Natural language processing Quiz 2

Shane Kao

Saturday, February 21, 2015

```r
setwd("C:/Users/asus/Downloads/final/en_US")
twitter=readLines("en_US.twitter.txt")
blogs=readLines("en_US.blogs.txt")
news=readLines("en_US.news.txt")
data=c(twitter,blogs,news)
```

Question 1

When you breathe, I want to be the air for you. I'll be there for you, I'd live and I'd

```r
x="and I'd"
data_v1=data[grep(x,data)]
n=length(data_v1)
pred=function(y){
        data_v2=data_v1[grep(paste(x,y,sep=" "),data_v1)]
        length(data_v2)/n
}
candidate=c("sleep","die","eat","give")
prob=apply(cbind(candidate),1,pred)
output=data.frame("candidate"=candidate,"probability"=prob)
output[order(output$probability,decreasing=TRUE),]
```

```
##   candidate probability
## 4      give 0.009861933
## 3       eat 0.001972387
## 1     sleep 0.000000000
## 2       die 0.000000000
```

Question 2

Guy at my table's wife got up to go to the bathroom and I asked about dessert and he started telling me about his

```
##        candidate  probability
## 4      spiritual 0.0005319149
## 1      financial 0.0002659574
## 2 horticultural 0.0000000000
## 3        marital 0.0000000000
```

Question 3

I'd give anything to see arctic monkeys this

```
##   candidate  probability
## 1   morning 0.0297452409
## 4   weekend 0.0265439001
## 2     month 0.0125833049
## 3     decade 0.0003278004
```

Question 4

Talking to your mom has the same effect as a hug and helps reduce your

```
##     candidate probability
## 3      stress  0.05504587
## 1      hunger  0.00000000
## 2   happiness  0.00000000
## 4   sleepiness  0.00000000
```

Question 5

When you were in Holland you were like 1 inch away from me but you hadn't time to take a

```
##   candidate probability
## 4      look 0.037550015
## 2   picture 0.031702062
## 3      walk 0.006925208
## 1    minute 0.002770083
```

Question 6

I'd just like all of these questions answered, a presentation of evidence, and a jury to settle the

```
##   candidate probability
## 3      case 0.173913043
## 2    matter 0.043478261
## 4   account 0.008695652
## 1  incident 0.000000000
```

Question 7

I can't deal with unsymetrical things. I can't even hold an uneven number of bags of groceries in each

```
##   candidate probability
## 1      hand 0.009329187
## 2       arm 0.001332741
## 3    finger 0.000000000
## 4       toe 0.000000000
```

Question 8

Every inch of you is perfect from the bottom to the

```
##   candidate probability
## 4       top 0.006415214
## 2      side 0.003101587
## 3    center 0.001525848
## 1    middle 0.001434380
```

Question 9

I'm thankful my childhood was filled with imagination and bruises from playing

```
##   candidate probability
## 4   outside  0.01595745
## 1    weekly  0.00000000
## 2     daily  0.00000000
## 3    inside  0.00000000
```

Question 10

I like how the same people are in almost all of Adam Sandler's

```
##   candidate  probability
## 1   stories 1.673149e-04
## 2    novels 7.648683e-05
## 3  pictures 6.533250e-05
## 4    movies 6.373903e-05
```