

# Evaluación MLII: Ejercicio 3

Selección de atributos

*Inmaculada Perea Fernández*

*junio 2017*

Realice las siguientes acciones sobre la base de datos *spam* de la librería *kernlab*:

Carga de librerías

```
if (!require('FSelector')) install.packages('FSelector'); library('FSelector')
if (!require('kernlab')) install.packages('kernlab'); library('kernlab')
if (!require('Biocomb')) install.packages('Biocomb'); library('Biocomb')
if (!require('plyr')) install.packages('plyr'); library('plyr')
```

Establecimiento de la semilla

```
set.seed(123456789)
```

Carga, inspección y preparación de los datos

El conjunto de datos *spam* consta de 4601 observaciones y 58 variables:

```
data(spam)
dim(spam)
```

```
## [1] 4601 58
```

```
summary(spam)
```

```
##      make      address      all      num3d
## Min.   :0.0000   Min.    : 0.000   Min.   :0.0000   Min.    : 0.00000
## 1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.: 0.00000
## Median :0.0000   Median : 0.000   Median :0.0000   Median : 0.00000
## Mean   :0.1046   Mean    : 0.213   Mean    :0.2807   Mean    : 0.06542
## 3rd Qu.:0.0000   3rd Qu.: 0.000   3rd Qu.:0.4200   3rd Qu.: 0.00000
## Max.   :4.5400   Max.    :14.280   Max.    :5.1000   Max.    :42.81000
##      our      over      remove      internet
## Min.   : 0.0000   Min.   :0.0000   Min.   :0.0000   Min.   : 0.0000
## 1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.0000
## Median : 0.0000   Median :0.0000   Median :0.0000   Median : 0.0000
## Mean   : 0.3122   Mean    :0.0959   Mean    :0.1142   Mean    : 0.1053
## 3rd Qu.: 0.3800   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.: 0.0000
## Max.   :10.0000   Max.    :5.8800   Max.    :7.2700   Max.    :11.1100
##      order      mail      receive      will
## Min.   :0.00000   Min.   : 0.0000   Min.   :0.00000   Min.   :0.0000
## 1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.:0.0000
## Median :0.00000   Median : 0.0000   Median :0.00000   Median :0.1000
## Mean   :0.09007   Mean    : 0.2394   Mean    :0.05982   Mean    :0.5417
## 3rd Qu.:0.00000   3rd Qu.: 0.1600   3rd Qu.:0.00000   3rd Qu.:0.8000
## Max.   :5.26000   Max.    :18.1800   Max.    :2.61000   Max.    :9.6700
##      people      report      addresses      free
## Min.   :0.00000   Min.   : 0.00000   Min.   :0.0000   Min.   : 0.0000
## 1st Qu.:0.00000   1st Qu.: 0.00000   1st Qu.:0.0000   1st Qu.: 0.0000
## Median :0.00000   Median : 0.00000   Median :0.0000   Median : 0.0000
```

