Markov Text Model

Yile Shi, Letian Yin and Danni Chen

2021/9/30

Change working dictionary and read file to R

```r
setwd("E:/Edi/StP/StP-practical_1-Gourp_8")
Bible <- scan(file = "Bible.txt", what = "character", skip = 156)
n_Bible <-length(Bible)
Bible <- Bible[-((n_Bible - 2909) : n_Bible)] ## strip license
Sys.setlocale('LC_ALL', 'English_United States.1252')
```

```
## [1] "LC_COLLATE=English_United States.1252;LC_CTYPE=English_United States.1252;LC_MONETARY=English_U
```

Create function split_punct() to seperate punctuation marks and words

```r
split_punct <- function(text){
  index_punct <- NULL
  punct <- c(",$", "\\.$", ";$", "!$", ":$", "\\?$")
  #n_index = NULL
  for (i in 1 : length(punct)) {
    index_punct[[i]]<- grep(punct[i], text)
    #n_index[i] = length(index_punct[[i]])
  }
  #cat("The number of words with different punctuations should be:", n_index, collapse = "\n")
  index_punct <- sort(unlist(index_punct, use.names = FALSE))
  index_punct_new <- index_punct + 1 : length(index_punct)
  text_new <- rep("", length(index_punct) + length(text))
  text_new[index_punct_new] <- substr(text[index_punct], nchar(text)[index_punct], nchar(text)[index_pu
  text_new[index_punct_new - 1] <- substr(text[index_punct], 1, nchar(text)[index_punct] - 1)
  text_new[-c(index_punct_new, index_punct_new - 1)] <- text[-index_punct]
  return(text_new)
}
```

Use split_punct() to get new text

```r
Bible_new <- split_punct(Bible)
```