Week 13 pdf

Khue Tran

4/21/2020

Libraries

```
library(tidyverse)
library(twitteR)
library(tm)
library(SnowballC)
library(textstem)
library(wordcloud)
library(ldatuning)
library(topicmodels)
library(tidytext)
library(caret)
library(LiblineaR)
```

Data Import and Cleaning

```
api <- "FGo4ZHuPDMcADDGSRC1ZjXGUk"
apiSecret <- "m4zWMREha1fUePelhsck60CQMuO8qKhQucObPycSXvLNneXW5o"
access <- "1243552285424340994-jeL8kxj8zRP95M9khVPCptH6NaWp7m"
accessSecret <- "mMoP3SJ2dGY67j1yT9CLLeXTQ60X48ThtHjq1nJOPds6O"
setup_twitter_oauth(api, apiSecret,access,accessSecret)
```

```
## [1] "Using direct authentication"
```

```
imported <- searchTwitter("#baking", 5000)
imported_tbl <- twListToDF(imported) %>%
  dplyr::filter(isRetweet == F)
imported_tbl$text <- imported_tbl$text %>%
  iconv("UTF-8", "ASCII", sub="")
```

```
# preprocessed lemmas
twitter_cp <- VCorpus(VectorSource(imported_tbl))

stops <- c(stopwords(kind = 'en'), '#baking', 'baking', 'false', 'true')
removeURL <- function(x) {
  gsub("http.*", "", x)
  gsub("href.*", "", x)}
twitter_cp <- tm_map(twitter_cp, PlainTextDocument)
twitter_cp <- tm_map(twitter_cp, content_transformer(str_to_lower))
twitter_cp <- tm_map(twitter_cp, removeWords, stops)
```

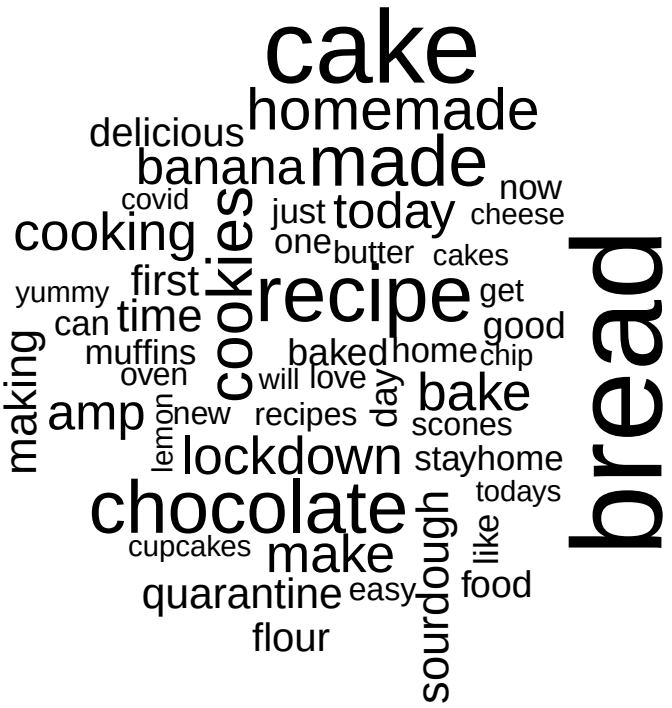
```
twitter_cp <- tm_map(twitter_cp, removeNumbers)
twitter_cp <- tm_map(twitter_cp, removePunctuation)
twitter_cp <- tm_map(twitter_cp, stripWhitespace)
twitter_cp <- tm_map(twitter_cp, content_transformer(removeURL))



# unigram and bigram DTM
# RWeka package does not run
# myTokenizer <- function(x) {NGramTokenizer(x,
#                                            Weka_control(min=1, max=2))}
# twitter_dtm <- DocumentTermMatrix(twitter_cp,
#                                   control = list(
#                                     tokenize = myTokenizer))


twitter_dtm <- DocumentTermMatrix(twitter_cp)

# eliminate sparse terms
twitter_slimmed <- removeSparseTerms(twitter_dtm, .95)

tokenCounts <- apply(twitter_slimmed, 1, sum)
twitter_cleaned_dtm <- twitter_slimmed[tokenCounts > 0,]
twitter_tbl <- as.tibble(as.matrix(twitter_cleaned_dtm))
```

```
## Warning: `as.tibble()` is deprecated, use `as_tibble()` (but mind the new semantics).
## This warning is displayed once per session.
```

```
# delete cases for tweets where no tokens were retained
dropped_tbl <- imported_tbl %>%
  unnest_tokens(token, text, token = "ngrams", n = 1) %>%
  filter(token %in% names(twitter_tbl)) %>%
  arrange(token)
```