##	Mean	:0.09393	Mean	: 0.05863	Mean	:0.0492	Mean	: 0.2488
##	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	0.0000	3rd Qu.:	0.1000
##	Max.	:5.55000	Max.	:10.00000	Max.	:4.4100	Max.	:20.0000
##	business		email		you		credit	
##	Min.	:0.0000	Min.	:0.0000	Min.	: 0.000	Min.	: 0.00000
##	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.000	1st Qu.:	0.00000
##	Median	:0.0000	Median	:0.0000	Median	: 1.310	Median	: 0.00000
##	Mean	:0.1426	Mean	:0.1847	Mean	: 1.662	Mean	: 0.08558
##	3rd Qu.:	0.0000	3rd Qu.:	0.0000	3rd Qu.:	2.640	3rd Qu.:	0.00000
##	Max.	:7.1400	Max.	:9.0900	Max.	:18.750	Max.	:18.18000
##	your		font		num000		money	
##	Min.	: 0.0000	Min.	: 0.0000	Min.	:0.0000	Min.	: 0.00000
##	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.00000
##	Median	: 0.2200	Median	: 0.0000	Median	:0.0000	Median	: 0.00000
##	Mean	: 0.8098	Mean	: 0.1212	Mean	:0.1016	Mean	: 0.09427
##	3rd Qu.:	1.2700	3rd Qu.:	0.0000	3rd Qu.:	0.0000	3rd Qu.:	0.00000
##	Max.	:11.1100	Max.	:17.1000	Max.	:5.4500	Max.	:12.50000
##	hp		hpl		george		num650	
##	Min.	: 0.0000	Min.	: 0.0000	Min.	: 0.0000	Min.	:0.0000
##	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.0000
##	Median	: 0.0000	Median	: 0.0000	Median	: 0.0000	Median	:0.0000
##	Mean	: 0.5495	Mean	: 0.2654	Mean	: 0.7673	Mean	:0.1248
##	3rd Qu.:	0.0000	3rd Qu.:	0.0000	3rd Qu.:	0.0000	3rd Qu.:	0.0000
##	Max.	:20.8300	Max.	:16.6600	Max.	:33.3300	Max.	:9.0900
##	lab		labs		telnet		num857	
##	Min.	: 0.00000	Min.	:0.0000	Min.	: 0.00000	Min.	:0.00000
##	1st Qu.:	0.00000	1st Qu.:	0.0000	1st Qu.:	0.00000	1st Qu.:	0.00000
##	Median	: 0.00000	Median	:0.0000	Median	: 0.00000	Median	:0.00000
##	Mean	: 0.09892	Mean	:0.1029	Mean	: 0.06475	Mean	:0.04705
##	3rd Qu.:	0.00000	3rd Qu.:	0.0000	3rd Qu.:	0.00000	3rd Qu.:	0.00000
##	Max.	:14.28000	Max.	:5.8800	Max.	:12.50000	Max.	:4.76000
##	data		num415		num85		technology	
##	Min.	: 0.00000	Min.	:0.00000	Min.	: 0.0000	Min.	:0.00000
##	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.0000	1st Qu.:	0.00000
##	Median	: 0.00000	Median	:0.00000	Median	: 0.0000	Median	:0.00000
##	Mean	: 0.09723	Mean	:0.04784	Mean	: 0.1054	Mean	:0.09748
##	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	0.0000	3rd Qu.:	0.00000
##	Max.	:18.18000	Max.	:4.76000	Max.	:20.0000	Max.	:7.69000
##	num1999		parts		pm		direct	
##	Min.	:0.000	Min.	:0.0000	Min.	: 0.00000	Min.	:0.00000
##	1st Qu.:	0.000	1st Qu.:	0.0000	1st Qu.:	0.00000	1st Qu.:	0.00000
##	Median	:0.000	Median	:0.0000	Median	: 0.00000	Median	:0.00000
##	Mean	:0.137	Mean	:0.0132	Mean	: 0.07863	Mean	:0.06483
##	3rd Qu.:	0.000	3rd Qu.:	0.0000	3rd Qu.:	0.00000	3rd Qu.:	0.00000
##	Max.	:6.890	Max.	:8.3300	Max.	:11.11000	Max.	:4.76000
##	cs		meeting		original		project	
##	Min.	:0.00000	Min.	: 0.0000	Min.	:0.0000	Min.	: 0.0000
##	1st Qu.:	0.00000	1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.0000
##	Median	:0.00000	Median	: 0.0000	Median	:0.0000	Median	: 0.0000
##	Mean	:0.04367	Mean	: 0.1323	Mean	:0.0461	Mean	: 0.0792
##	3rd Qu.:	0.00000	3rd Qu.:	0.0000	3rd Qu.:	0.0000	3rd Qu.:	0.0000
##	Max.	:7.14000	Max.	:14.2800	Max.	:3.5700	Max.	:20.0000
##	re		edu		table		conference	
##	Min.	: 0.0000	Min.	: 0.0000	Min.	:0.000000	Min.	: 0.00000

```
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.:0.000000 1st Qu.: 0.00000
