

Human Activity Recognition

Peggy Courtois

8 August 2017

Abstract

This work reports the model used to predict the quality of an exercise done by individuals carrying portable devices.

Introduction

We usually quantify how much of a particular activity is done, but rarely assess the quality of this exercise. In this report, we use data from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants, to predict the manner in which the exercise is done. The participants were asked to perform barbell lifts correctly and incorrectly in 5 different ways. More information is available from the website here, in the section “Weight Lifting Exercise Dataset”.

Weight lifting exercises dataset

Description

We define quality of execution and investigate three aspects that pertain to qualitative activity recognition:

- the problem of specifying correct execution
- the automatic and robust detection of execution mistakes
- how to provide feedback on the quality of execution to the user.

Six young health participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different fashions:

- class A: exactly according to the specification
- class B: throwing the elbows to the front
- class C: lifting the dumbbell only halfway
- class D: lowering the dumbbell only halfway
- class E: throwing the hips to the front

Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes. Participants were supervised by an experienced weight lifter to make sure the execution complied to the manner they were supposed to simulate. The exercises were performed by six male participants aged between 20-28 years, with little weight lifting experience. We made sure that all participants could easily simulate the mistakes in a safe and controlled manner by using a relatively light dumbbell (1.25kg). We calculated features on the Euler angles (roll, pitch and yaw), as well as the raw accelerometer, gyroscope and magnetometer readings. For the Euler angles of each of the four sensors we calculated eight features: mean, variance, standard deviation, max, min, amplitude, kurtosis and skewness.

Cleaning

The training data file is pml-training.csv and the test data file is pml-testing.csv

The training dataset contains 19622 observations and 160 variables. Over the 160 variables, only 54 were kept for the following work. The variables we removed were composed of more than 95% of NAs. We also removed unnecessary variables such as the names of the participants, or the date of the practice.

```
## Read the data
data <- read.csv(file = "pml-training.csv")

## Discard the variables with too many NAs:
var_names <- names(data) #list of variable names
#Function calculating the number of NAs in a variable:
na_count <- function(var) {
  count <- sum(is.na(var)=="TRUE")
  return(count)
}

count <- 0
keeps <- ""
for(i in 1:length(var_names)){
  if(is.factor(data[[var_names[i]]])=="TRUE"){
    temp <- as.numeric(as.character(data[[var_names[i]]]))
  } else{
    temp <- data[[var_names[i]]]
  }
  count[i] <- na_count(temp)
  if(count[i]/19622 < 0.95){
    keeps <- c(keeps,var_names[i])
  }
}
# We remove the unwanted variables (time, participants' names)
keeps <- keeps[-c(1:5)]
# The final data:
final_data <- data[keeps]
final_data$classe <- data$classe

## Visualization of the data:
dim(final_data)
summary(final_data)
```

Random Forest Prediction Model

Cross Validation

First, we split our data into training and testing sets.

```
## Split the data into training and testing datasets.
library("caret")
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
inTrain <- createDataPartition(y=final_data$classe, p=0.7, list=FALSE)
training <- final_data[inTrain,]
testing <- final_data[-inTrain,]
```

We use the random forest prediction model for its better accuracy compared to other models.

```
# Prediction model:
library(caret)
library(randomForest)

## randomForest 4.6-12
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:ggplot2':
##
##     margin
set.seed(33833)
#modfit <- train(classe ~ ., data = training, method = "rf") #too time consuming
modfit <- randomForest(classe ~., data=training)
predic <- predict(modfit, testing)
confusionMatrix(predic,testing$classe)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##           A 1671   10    0    0    0
##           B    2 1126    8    0    0
##           C    0    3 1018   21    1
##           D    0    0    0  943    3
##           E    1    0    0    0 1078
##
## Overall Statistics
##
##               Accuracy : 0.9917
##               95% CI : (0.989, 0.9938)
##       No Information Rate : 0.2845
##       P-Value [Acc > NIR] : < 2.2e-16
##
##               Kappa : 0.9895
##  McNemar's Test P-Value : NA
##
## Statistics by Class:
##
##               Class: A Class: B Class: C Class: D Class: E
## Sensitivity           0.9982  0.9886  0.9922  0.9782  0.9963
## Specificity           0.9976  0.9979  0.9949  0.9994  0.9998
## Pos Pred Value        0.9941  0.9912  0.9760  0.9968  0.9991
## Neg Pred Value        0.9993  0.9973  0.9983  0.9957  0.9992
## Prevalence            0.2845  0.1935  0.1743  0.1638  0.1839
## Detection Rate        0.2839  0.1913  0.1730  0.1602  0.1832
## Detection Prevalence  0.2856  0.1930  0.1772  0.1607  0.1833
```

```
## Balanced Accuracy      0.9979   0.9932   0.9935   0.9888   0.9980
```

We obtain an accuracy of 99.47% of accuracy.

Final application

First, we clean and pre-process the data the same way we did for the training data.

```
## Read the data
data2 <- read.csv(file = "pml-testing.csv")
final_data2 <- data2[keeps]
```

```
## Visualization of the data:
```

```
#dim(final_data2)
#summary(final_data2)
```

```
predic2 <- predict(modfit, final_data2)
predic2
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
## Levels: A B C D E
```

Acknowledgment

The data for this project come from this source: <http://groupware.les.inf.puc-rio.br/har>. If you use the document you create for this class for any purpose please cite them as they have been very generous in allowing their data to be used for this kind of assignment.

Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013.

Read more: <http://web.archive.org/web/20161224072740/http://groupware.les.inf.puc-rio.br/har#ixzz4p7qCIbCe>