**Возможности Sentiment-анализа в R**

*Родионова Мария*

*13.06.2019*

Помимо формального анализа текстовых корпусов: подсчетов частот встречаемости для слов, словосочетаний, более сложных методов – например, тематического моделирования, в социальных науках полезным представляется и функционал Sentiment-анализа.

Сентимент-анализ, в его общем понимании, представляет собой формальную количественную эмоциональную оценку тональности текста. Расчет может производиться разными способами, но зачастую используются готовые словари с эмоциональными оценками лемм. Для русского языка используются словари http://www.labinform.ru/pub/rusentilex/index.htm и http://linis-crowd.org/, которые могут быть загружены в качестве таблицы в привычную программу для работы с данными, будь то, например, Python или R.

В R также существует небезынтересный пакет syuzhet, позволяющий осуществлять сентимент-анализ, учитывая темпоральность текста, что хорошо подходит для книг, в которых сюжет в разных главах меняет эмоциональную тональность.

Совмещая функционал пакета «сюжет» и ggplot2, можно строить интересные визуализации, например, так выглядит сентимент-анализ твиттов, упоминающих ПАО «Газпром».

На семинаре мы рассмотрим более подробный пример построения таких моделей.



***Приложение 1.*** Код.

```{r}

gpmentions <- read.csv("Gazprom.csv", sep = ",", header = TRUE, fileEncoding = "UTF-8")

gpmentions$ifduplicate <- duplicated(gpmentions$text)

sum(gpmentions$ifduplicate, na.rm=TRUE)

gpmentions\_clear <- subset(gpmentions, gpmentions$ifduplicate == FALSE)

```

```{r}

tweets\_mentions <-iconv(gpmentions\_clear$text, to = "UTF-8")

emotions <- get\_nrc\_sentiment(tweets\_mentions)

View(emotions)

emotions\_clear <- subset(emotions, !anger == "0" | !anticipation == "0" | !disgust == "0" | !fear == "0" | !joy == "0" | !sadness == "0" | !surprise == "0" | !trust == "0" | !negative == "0" | !positive == "0")

View(emotions\_clear)

```

```{r, fig.align=center, fig.width=9}

sumdata <- data.frame(count=apply(emotions\_clear,2,sum))

sumdata$emotions <- rownames(sumdata)

ggplot(data=sumdata, aes(x=emotions, y=count, fill=emotions)) +

 labs(title = "Gazprom: Sentiment Scores", x="Sentiments", y="Count") +

 geom\_bar(colour="black", stat="identity") +

 theme\_minimal() +

 scale\_fill\_brewer(palette="RdBu") +

 geom\_text(aes(label=count), vjust=-0.3, size=3.5)

```

```{r}

eminem <- get\_text\_as\_string("eminem.txt")

eminem\_v <- get\_sentences(eminem)

eminem\_sentiment <- get\_sentiment(eminem\_v)

```

```{r}

simple\_plot(eminem\_sentiment)

```

```{r}

eminem\_nrc\_full <- get\_nrc\_sentiment(eminem\_v)

eminem\_nrc\_prop <- data.frame(percentage=colSums(prop.table(eminem\_nrc\_full[, 1:8])))

eminem\_nrc\_prop$emotion <- row.names(eminem\_nrc\_prop)

eminem\_nrc\_prop$emotion <- factor(eminem\_nrc\_prop$emotion,

 levels = eminem\_nrc\_prop$emotion[order(eminem\_nrc\_prop$percentage)])

nrc\_plot <- ggplot(data=eminem\_nrc\_prop, aes(x=emotion, y=percentage)) +

 geom\_bar(stat= "identity") +

 xlab("") + ylab("Percentage") +

 coord\_flip() +

 theme\_minimal()

nrc\_plot

```