M11\_Assignment

Xueyi Fan

July 19, 2016

Assignment:

If (and only if) you can’t use some form of Time series analysis help in your research project, then apply a form of Time series analysis to the data the Twitter time series data set M11\_Tweets\_Miley\_Nikki\_Taylor.csv (it is online at ‘<http://nikbearbrown.com/YouTube/MachineLearning/M11/M11_Tweets_Miley_Nikki_Taylor.csv’>). Does it help? Note you only need to use ONE forecasting approach from Module 11, so there will be only ONE assignment for all the modules and the same assignment for all the modules.

# Answer:

My project is to use only sequencing data and couldn't use one of the Time series analysis. So here I will use the data the Twitter time series data set [M11\_Tweets\_Miley\_Nikki\_Taylor.csv](http://nikbearbrown.com/YouTube/MachineLearning/M11/M11_Tweets_Miley_Nikki_Taylor.csv)

## Loading the data

tweets\_url <- "http://nikbearbrown.com/YouTube/MachineLearning/M11/M11\_Tweets\_Miley\_Nikki\_Taylor.csv"  
  
tweets <- read.csv(url(tweets\_url), header=F)  
names(tweets) <- c("unknow", "Time", "tweets")  
  
head(tweets)

## unknow Time  
## 1 6.60163E+17 Fri Oct 30 18:33:50 +0000 2015  
## 2 6.60163E+17 Fri Oct 30 18:33:51 +0000 2015  
## 3 6.60163E+17 Fri Oct 30 18:33:52 +0000 2015  
## 4 6.60163E+17 Fri Oct 30 18:33:54 +0000 2015  
## 5 6.60163E+17 Fri Oct 30 18:33:55 +0000 2015  
## 6 6.60163E+17 Fri Oct 30 18:33:56 +0000 2015  
## tweets  
## 1 https://t.co/V2kb7Qk6VD Check out my #remix here for #taylorswift if you didnt catch it already.  
## 2 Nicki Minaj Strips Down To Bra &amp;amp; Thong For 2015 Calendar!... https://t.co/G5qZeKCTEu https://t.co/8Yz9LRhYbk  
## 3 RT @FactsAboutNM: Nicki Minaj is on Jhene Aiko's New Album featured. Imagine the SLAYAGE https://t.co/KWrhdtNann  
## 4 Taylor Swift sings the right songs DO NOT ARGUE  
## 5 You taught me about your past, thinking your future was me. - Taylor Swift  
## 6 RT @SelfLearnings: "I can't say hello to you and risk another goodbye." - Taylor Swift

str(tweets)

## 'data.frame': 95939 obs. of 3 variables:  
## $ unknow: Factor w/ 1694 levels ""," @NICKIMINAJ",..: 49 49 49 49 49 49 49 49 49 49 ...  
## $ Time : Factor w/ 73776 levels "","Fri Nov 06 18:21:03 +0000 2015",..: 11330 11331 11332 11333 11334 11335 11336 11337 11338 11338 ...  
## $ tweets: Factor w/ 50203 levels ""," @MTV used to be decient now there is cuss words, Nipple slip by @MileyCyrus I will not be watch mtv no more. https://t.co/biYo"| \_\_truncated\_\_,..: 23559 30812 35668 46430 50006 42120 22098 48655 24324 45290 ...

library("RCurl")

## Loading required package: bitops

library("plyr")

## Warning: package 'plyr' was built under R version 3.2.5

library("forecast")

## Warning: package 'forecast' was built under R version 3.2.5

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.2.5

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: timeDate

## This is forecast 7.1

split\_time <- function(x){  
 time\_list <- strsplit(x, "[ ]")[[1]]  
 time\_list <- t(as.data.frame(time\_list))  
 time\_list  
}  
  
  
time <- as.data.frame(tweets[,2])  
  
new\_time<- apply(time, 1, split\_time)  
  
new\_time<- data.frame(matrix(unlist(new\_time), ncol=6, byrow = T))  
colnames(new\_time) <- c("week", "month", "day", "hr:mn:sec", "year")  
head(new\_time)

## week month day hr:mn:sec year NA  
## 1 Fri Oct 30 18:33:50 +0000 2015  
## 2 Fri Oct 30 18:33:51 +0000 2015  
## 3 Fri Oct 30 18:33:52 +0000 2015  
## 4 Fri Oct 30 18:33:54 +0000 2015  
## 5 Fri Oct 30 18:33:55 +0000 2015  
## 6 Fri Oct 30 18:33:56 +0000 2015

