Appendix C: Sentitopics Tutorial

Contents

Introduction

In this tutorial we show how interested reasearchers can easily get started with JST and rJST models using the sentitopics R package. We demonstrate an example workfrom from start (reading in a text corpus) to finish (visualizing JST and rJST results).

The repository can be found at: https://github.com/cpipal/sentitopics-tutorial. There you can always find the up-to-date version.

Preparing the data

We use two data sources:

- `EUSpeech v2` = Dataset of EU leader speeches Schumacher et al. (2020)
- `LSD2015` = Lexicoder sentiment dictionary as provided by quanteda

Load the EUSpeech v2 corpus and select speeches from the United Kingdom:

```
load("data/EUspeech_V2.RData")
corpus <- corpus %>%
  quanteda::corpus_subset(country == "Great Britain")
```

Then turn the texts into a `quanteda` dfm (document-feature matrix). We apply a couple of preprocessing steps such as stopword removal, lowercasing, stemming, and the removal of very rarely and frequently used terms (We remove words that appear less than 10 times and words that appear in more than 95% of the documents).

```
dfm <- corpus %>%
    dfm(verbose = TRUE,
        tolower = TRUE,
        remove = stopwords("english"),
        remove_punct = TRUE,
        remove_numbers = TRUE) %>%
    dfm_wordstem(language = "english")

  dfm <- dfm %>%
    dfm_trim(min_termfreq = 10, termfreq_type = "count") %>%
    dfm_trim(max_docfreq = 0.95, docfreq_type = "prop")
```

We also create a `dtm` object so we can show that the `sentitopics` package can also work with a document term matrix as input (this comes in handy if you prefer to use the `tidytext` or `tm` text analysis packages).

```
dtm <- dfm %>% quanteda::convert(to = "tm")
```

Now the only thing left is that we have to load a dictionary that we want to use as the supervised input for the JST/rJST models. Here we are going to use the Lexcoder dictionary that comes with the `quanteda` package.

```
dict <- quanteda::data_dictionary_LSD2015[1:2]
```

Estimating speech-level sentiment with JST

We can estimate the speech-level sentiment using the `jst()` function of the `sentitopics` package. Similarl to LDA, we have to choose the number of topics and iterations. We can also experiment with hyperparameter settings, but going with the default values is usually fine.

```
set.seed(1899)
jst_out <- sentitopics::jst(dfm, dict, numTopics = 30, numIters = 100)
```

That's it! We can now easily inspect the different model results using the `get_parameter()` function. Let's try this to get the speech-level sentiment estimates for each speech in our dataset:

```
pi <- sentitopics::get_parameter(jst_out, "pi")
pi %>%
  select(sent1, sent2, sent3) %>%
  head()
```

```
##              sent1     sent2     sent3
## text5343 0.2737402 0.2962636 0.4299962
## text5344 0.5426177 0.1598298 0.2975525
## text5345 0.4159158 0.2715969 0.3124873
## text5346 0.2593740 0.1888950 0.5517310
## text5347 0.4213480 0.2638480 0.3148039
## text5348 0.2047637 0.5257382 0.2694981
```

What do thoese labels sent1, sent2, sent3 mean? JST is able to estimate 2 (positive, negative) or 3 (neutral, positive, negative) sentiment estimates. Because we opted for the default parameters when running the

model, we estimated all three categories. Essentially, JST results are probabilities that a document belongs to one of the 2 (or in our case 3) sentiment categories. For instance, JST estimated that the probability of the first text (text5243) being neutral is 0.42 (sent1), the probability of it being positive is 0.22 (sent2), and the probability of the text being negative is 0.36 (sent3). We can also use these probabilities to calculate an overall sentiment score. For this we substract the negative score of a text from its positive score.

```
pi %>%
  mutate(sentiment = sent2 - sent3) %>%
  select(sentiment) %>%
  head()
```

```
##            sentiment
## text5343 -0.13373256
## text5344 -0.13772263
## text5345 -0.04089035
## text5346 -0.36283594
## text5347 -0.05095589
## text5348  0.25624011
```

We can also repeat this using the `dtm` object we just created instead of the `dfm`. The results are the same.

```
set.seed(1899)
jst_out <- sentitopics::jst(dtm, dict, numTopics = 30, numIters = 100)
pi <- sentitopics::get_parameter(jst_out, "pi")
pi %>%
  select(sent1, sent2, sent3) %>%
  head()
```

```
##              sent1     sent2     sent3
## text5343 0.2737402 0.2962636 0.4299962
## text5344 0.5426177 0.1598298 0.2975525
## text5345 0.4159158 0.2715969 0.3124873
## text5346 0.2593740 0.1888950 0.5517310
## text5347 0.4213480 0.2638480 0.3148039
## text5348 0.2047637 0.5257382 0.2694981
```