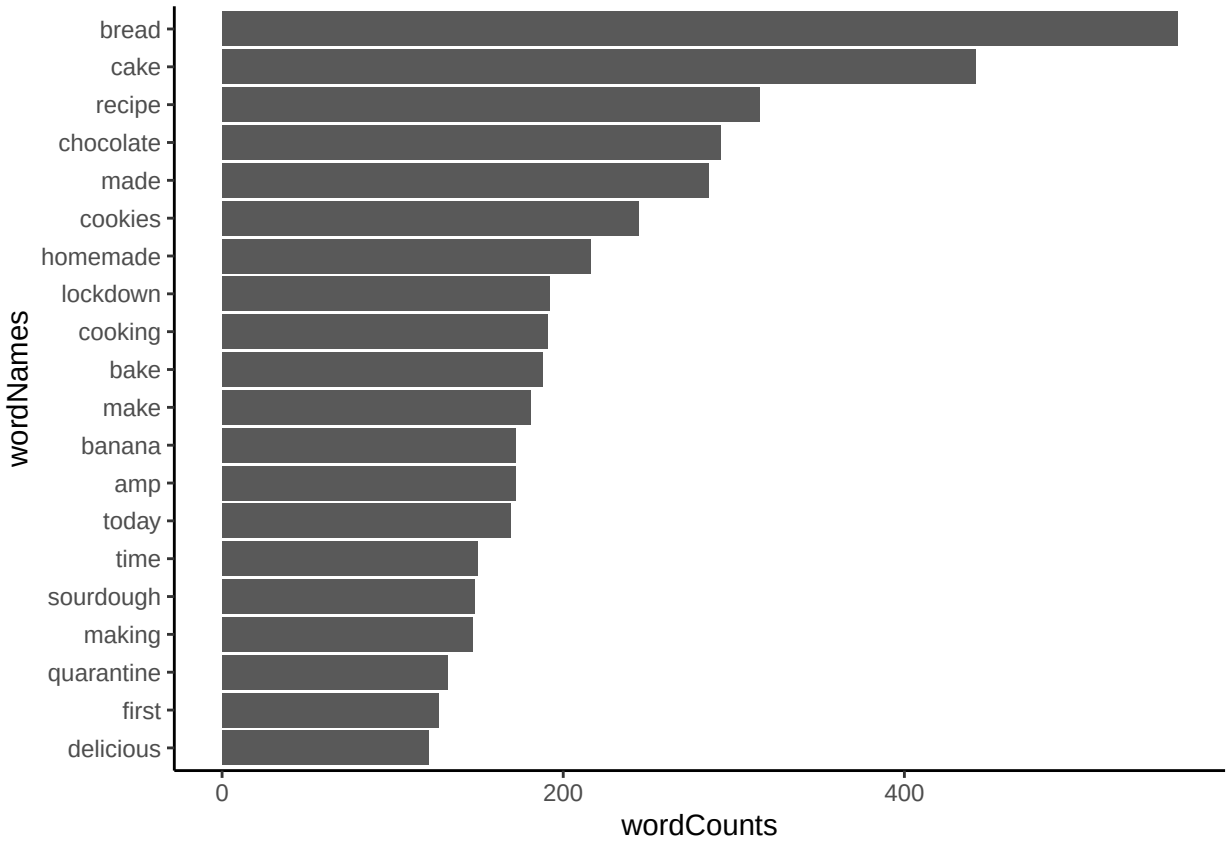
Visualization

```
# wordcloud
wordCounts <- colSums(twitter_tbl)
wordNames <- names(twitter_tbl)
wordcloud(wordNames, wordCounts, max.words = 50)
```




```
# bar chart
tibble(wordNames, wordCounts) %>%
  arrange(desc(wordCounts)) %>%
  top_n(20) %>%
  mutate(wordNames = reorder(wordNames, wordCounts)) %>%
  ggplot(aes(x = wordNames, y = wordCounts)) + geom_col() + coord_flip() + theme_classic()
```

```
## Selecting by wordCounts
```



