David & Goliath: Sentiment Analysis

Ricky Lim

November 13, 2013

```
Filename: sentimentPlot.Rnw
Working directory: /Users/RickyLim/Dropbox/RandomWalk/DavidGoliath/Ranalysis
```

1 Background

The background of this analysis was described here

2 Content

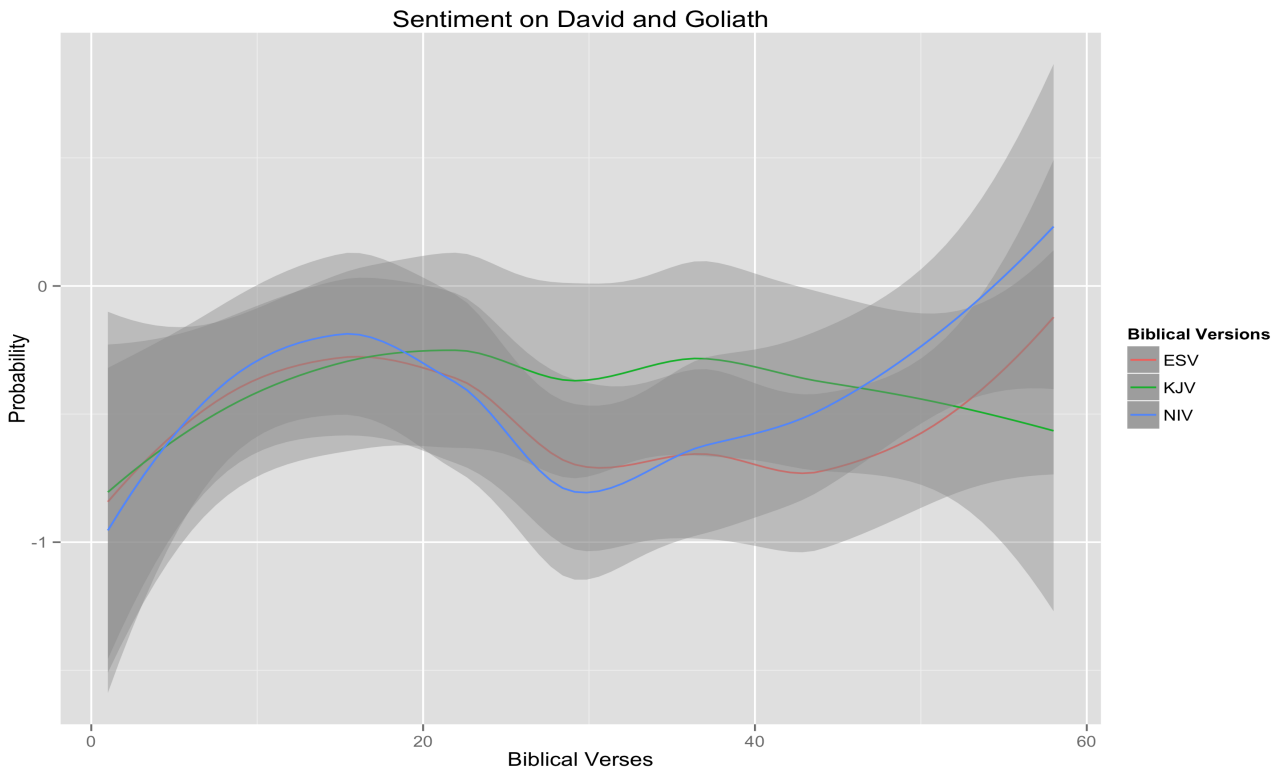
2.1 Data Loaded

```
# from different bible versions
ESV_sentiment <- read.delim("../data/ESV_sentimentAnalysis.txt", header = FALSE)
KJV_sentiment <- read.delim("../data/KJV_sentimentAnalysis.txt", header = FALSE)
NIV_sentiment <- read.delim("../data/NIV_sentimentAnalysis.txt", header = FALSE)
```