Running JST many times

Similar to LDA models, JST model results usually differ across model model runs to some degree. We can use this variation to compute uncertainty estimates around sentiment scores by running JST several times. To run the model several times, we can use the `jstManyRuns()` function. Here we just have to specify how often we want to run the model. It is important to note here that the function only returns the averaged results of the document-level sentiment scores and their associated uncertainty measures. Keeping all model information would quickly result into reaching RAM limits. In our example we run the model 10 times, and use the default settings for the number of CPU cores (available cores - 3). We could change those settings by using the parameter ncores.

```
set.seed(1899)
res <- sentitopics::jstManyRuns(dfm, dict, numIters = 10, n = 10)

res %>%
  select(sent2_mean, sent2_sd, sent2_se, sent2_ci_high, sent2_ci_low) %>%
  head()
```

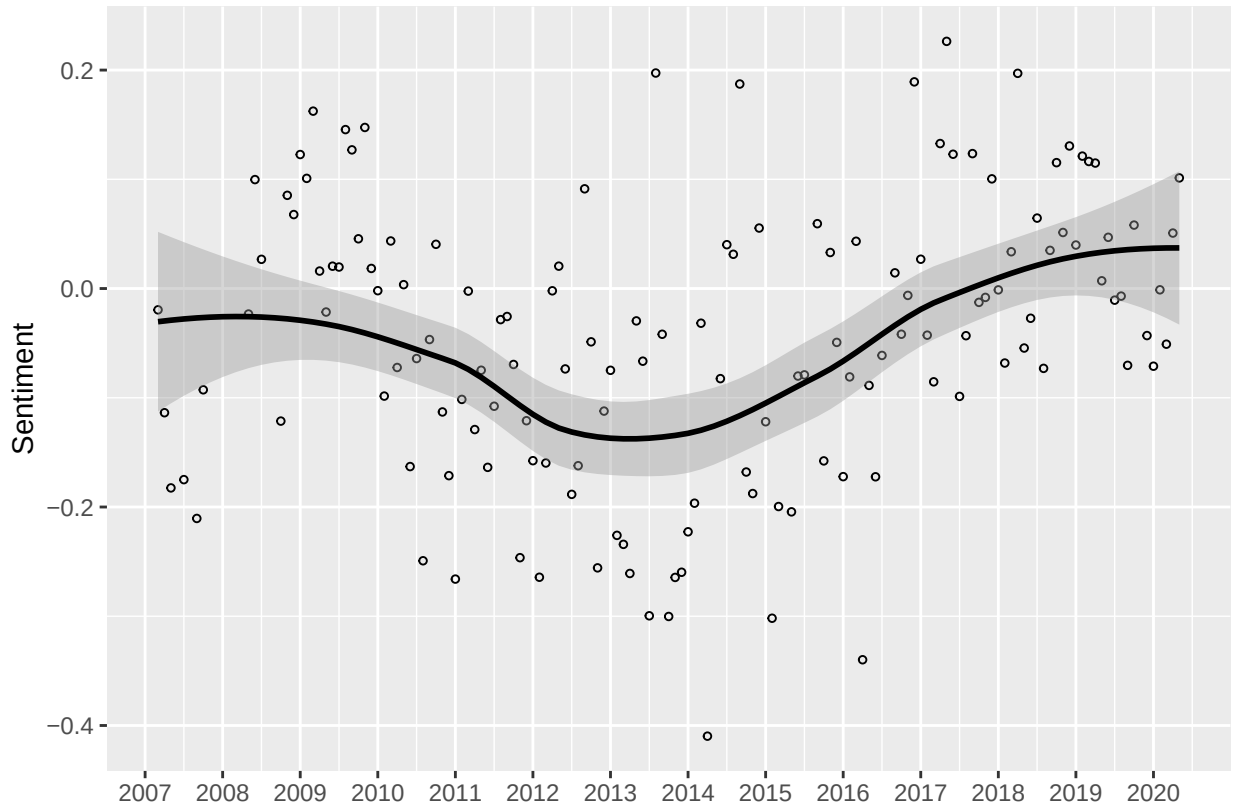
```
##          sent2_mean   sent2_sd   sent2_se sent2_ci_high sent2_ci_low
## text5343  0.3250513 0.05070958 0.01603578     0.3613267    0.2887759
## text5344  0.3528440 0.07379460 0.02333590     0.4056335    0.3000546
## text5345  0.3509723 0.04184602 0.01323287     0.3809072    0.3210375
## text5346  0.2056664 0.03367436 0.01064877     0.2297556    0.1815772
## text5347  0.3555686 0.05701957 0.01803117     0.3963580    0.3147793
## text5348  0.3015955 0.08628857 0.02728684     0.3633227    0.2398684
```