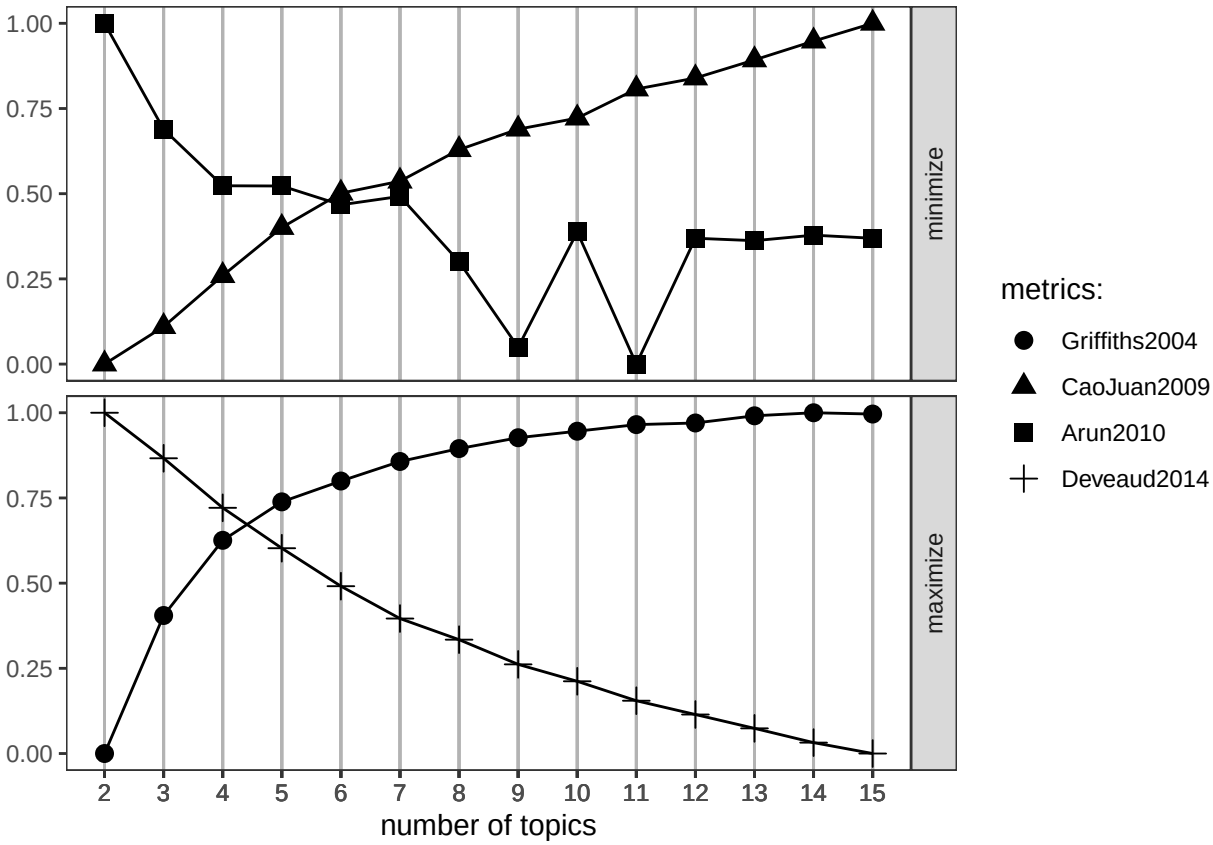
Analysis

Topic Modeling

```
tuning <- FindTopicsNumber(twitter_cleaned_dtm,
                           topics = seq(2,15,1),
                           metrics = c("Griffiths2004", "CaoJuan2009","Arun2010", "Deveaud2014"),
                           verbose = T)
```

```
## fit models... done.
## calculate metrics:
##   Griffiths2004... done.
##   CaoJuan2009... done.
##   Arun2010... done.
##   Deveaud2014... done.
```

```
FindTopicsNumber_plot(tuning)
```




```
lda_results <- LDA(twitter_cleaned_dtm, 4) # num topics based on plot
lda_betas <- tidy(lda_results, matrix="beta")
lda_gammas <- tidy(lda_results, matrix="gamma")

(topic_top <- lda_betas %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  arrange(topic, -beta))
```

```
## # A tibble: 40 x 3
## # Groups:   topic [4]
##    topic term               beta
##    <int> <chr>             <dbl>
##  1     1 realbread       0.00495
##  2     1 buttermilklips  0.00495
##  3     1 kitchenlevelup  0.00495
##  4     1 bakingcast      0.00340
##  5     1 punkrockpastrys 0.00278
##  6     1 jojosekri       0.00247
##  7     1 runcertain      0.00247
##  8     1 multiplexer     0.00217
##  9     1 oliverdevoti    0.00217
## 10     1 kyleecooks      0.00186
## # ... with 30 more rows
```

Topics 1, 2, and 3 are similar, with Topic 2 related more to actions related to baking and Topic 1 and 3 related more to baking recipes and ingredients. Topics 4 seems to include hashtags related to baking.

Machine Learning

```
dropped <- dropped_tbl %>%