2.2 Plot Constructed

```
# construct plot for sentiment probabilities from these three versions
library(ggplot2)
p <- ggplot(ESV_sentiment, aes(x=V1, y=V2))+
     stat_smooth(aes(color="ESV"))+
     stat_smooth(data=KJV_sentiment, aes(color="KJV"))+
     stat_smooth(data=NIV_sentiment, aes(color="NIV"))+
     labs(color="Biblical Versions") +
     xlab('Biblical Verses') +
     ylab('Probability') +
     opts(title='Sentiment on David and Goliath')
ggsave(filename='../img/sentimentPlot.png', plot = p)
```

```
# insert saved figure from above
library(knitr)
f <- "../img/sentimentPlot"
```



2.3 Correlation Checked

```
# merge three biblical versions based on their verses
ESV_KJV <- merge(ESV_sentiment, KJV_sentiment, by.x = "V1", by.y = "V1")
ESV_KJV_NIV <- merge(ESV_KJV, NIV_sentiment, by.x = "V1", by.y = "V1")

# label this dataframe
colnames(ESV_KJV_NIV) <- c("Verse", "ESV", "KJV", "NIV")
rownames(ESV_KJV_NIV) <- ESV_KJV_NIV[, 1]
ESV_KJV_NIV <- ESV_KJV_NIV[, 2:4]

# compute the correlation using spearman
cor_tab <- cor(ESV_KJV_NIV, method = "spearman")

# create table for latex
library(xtable)
print(xtable(cor_tab, caption = "Correlation among Three Versions"))
```

	ESV	KJV	NIV
ESV	1.00	0.58	0.69
KJV	0.58	1.00	0.24
NIV	0.69	0.24	1.00

Table 1: Correlation among Three Versions

2.4 Overall Sentiment

```
print(xtable(summary(ESV_KJV_NIV), caption = "Stat Overview"))
```

	ESV	KJV	NIV
1	Min. :-1.000	Min. :-0.999	Min. :-0.999
2	1st Qu.:-0.923	1st Qu.:-0.922	1st Qu.:-0.911
3	Median :-0.804	Median :-0.761	Median :-0.725
4	Mean :-0.527	Mean :-0.397	Mean :-0.413
5	3rd Qu.:-0.603	3rd Qu.: 0.572	3rd Qu.: 0.244
6	Max. : 0.991	Max. : 0.973	Max. : 0.988

Table 2: Stat Overview

```
MostPositive <- apply(ESV_KJV_NIV, 2, which.max)
MostNegative <- apply(ESV_KJV_NIV, 2, which.min)
MostSentiment <- cbind(MostPositive, MostNegative)
print(xtable(MostSentiment, caption = "Most Sentimented Verses"))
```

	MostPositive	MostNegative
ESV	18	51
KJV	38	8
NIV	18	7

Table 3: Most Sentimented Verses

3 Metainfo

```
sessionInfo()

## R version 3.0.2 (2013-09-25)
## Platform: x86_64-apple-darwin11.4.2 (64-bit)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] reshape2_1.2.2  codetools_0.2-8 knitr_1.5       xtable_1.7-1    ggplot2_0.9.3.1
##
## loaded via a namespace (and not attached):
##  [1] Cairo_1.5-2        colorspace_1.2-4   dichromat_2.0-0    digest_0.6.3
##  [5] evaluate_0.5.1     formatR_0.10       grid_3.0.2         gtable_0.1.2
##  [9] highr_0.3          labeling_0.2       MASS_7.3-29        munsell_0.4.2
## [13] plyr_1.8           proto_0.3-10       RColorBrewer_1.0-5 scales_0.2.3
## [17] stringr_0.6.2      tools_3.0.2
```

```
library(knitr)
knit("sentimentPlot.Rnw")  # compile to tex

## Error:  duplicate label 'setup'

purl("sentimentPlot.Rnw", documentation = 0)  # extract R code only
knit2pdf("sentimentPlot.Rnw")

## Error:  duplicate label 'setup'
```