## Median : 0.0000 Median : 0.0000 Median :0.000000 Median : 0.00000
## Mean : 0.3012 Mean : 0.1798 Mean :0.005444 Mean : 0.03187
## 3rd Qu.: 0.1100 3rd Qu.: 0.0000 3rd Qu.:0.000000 3rd Qu.: 0.00000
## Max. :21.4200 Max. :22.0500 Max. :2.170000 Max. :10.00000
## charSemicolon charRoundbracket charSquarebracket charExclamation
## Min. :0.00000 Min. :0.000 Min. :0.00000 Min. : 0.0000
## 1st Qu.:0.00000 1st Qu.:0.000 1st Qu.:0.00000 1st Qu.: 0.0000
## Median :0.00000 Median :0.065 Median :0.00000 Median : 0.0000
## Mean :0.03857 Mean :0.139 Mean :0.01698 Mean : 0.2691
## 3rd Qu.:0.00000 3rd Qu.:0.188 3rd Qu.:0.00000 3rd Qu.: 0.3150
## Max. :4.38500 Max. :9.752 Max. :4.08100 Max. :32.4780
## charDollar charHash capitalAve capitalLong
## Min. :0.00000 Min. : 0.00000 Min. : 1.000 Min. : 1.00
## 1st Qu.:0.00000 1st Qu.: 0.00000 1st Qu.: 1.588 1st Qu.: 6.00
## Median :0.00000 Median : 0.00000 Median : 2.276 Median : 15.00
## Mean :0.07581 Mean : 0.04424 Mean : 5.191 Mean : 52.17
## 3rd Qu.:0.05200 3rd Qu.: 0.00000 3rd Qu.: 3.706 3rd Qu.: 43.00
## Max. :6.00300 Max. :19.82900 Max. :1102.500 Max. :9989.00
## capitalTotal type
## Min. : 1.0 nonspam:2788
## 1st Qu.: 35.0 spam :1813
## Median : 95.0
## Mean : 283.3
## 3rd Qu.: 266.0
## Max. :15841.0
```

```
str(spam)
```

```
## 'data.frame': 4601 obs. of 58 variables:
## $ make : num 0 0.21 0.06 0 0 0 0 0 0.15 0.06 ...
## $ address : num 0.64 0.28 0 0 0 0 0 0 0 0.12 ...
## $ all : num 0.64 0.5 0.71 0 0 0 0 0 0.46 0.77 ...
## $ num3d : num 0 0 0 0 0 0 0 0 0 0 ...
## $ our : num 0.32 0.14 1.23 0.63 0.63 1.85 1.92 1.88 0.61 0.19 ...
## $ over : num 0 0.28 0.19 0 0 0 0 0 0 0.32 ...
## $ remove : num 0 0.21 0.19 0.31 0.31 0 0 0 0.3 0.38 ...
## $ internet : num 0 0.07 0.12 0.63 0.63 1.85 0 1.88 0 0 ...
## $ order : num 0 0 0.64 0.31 0.31 0 0 0 0.92 0.06 ...
## $ mail : num 0 0.94 0.25 0.63 0.63 0 0.64 0 0.76 0 ...
## $ receive : num 0 0.21 0.38 0.31 0.31 0 0.96 0 0.76 0 ...
## $ will : num 0.64 0.79 0.45 0.31 0.31 0 1.28 0 0.92 0.64 ...
## $ people : num 0 0.65 0.12 0.31 0.31 0 0 0 0 0.25 ...
## $ report : num 0 0.21 0 0 0 0 0 0 0 0 ...
## $ addresses : num 0 0.14 1.75 0 0 0 0 0 0 0.12 ...
## $ free : num 0.32 0.14 0.06 0.31 0.31 0 0.96 0 0 0 ...
## $ business : num 0 0.07 0.06 0 0 0 0 0 0 0 ...
## $ email : num 1.29 0.28 1.03 0 0 0 0.32 0 0.15 0.12 ...
## $ you : num 1.93 3.47 1.36 3.18 3.18 0 3.85 0 1.23 1.67 ...
## $ credit : num 0 0 0.32 0 0 0 0 0 3.53 0.06 ...
## $ your : num 0.96 1.59 0.51 0.31 0.31 0 0.64 0 2 0.71 ...
## $ font : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num000 : num 0 0.43 1.16 0 0 0 0 0 0 0.19 ...
## $ money : num 0 0.43 0.06 0 0 0 0 0 0.15 0 ...
## $ hp : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ hpl : num 0 0 0 0 0 0 0 0 0 0 ...
## $ george : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num650 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ lab : num 0 0 0 0 0 0 0 0 0 0 ...
## $ labs : num 0 0 0 0 0 0 0 0 0 0 ...
## $ telnet : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num857 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ data : num 0 0 0 0 0 0 0 0 0.15 0 ...
## $ num415 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num85 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ technology : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num1999 : num 0 0.07 0 0 0 0 0 0 0 0 ...
## $ parts : num 0 0 0 0 0 0 0 0 0 0 ...
## $ pm : num 0 0 0 0 0 0 0 0 0 0 ...
## $ direct : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ cs : num 0 0 0 0 0 0 0 0 0 0 ...
## $ meeting : num 0 0 0 0 0 0 0 0 0 0 ...
## $ original : num 0 0 0.12 0 0 0 0 0 0.3 0 ...
## $ project : num 0 0 0 0 0 0 0 0 0 0.06 ...