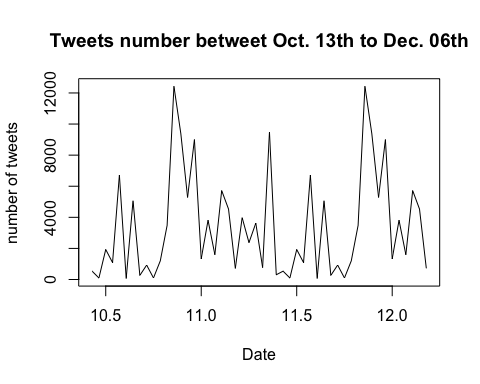
dim(new\_time)

## [1] 95817 6

seismic <- count(new\_time, c("month","day"))  
  
seismic$month <- gsub("Oct", 10, seismic[,1])  
seismic$month <- gsub("Nov", 11, seismic[,1])  
seismic$month <- gsub("Dec", 12, seismic[,1])  
seismic$month <- as.numeric(seismic$month)  
#sort the data as data   
  
seismic <- seismic[order(seismic$month),]  
seismic

## month day freq  
## 25 10 13 536  
## 26 10 14 99  
## 27 10 30 1930  
## 28 10 31 1089  
## 4 11 02 6711  
## 5 11 03 71  
## 6 11 05 5064  
## 7 11 06 265  
## 8 11 07 928  
## 9 11 08 100  
## 10 11 09 1202  
## 11 11 10 3482  
## 12 11 11 12422  
## 13 11 13 9412  
## 14 11 14 5283  
## 15 11 16 8999  
## 16 11 17 1335  
## 17 11 19 3809  
## 18 11 20 1600  
## 19 11 21 5719  
## 20 11 23 4535  
## 21 11 24 717  
## 22 11 27 3976  
## 23 11 29 2374  
## 24 11 30 3622  
## 1 12 02 773  
## 2 12 05 9460  
## 3 12 06 304

seismic\_timeseries <- ts(seismic$freq, start = c(10,13), end = c(12,06), frequency = 28)  
plot(seismic\_timeseries, xlab = "Date", ylab="number of tweets", main="Tweets number betweet Oct. 13th to Dec. 06th")



#using ARIMA Model  
  
d<- 0:2  
p <- 0:10  
q <- 0:10  
seismic\_models <- expand.grid(d=d,p=p,q=q)  
head(seismic\_models, n=4)

## d p q  
## 1 0 0 0  
## 2 1 0 0  
## 3 2 0 0  
## 4 0 1 0

getTSModelAIC <- function(ts\_data, p,d,q){  
 ts\_model <- arima(ts\_data, order = c(p,d,q))  
 return(ts\_model$aic)  
}  
  
  
getTSModelAICSafe <- function(ts\_data,p,d,q){  
 result <- tryCatch({getTSModelAIC(ts\_data,p,d,q)},error = function(e){Inf})  
}  
  
  
#Pick the best model that has the smallest aic  
seismic\_models$aic <- mapply(function(x,y,z)getTSModelAICSafe(seismic\_timeseries,x,y,z), seismic\_models$p, seismic\_models$d, seismic\_models$q)

## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1

## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1

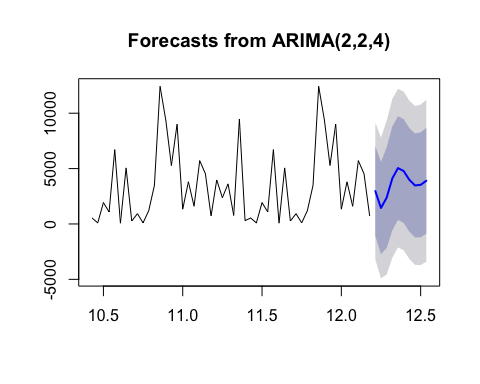
subset(seismic\_models, aic == min(aic))

## d p q aic  
## 141 2 2 4 933.0406

#ARIMA model for best p,d,q, order model  
seismic\_model<- arima(seismic\_timeseries, order = c(2,2,4))  
summary(seismic\_model)

##   
## Call:  
## arima(x = seismic\_timeseries, order = c(2, 2, 4))  
##   
## Coefficients:  
## ar1 ar2 ma1 ma2 ma3 ma4  
## 0.8587 -0.6146 -2.6496 3.1396 -2.3224 0.8361  
## s.e. 0.1792 0.2151 0.2109 0.4960 0.5382 0.2356  
##   
## sigma^2 estimated as 9504751: log likelihood = -459.52, aic = 933.04  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -420.0101 3020.689 2233.792 -494.7558 521.1218 0.6452582  
## ACF1  
## Training set -0.07686123

plot(forecast(seismic\_model,10))



Using the ARIMA model to predict the number of tweets in the future. The number of posted tweets will decline first then increase. However, the peak of prediction is about 5000.