信用卡帳單-簡易分析

Andrew Tang

2017/5/29

前言

懶得一筆一筆的記帳，但實在是不知道錢都花去哪了。若能把信用卡的帳單自動化的整理出來，那該有多好，於是挑了幾張卡的帳單來試試，以下提供玉山及元大的成果及Code。

玉山

- 範例帳單連結

```
ESUN <- function(path){
  options(stringsAsFactors = F)
  require(textreadr)
  require(stringr)
  require(magrittr)
  pdftext <- pdftools::pdf_text(path)
  tmp <- str_extract_all(pdftext[1], ".*")[[1]] %>%
    .[-(1:grep("本期消費明細", .))] %>%
    grep(" [0-9]{2,2}/[0-9]{2,2}", ., value = T) %>%
    gsub(",","",.)

  detail <- strsplit(tmp, "[0-9]{2,2}/[0-9]{2,2}|TWD.*") %>%
              sapply(function(x)paste(x,collapse="")) %>%
              str_trim() %>% gsub(" {1,}","_",.)

  result <- strsplit(gsub(" {2,}"," ",tmp), " ") %>%
    sapply(function(x)x[c(2,3,length(x)-1,length(x))]) %>%
    {data.frame(detail,t(.))} %>% set_colnames(c("消費摘要",
                                                 "消費日",
                                                 "入帳日",
                                                 "繳款幣別",
                                                 "臺幣金額"))

  result$臺幣金額 %<>% as.numeric()
  return(result[c("消費摘要",
                  "消費日",
                  "臺幣金額")])
}
```

簡易分析

```
df <- ESUN("demo/ESUN_Estatement_10603.pdf")
sum(df$臺幣金額)
```

```
## [1] 15006
```

```
library(dplyr)
df$月份 <- substr(df$消費日,1,2)
```

```
group_by(df, 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 2 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    03            9293
## 2    04            5713
```

```
group_by(df[grep("悠遊卡",df$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 2 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    03            1000
## 2    04            1500
```

```
加油 <- paste(c("加油","台亞","台塑石油"), collapse = "|")
group_by(df[grep(加油,df$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 1 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    04             991
```

元大

- 範例帳單連結
- 原始帳單需要密碼，使用參數pwd即可

```
Yuanta <- function(path, pwd=""){
  options(stringsAsFactors = F)
  require(textreadr)
  require(stringr)
  require(magrittr)
  pdftext <- pdftools::pdf_text(path, upw=pwd)
  tmp <- str_extract_all(pdftext[2], ".*")[[1]] %>%
    .[-(1:grep("卡號：", .))] %>%
    grep(" [0-9]{2,2}/[0-9]{2,2}", ., value = T) %>%
    gsub(",","",.)

  detail <- sapply(strsplit(substr(tmp, 14,nchar(tmp)), " {10,}"), "[", 1) %>% gsub(" ","_",.)
  tmp <- str_extract(tmp, ".*[A-Z]{2,2}/[A-Z]{3,3}")
  tmp <- strsplit(gsub(" {2,}"," ",tmp), " ") %>%
sapply(function(x)x[c(2,3,length(x)-1,length(x))])
  result <- data.frame(detail,t(tmp)) %>%
              set_colnames(c("消費摘要",

                             "消費日",
                             "入帳日",
                             "臺幣金額",
                             "國家/幣別"))

  result$臺幣金額 %<>% as.numeric()
```

```
  result$臺幣金額 %>% as.numeric()
  return(result[c("消費摘要",
                  "消費日",
                  "臺幣金額")])
}
```

簡易分析

```
df <- Yuanta("demo/Yuanta-CreditCard_estatement_10602.pdf")
sum(df$臺幣金額)
```

```
## [1] 8416
```

```
library(dplyr)
df$月份 <- substr(df$消費日,1,2)
group_by(df, 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 2 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    02            7610
## 2    03             806
```

```
group_by(df[grep("悠遊卡",df$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 0 × 2
## # ... with 2 variables: 月份 <chr>, sum(臺幣金額) <dbl>
```

```
加油 <- paste(c("加油","台亞","台塑石油"), collapse = "|")
group_by(df[grep(加油,df$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 1 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    02            3251
```

合併分析

```
est <- list()
est[[1]] <- Yuanta("demo/Yuanta-CreditCard_estatement_10602.pdf")
est[[2]] <- ESUN("demo/ESUN_Estatement_10603.pdf")
result <- do.call("rbind", est)
sum(result$臺幣金額)
```

```
## [1] 23422
```

```
library(dplyr)
result$月份 <- substr(result$消費日,1,2)
group_by(result, 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 3 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    02            7610
## 2    03           10099
## 3    04            5713
```

```
group_by(result[grep("悠遊卡",result$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 2 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    03            1000
## 2    04            1500
```

```
加油 <- paste(c("加油","台亞","台塑石油"), collapse = "|")
group_by(result[grep(加油,result$消費摘要),], 月份) %>% summarise(sum(臺幣金額))
```

```
## # A tibble: 2 × 2
##    月份 `sum(臺幣金額)`
##   <chr>           <dbl>
## 1    02            3251
## 2    04             991
```

結論

- 不務正業的弄這些東西滿花時間，真希望銀行可以開放API供查詢
- 凱基銀行的帳單要用OCR的方式解，我投降。