

ステップ5： AI予想と本当の相場を比べる

佐藤 聖

今回の結果を確かめる

● ターゲット…2016年9月の米国経済指標

第5章で作ったプログラムの効果を検証します。ニュースの分析結果と、株式指標や為替価格の動きとを、2016年9月18日～24日の1週間分のデータを用いて比較します。

● AI予想結果を確かめるために単語抽出処理をやり直す

最初にダウンロードしたツイートからサンプリングをやり直します。第2章リスト22を元に、中央付近の「# データ期間を指定」を、2016年9月18日～24日の1週間分を抽出と書き換え、「# サンプリング」部分を削除したのがリスト1です。

サンプリング・データをCSVファイルに書き出した後、第5章リスト1で感情査定とクラスタ分析を行いました。クラスタ分析の結果、出力されたのが図1です。

ツイートは米国で発信された経済ニュースとします。ここでは一般紙であるUSAトゥデイとワシントン・ポストは除外しました。クラスタ数はGAP統計量を元に算出しました。

● 図の考察1…形状から

図1(c)のIBDインベスターズ以外は、3角形のようにデータが分布しています。この形からもレンジ相場(売りと買いが均衡し一定の値幅にて上下を繰り返している)である可能性がうかがえます。

図1から、関心度の高いニュースはほんの一部しかなかったようです。米国の納税シーズンである6月く

リスト1 AI予想結果を確かめるために単語抽出処理のプログラムを作り直した(9月18日から24日まで抽出)

```
# サンプリングのためのライブラリを読み込み
library(dplyr)

# テキスト・クレンジング処理を関数化
# 関数名:cleand.data()
cleand.data = function(sentences){
  cl_text <- iconv(sentences, "latin1", "ASCII", sub="")
  cl_text <- gsub("<.?*>", "", cl_text)
  cl_text <- gsub("&#x27;", "'", cl_text)
  cl_text <- gsub("(f|ht)tp(?:S+|Ys)*", "", cl_text)
  cl_text <- gsub("(RT|via)((?:\b\S+/)+)", "", cl_text)
  cl_text <- gsub("@\b\S+", "", cl_text)
  cl_text <- gsub("[:punct:]", "", cl_text)
  cl_text <- gsub("[:digit:]", "", cl_text)
  cl_text <- gsub("[ \t]{2,}", " ", cl_text)
  cl_text <- gsub("&#x27;", "'", cl_text)
  df_tw$created <- as.POSIXct(df_tw$created, tz = "EST")
  return(cl_text)
}

# ユーザー名を指定
usernames <- c("WSJmarkets", "FXstreetNews", "IBDinvestors",
               "ForexLive", "USATODAY", "washingtonpost")

# 全てのユーザー分に対して処理を繰り返す
for (user in usernames) {
  # 事前にPCへダウンロードしたタイムラインのデータを読み込み
  filename <- paste(user, ".csv", sep = "")
  df_tw <- read.csv(filename, header = TRUE, skip = 0)

  # データ期間を絞り込み、データを抽出
  grep_text1 <- grep("2016-09-18", df_tw$created)
  grep_text2 <- grep("2016-09-19", df_tw$created)
  grep_text3 <- grep("2016-09-20", df_tw$created)
  grep_text4 <- grep("2016-09-21", df_tw$created)
  grep_text5 <- grep("2016-09-22", df_tw$created)
  grep_text6 <- grep("2016-09-23", df_tw$created)
  grep_text7 <- grep("2016-09-24", df_tw$created)
  df_list <- c(grep_text1, grep_text2, grep_text3, grep_text4,
              grep_text5, grep_text6, grep_text7)
  df_tw <- data.frame(df_tw[df_list,])
  print(user)
  print(nrow(df_tw))

  # テキスト・クレンジング処理
  df_tw <- df_tw[complete.cases(df_tw$text),]

  # 関数cleand.data()を呼び出してクレンジング処理
  df_tw$text <- cleand.data(df_tw$text)

  # 日時でソートし、行番号を振り直し
  sortlist <- order(df_tw$created)
  df_tw <- df_tw[sortlist,]
  rownames(df_tw) <- c(1:nrow(df_tw))

  # サンプリング・データをCSVファイルに書き出し
  filename <- paste(user, "_sample.csv", sep = "")
  write.csv(df_tw, filename, quote = FALSE, row.names = FALSE)
}
```