  group_by(token) %>%
  summarise(popularity = sum(favoriteCount))

transpose_df <- function(df) {
  t_df <- data.table::transpose(df)
  colnames(t_df) <- rownames(df)
  rownames(t_df) <- colnames(df)
  t_df <- t_df %>%
    tibble::rownames_to_column(.data = .) %>%
    tibble::as_tibble(.)
  return(t_df)
}

twitter_tbl_ml <- transpose_df(twitter_tbl) %>%
  inner_join(dropped, by = c("rowname" = "token")) %>%
  inner_join(lda_betas, by = c("rowname" = "term")) %>%
  rename(token = rowname) %>%
  mutate(topic = factor(topic))
```

```
svm_mod1 <- train(popularity ~ token,
                 data = twitter_tbl_ml,
                 # SVM
                 method = "svmLinear3",
                 # missing values
                 # na.action = na.pass,
                 # Set cross-validation to be 10 fold
                 trControl = trainControl("cv", number = 10))
```

```
## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1
```

```
## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1
```

```
## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1
```

```
## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1
```

```r
svm_mod2 <- train(popularity ~ token + topic,
                 data = twitter_tbl_ml,
```

```r
                # SVM
                method = "svmLinear3",
                # Set cross-validation to be 10 fold
                trControl = trainControl("cv", number = 10))
```

```
## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
```

```
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
```