## $ re : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ edu : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ table : num 0 0 0 0 0 0 0 0 0 0 ...
## $ conference : num 0 0 0 0 0 0 0 0 0 0 ...
## $ charSemicolon : num 0 0 0.01 0 0 0 0 0 0 0.04 ...
## $ charRoundbracket : num 0 0.132 0.143 0.137 0.135 0.223 0.054 0.206 0.271 0.03 ...
## $ charSquarebracket : num 0 0 0 0 0 0 0 0 0 0 ...
## $ charExclamation : num 0.778 0.372 0.276 0.137 0.135 0 0.164 0 0.181 0.244 ...
## $ charDollar : num 0 0.18 0.184 0 0 0 0.054 0 0.203 0.081 ...
## $ charHash : num 0 0.048 0.01 0 0 0 0 0 0.022 0 ...
## $ capitalAve : num 3.76 5.11 9.82 3.54 3.54 ...
## $ capitalLong : num 61 101 485 40 40 15 4 11 445 43 ...
## $ capitalTotal : num 278 1028 2259 191 191 ...
## $ type : Factor w/ 2 levels "nonspam","spam": 2 2 2 2 2 2 2 2 2 2 ...
```

```
head(spam)
```

```
## make address all num3d our over remove internet order mail receive
## 1 0.00 0.64 0.64 0 0.32 0.00 0.00 0.00 0.00 0.00 0.00
## 2 0.21 0.28 0.50 0 0.14 0.28 0.21 0.07 0.00 0.94 0.21
## 3 0.06 0.00 0.71 0 1.23 0.19 0.19 0.12 0.64 0.25 0.38
## 4 0.00 0.00 0.00 0 0.63 0.00 0.31 0.63 0.31 0.63 0.31
## 5 0.00 0.00 0.00 0 0.63 0.00 0.31 0.63 0.31 0.63 0.31
## 6 0.00 0.00 0.00 0 1.85 0.00 0.00 1.85 0.00 0.00 0.00
## will people report addresses free business email you credit your font
## 1 0.64 0.00 0.00 0.00 0.32 0.00 1.29 1.93 0.00 0.96 0
## 2 0.79 0.65 0.21 0.14 0.14 0.07 0.28 3.47 0.00 1.59 0
## 3 0.45 0.12 0.00 1.75 0.06 0.06 1.03 1.36 0.32 0.51 0
## 4 0.31 0.31 0.00 0.00 0.31 0.00 0.00 3.18 0.00 0.31 0
## 5 0.31 0.31 0.00 0.00 0.31 0.00 0.00 3.18 0.00 0.31 0
## 6 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0
## num000 money hp hpl george num650 lab labs telnet num857 data num415
## 1 0.00 0.00 0 0 0 0 0 0 0 0 0 0
## 2 0.43 0.43 0 0 0 0 0 0 0 0 0 0
## 3 1.16 0.06 0 0 0 0 0 0 0 0 0 0
## 4 0.00 0.00 0 0 0 0 0 0 0 0 0 0
```

```

## 5  0.00  0.00  0  0      0      0  0  0      0      0  0  0
## 6  0.00  0.00  0  0      0      0  0  0      0      0  0  0
##   num85 technology num1999 parts pm direct cs meeting original project
## 1    0          0    0.00    0 0  0.00 0      0    0.00    0
## 2    0          0    0.07    0 0  0.00 0      0    0.00    0
## 3    0          0    0.00    0 0  0.06 0      0    0.12    0
## 4    0          0    0.00    0 0  0.00 0      0    0.00    0
## 5    0          0    0.00    0 0  0.00 0      0    0.00    0
## 6    0          0    0.00    0 0  0.00 0      0    0.00    0
##   re  edu table conference charSemicolon charRoundbracket
## 1 0.00 0.00    0          0          0.00          0.000
## 2 0.00 0.00    0          0          0.00          0.132
## 3 0.06 0.06    0          0          0.01          0.143
## 4 0.00 0.00    0          0          0.00          0.137
## 5 0.00 0.00    0          0          0.00          0.135
## 6 0.00 0.00    0          0          0.00          0.223
##   charSquarebracket charExclamation charDollar charHash capitalAve
## 1          0          0.778      0.000  0.000      3.756
## 2          0          0.372      0.180  0.048      5.114
## 3          0          0.276      0.184  0.010      9.821
## 4          0          0.137      0.000  0.000      3.537
## 5          0          0.135      0.000  0.000      3.537
## 6          0          0.000      0.000  0.000      3.000
##   capitalLong capitalTotal type
## 1         61         278 spam
## 2        101        1028 spam
## 3        485        2259 spam
## 4         40         191 spam
## 5         40         191 spam
## 6         15          54 spam

