE)
```