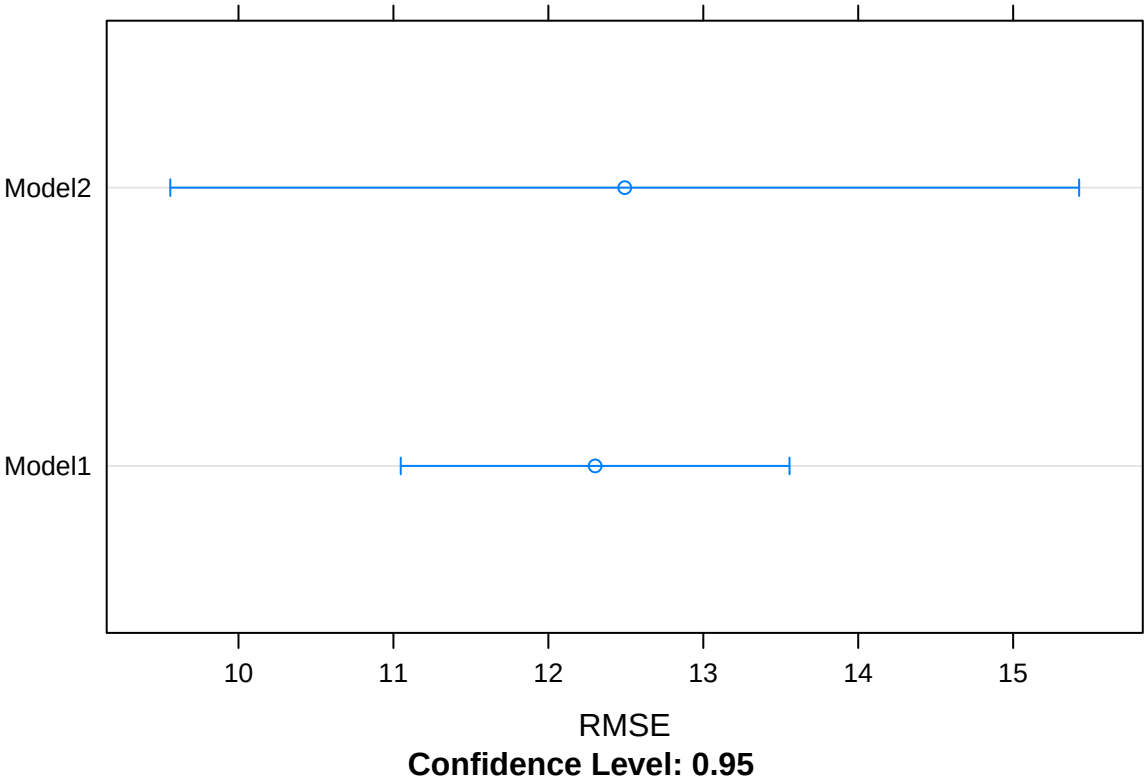
```
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
```

```
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1

## Warning in LiblineaR::LiblineaR(data = as.matrix(x), target = y, cost =
## param$cost, : No value provided for svr_eps. Using default of 0.1
```

```r
# comparing cross-validated
summary(resamples(list(svm_mod1, svm_mod2)))
```

```
##
## Call:
## summary.resamples(object = resamples(list(svm_mod1, svm_mod2)))
##
## Models: Model1, Model2
## Number of resamples: 10
##
## MAE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## Model1 3.286181 3.387771 3.635363 3.606076 3.826685 3.881414    0
## Model2 3.077836 3.464462 3.632748 3.646735 3.820913 4.145073    0
##
## RMSE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## Model1 10.33156 10.84521 12.19497 12.30191 13.50758 15.64975    0
## Model2  7.51368 10.25347 11.24652 12.49291 14.06242 20.94662    0
##
## Rsquared
##             Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## Model1 0.9984401 0.9990145 0.9991879 0.9991482 0.9992813 0.9995731    0
## Model2 0.9969370 0.9981374 0.9990120 0.9986219 0.9990918 0.9995509    0
```

```r
dotplot(resamples(list(svm_mod1, svm_mod2)), metric="RMSE")
```



Final Interpretation

Based on the comparison of the two models and the topics extracted from the analysis, it seems like topics are not very meaningful and too noisy, leading to the model 2 with topic as the added predictors are not