```

Tras la observación de los datos podemos concluir lo siguiente:

- La variable objetivo es *type* y consta de 2 clases (*nonspam* y *spam*).
- Todas las variables regresoras son numéricas.
- Todos los datos están completos, no hay valores perdidos.
- Los datos están ligeramente desbalanceados, porque existen más observaciones de la clase *nonspam*.

## Apartado 1

Utilice los siguientes filtros univariantes de la librería *Fselector* para determinar los 10 atributos más relevantes en cada caso:

- `chi-squared`
- `gain.ratio`
- `oneR`
- `random.forest.importance`
- `relief`

### 1.1 chi-squared

```
pesos.chi = chi.squared(type~., spam)
(sel.chi=cutoff.k(pesos.chi, 10))
```

```
## [1] "charExclamation" "charDollar"      "capitalLong"
## [4] "your"              "remove"          "capitalAve"
## [7] "free"              "money"           "capitalTotal"
## [10] "num000"
```

## 1.2 gain.ratio

```
pesos.gain = gain.ratio(type~., spam)
(sel.gain=cutoff.k(pesos.gain, 10))
```

```
## [1] "remove"      "money"      "num000"
## [4] "free"        "charDollar" "credit"
## [7] "hpl"         "george"     "charExclamation"
## [10] "receive"
```

## 1.3 oneR

```
pesos.one = oneR(type~., spam)
(sel.one=cutoff.k(pesos.one, 10))
```

```
## [1] "remove" "money" "parts" "table" "cs" "meeting" "direct"
## [8] "num857" "lab" "telnet"
```