Let's use these last results to investigate how the sentiment of prime minister speeches in the UK developed over time:

```
library(zoo)
data <- res %>%
  mutate(sentiment = sent2_mean - sent3_mean) %>%
  mutate(year_month = as.yearmon(date)) %>%
  group_by(year_month) %>%
  summarise(sentiment = mean(sentiment))


plot <- data %>%
  ggplot(aes(x = year_month, y = sentiment)) +
  geom_point(shape = 1, color = "black", size=1) +
  geom_smooth(color = "black", se = TRUE, size = 1, level = 0.95) +
  ylab("Sentiment") +
  xlab("") +
  scale_x_continuous(name = "",
                     breaks = c(2007:2020))

plot
```



Estimating topic-specific sentiment with rJST

While JST assumes that a document is first structured by its sentiment, rJST assumend that a text is structured by topics first. We can therefore use the rJST model to estimate topic-specific sentiment (e.g. how positive/negative is a text about the EU). Estimating a rJST model with the `sentitopics` can also easily be done with the `jst_reversed()` function. Again, we have to specify the number of topics we expect to find in the corpus. In addition, we also can play around with the hyperparameter settings. In this example we just use the default settings and run the model 100 times (You should use more iterations in a real application).

```
set.seed(1899)
rjst <- sentitopics::jst_reversed(dfm, dict, numTopics = 30, numIters = 1000, alpha = 1, gamma = 50, up
```