## 1.4 random.forest.importance

```
pesos.RF = random.forest.importance(type~.,
                                     spam,
                                     importance.type=1) # decrecimiento medio en accuracy
(sel.RF=cutoff.k(pesos.RF, 10))
```

```
## [1] "charExclamation" "remove"      "capitalAve"
## [4] "hp"              "charDollar"  "edu"
## [7] "capitalLong"     "free"        "capitalTotal"
## [10] "your"
```

## 1.5 relief

```
pesos.relief = relief(type~.,
                      spam,
                      neighbours.count = 6, # vecinos a analizar
                      sample.size = 15)    # instancias a analizar
(sel.relief=cutoff.k(pesos.relief, 10))
```

```
## [1] "num857" "all" "meeting" "lab" "num415"
## [6] "num000" "your" "num1999" "you" "charDollar"
```

Los filtros *chi cuadrado* y los *basados en entropía* determinan la importancia de atributos **discretos**, por tanto no es recomendable usarlo sobre el conjunto *spam*, ya que los atributos son en su mayoría continuos.

## Apartado 2

Aplique el filtro multivariante *CFS*

```
(sel.cfs = cfs(type~., spam))
```

```
## [1] "remove"          "free"             "your"
## [4] "num000"          "money"            "hp"
## [7] "george"          "charExclamation" "charDollar"
## [10] "capitalAve"
```

## Apartado 3

Aplique el filtro multivariante *FCBF*, utilizando un valor umbral 0.02 para la correlación del atributo con la variable de clase.

Para utilizar la función *select.fast.filter* se debe cumplir lo siguiente:

- La ultima columna del dataframe de entrada debe ser la variable de clase (factor)
- El número máximo de clases 10

El conjunto de datos *spam* tiene la variable *type* en la ultima posición y es de tipo factor con 2 clases (spam/nospam), por tanto no es necesario realizar ningún tipo de transformación.

```
(pesos.fcbf=select.fast.filter(spam,
                                disc.method="MDL",    # Algoritmo discretización
                                threshold=0.02,       # Valor umbral
                                attrs.nominal=58))     # El único atributo nominal es type
```

```
##      Biomarker Information.Gain NumberFeature
## 1 charExclamation      0.19444917           52
## 2 charDollar          0.18110529           53
## 3 capitalLong         0.15942948           56
## 4 remove              0.14783596            7
## 5 your                 0.14345068           21
## 6 free                 0.12983756           16
## 7 hp                   0.10703484           25
## 8 george               0.08856656           27
## 9 edu                  0.03566742           46
## 10 meeting            0.02775998           42
```

```
sel.fcbf=pesos.fcbf$Biomarker
```

## Apartado 4

Indique cuáles son los 5 atributos más frecuentemente seleccionados por los procedimientos aplicados.

Obtenemos la frecuencia con la que ha sido seleccionado cada atributo por los procedimientos aplicados

```
selected_features= c(sel.chi, sel.gain, sel.one, sel.RF, sel.relief, sel.cfs, as.character(sel.fcbf))
df = count(selected_features)
(best.sel = df[order(df[["freq"]], decreasing = TRUE),] )
```

```
##           x freq
## 5      charDollar    6
## 24      remove      6
## 6 charExclamation    5
## 11         free      5
## 28         your      5
## 17         money      4
## 18         num000      4
## 2      capitalAve      3
## 3      capitalLong      3
## 12         george      3
## 13          hp      3
## 16        meeting      3
## 4      capitalTotal      2
## 10          edu      2
## 15          lab      2
## 21        num857      2
## 1          all      1
## 7         credit      1
## 8           cs      1
## 9        direct      1
## 14          hpl      1
## 19        num1999      1
## 20        num415      1
## 22         parts      1
## 23        receive      1
## 25         table      1
## 26        telnet      1
## 27          you      1
```

Seleccionamos los 5 atributos más frecuentes

```
head(best.sel, 5)
```

```
##           x freq
## 5      charDollar    6
## 24      remove      6
## 6 charExclamation    5
## 11         free      5
## 28         your      5
```