How do rJST look like? First, we ce can extract the words that load highly on each topic-sentiment with the `top20words()` and `topNwords()` functions. These words list are similar to what you would get from an LDA mode, but with an important addition: For each topic we get three word lists: One each for neutral, positive, and negative topic-sentiment.

```
words <- rjst %>% sentitopics::top20words(topic = 1)
head(words)
```

```
##   topic1sent1 topic1sent2 topic1sent3 topic2sent1 topic2sent2 topic2sent3
## 1         can      action      gordon       chang        need       parti
## 2         now         now       brown    question      system       polit
```

```
## 3        time     announc       peopl        issu      propos      govern
## 4       futur      measur   difficult       polit         put  parliament
## 5        next     countri      andrew     process        made        hous
## 6      believ       month       peter        want      report       chang
##    topic3sent1 topic3sent2 topic3sent3 topic4sent1 topic4sent2 topic4sent3
## 1 infrastructur       innov      invest     ireland          eu        unit
## 2          plan   technolog        busi    northern          uk        work
## 3           now      scienc    industri   agreement        deal     kingdom
## 4        invest       world     britain       peopl   agreement      remain
## 5        london    research     economi      govern        leav       peopl
## 6          area       futur     british       irish      brexit     continu
##   topic5sent1 topic5sent2 topic5sent3 topic6sent1 topic6sent2 topic6sent3
## 1        busi         tax        make      diseas      servic      health
## 2         one          g8         can    dementia      public         nhs
## 3   enterpris     compani          go       peopl      govern        care
## 4       small    transpar        want        home         new      servic
## 5      govern     corrupt        need       medic       power      hospit
## 6     compani      agenda     countri       advic         way     patient
##   topic7sent1 topic7sent2 topic7sent3 topic8sent1 topic8sent2 topic8sent3
## 1        nato afghanistan        forc       china   terrorist     countri
## 2      defenc      afghan     british      chines      threat     britain
## 3       secur    pakistan        oper       islam      terror       world
## 4        alli       troop         arm       peopl       secur         can
## 5      capabl       secur     support       burma      extrem        open
## 6    aircraft      terror      servic       visit   extremist        take
##   topic9sent1 topic9sent2 topic9sent3 topic10sent1 topic10sent2 topic10sent3
## 1       first     countri      presid        peopl           go          get
## 2       today      climat        unit         want         need         make
## 3        said       chang       prime         just          say          can
## 4        year     develop        work          let        thing        chang
## 5      import   agreement       obama          now         dont         like
## 6         now  copenhagen       state         take         know         help
##   topic11sent1 topic11sent2 topic11sent3 topic12sent1 topic12sent2 topic12sent3
## 1        polic       attack         make      financi        trade       global
## 2     investig       famili        right         bank        world        world
## 3      inquiri          sir        clear        crisi      economi       intern
## 4       happen        offic        today      economi         deal        chang
## 5        press       report         home      countri          g20       togeth
## 6        decis         fire         safe       fiscal       global          new
##   topic13sent1 topic13sent2 topic13sent3 topic14sent1 topic14sent2 topic14sent3
## 1      economi       govern       reform          oil      problem         bank
## 2       growth       budget         fair        price        peopl         lend
## 3       econom       coalit         need         food      countri         busi
## 4          job        spend         mean       energi          can         help
## 5      britain          new        right         rise         deal      financi
## 6      deficit          cut        first       produc          now        money
##   topic15sent1 topic15sent2 topic15sent3 topic16sent1 topic16sent2 topic16sent3
## 1         know       minist        think         issu       immigr        libya
## 2        peopl        prime          got          can       migrat        syria
## 3          say     question     question        think         come       turkey
## 4          tri        peopl           go      discuss      migrant        peopl
## 5          lot         deal          get          see      control      transit
## 6         just         well       actual         want       system       libyan
##   topic17sent1 topic17sent2 topic17sent3 topic18sent1 topic18sent2 topic18sent3
```

```
## 1           uk         work         also        today         year          new
## 2        prime       togeth        thank            £          per          can
## 3 relationship  partnership       cooper      million         need        first
## 4       minist        share         like         year       across      support
## 5        trade        secur      british         fund        deliv         make
## 6      discuss           us          two      billion         last         also
##   topic19sent1 topic19sent2 topic19sent3 topic20sent1 topic20sent2 topic20sent3
## 1       school     opportun      univers     european      eurozon     european
## 2         educ        peopl      student      britain        franc        europ
## 3     children        young         educ        union        europ      council
## 4       parent         work      countri        europ      britain       russia
## 5        everi          get      britain           eu       togeth        union
## 6        child          can      graduat      countri         need         agre
##   topic21sent1 topic21sent2 topic21sent3 topic22sent1 topic22sent2 topic22sent3
## 1       nation         must         time       israel       region       africa
## 2           us         mani         free         peac         iraq          aid
## 3     challeng      respons         live  palestinian       govern      develop
## 4          let         also        great        state      support      million
## 5         just        ensur        everi       jewish        polit        world
## 6          one         work       better        secur      countri     children
##   topic23sent1 topic23sent2   topic23sent3 topic24sent1 topic24sent2
## 1         work         hous         peopl     internet        crime
## 2       famili         home          work        onlin        polic
## 3        money        build           job      compani       prison
## 4          pay        peopl         skill          use    communiti
## 5          tax         plan apprenticeship         data        peopl
## 6            £       afford         train     children       crimin
##   topic24sent3 topic25sent1 topic25sent2 topic25sent3 topic26sent1 topic26sent2
## 1        peopl      nuclear           mr      support        today         unit
## 2    communiti       weapon      speaker         help         year      kingdom
## 3      societi         iran         hous       commit        never     scotland
## 4       social          use    statement     progress        stand         wale
## 5         help       intern       member       intern       rememb     scottish
## 6          can       chemic      council         also          one        union
##   topic26sent3 topic27sent1 topic27sent2 topic27sent3 topic28sent1 topic28sent2
## 1        great        think         weve         well        india       energi
## 2        world        thing          got           im         citi       carbon
## 3         just          say         that       actual       london        renew
## 4        sport          one          get         dont        mayor          low
## 5         team       import         youv         good       indian       invest
## 6      footbal           go          now        right         want        green
##   topic28sent3 topic29sent1 topic29sent2 topic29sent3 topic30sent1 topic30sent2
## 1           uk       celebr        thank        women        peopl        world
## 2       nation        faith         mani commonwealth         valu        power
## 3      support      tonight      countri         girl          can          war
## 4          can      countri         year        equal       believ       nation
## 5         like      british        peopl      slaveri       across     interest
## 6         come    communiti       street          men           us        state
##   topic30sent3
## 1        right
## 2    democraci
## 3        peopl
## 4      freedom
## 5        human
```

```
## 6         free
```