

Machine Learning COSRX Twitter Sentiment Analysis By Cate de Leon

What are the specific topics that people are talking about with regards to COSRX?

For the machine learning portion of our sentiment analysis, we run a Latent Dirichlet allocation (topic modelling) from the text2vec package. This turns our word tokens into a document term matrix and produces bags of related, co-occurring words to reveal possible topics that are being talked about.

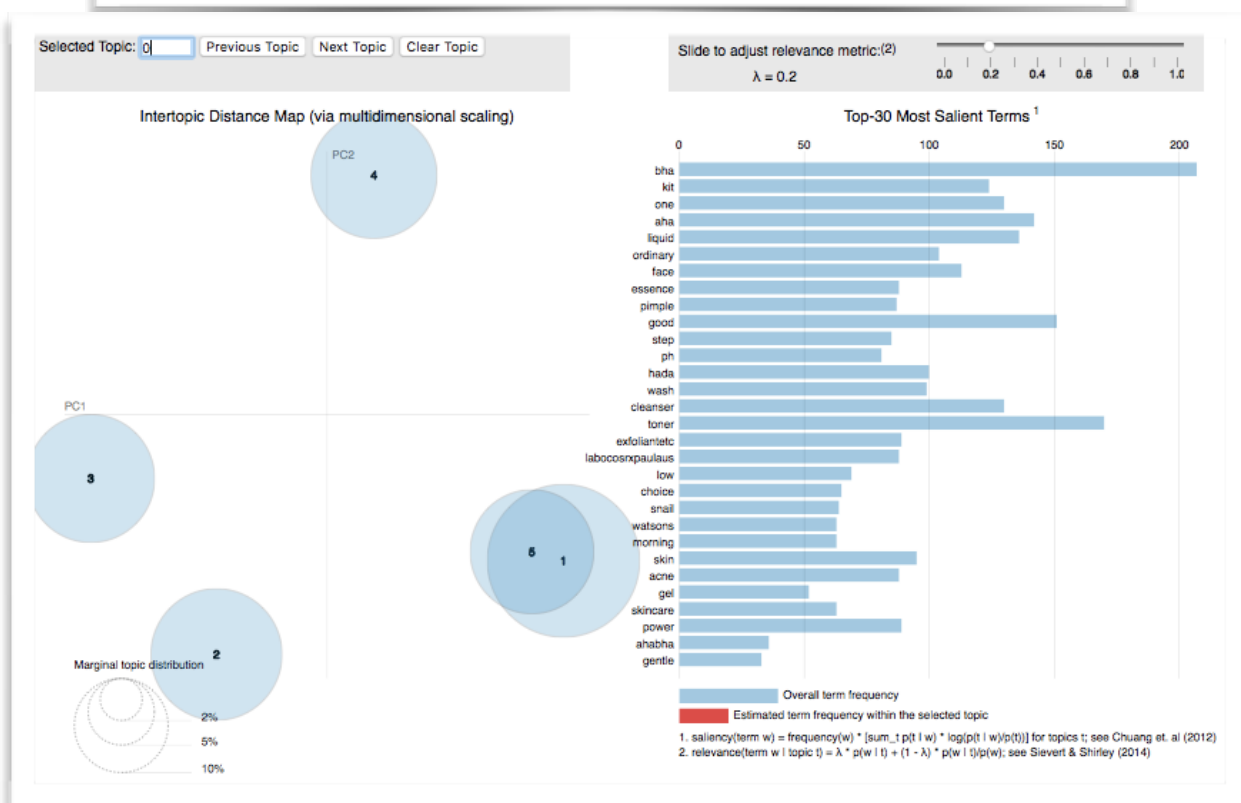
This provides us with a more nuanced view of the discussion, on top of the most frequent words.

```
# Topic modelling
# Run Latent Dirichlet allocation

tokens = tweetsDF$text %>%
  tolower %>%
  word_tokenizer
it = itoken(tokens, progressbar = FALSE)
v = create_vocabulary(it) %>%
  prune_vocabulary(term_count_min = 10, doc_proportion_max = 0.2)
vectorizer = vocab_vectorizer(v)
dtm = create_dtm(it, vectorizer, type = "dgTMatrix")
```

After a few experiments on the model formula, it seemed best to set the number of topics to 5.

```
set.seed(123)
lda_model = LDA$new(n_topics = 5, doc_topic_prior = 0.3, topic_word_prior = 0.0001)
doc_topic_distr =
  lda_model$fit_transform(x = dtm, n_iter = 1000,
    convergence_tol = 0.0001, n_check_convergence = 25,
    progressbar = FALSE)
barplot(doc_topic_distr[1, ], xlab = "topic",
  ylab = "proportion", ylim = c(0, 1),
  names.arg = 1:ncol(doc_topic_distr))
```



We now get the top 10 words for each topic:

```
> lda_model$get_top_words(n = 10, topic_number = c(1L, 2L, 3L, 4L, 5L), lambda = 0.2)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] "ordinary" "ahabha"  "kit"   "ph"   "bha"
[2,] "essence"  "gentle" "one"   "low"  "aha"
[3,] "choice"   "exfoliator" "pimple" "morning" "liquid"
[4,] "snail"    "mask"    "step"  "gel"   "face"
[5,] "paulas"   "two"     "watsons" "oil"   "hada"
[6,] "mucin"    "tight"   "fair"   "cleansing" "wash"
[7,] "c"        "salicylic" "master" "labo"   "exfoliantetc"
[8,] "scars"    "normal"  "advanced" "clear"  "labocosrxpaulaus"
[9,] "love"     "cleanserlg" "patch"  "water"  "toner"
[10,] "glycolic" "mild"    "pixi"   "lotion" "bottle"
> lda_model$plot()
```

From this, we can gather the following topics:

Topic 1 is about brightening:

This is what is commonly achieved by the substances Snail Mucin (aside from hydration), exfoliants such as Glycolic Acid, and Vitamin C. Brightening is commonly desired by customers who are trying to fade scars, such as dark post-acne marks, or have dull skin.

The brands The Ordinary and Paula's Choice were also mentioned in this topic.

Topic 2 is about gently carifying the skin:

This is what is achieved by the substances AHA (alpha hydroxy acid), BHA (beta hydroxy acid), and salicylic acid (which is a form of BHA)—all of which are exfoliators. Their Low pH Good Morning Gel Cleanser contains a form of BHA and is known for being both effective and gentle/mild—which is COSRX's claim to fame with its product line.

Other words in this topic are mask, two, tight, and normal.

Topic 3 is about quick fixes for pimples:

COSRX's One Step Pimple Clear Pads, One Step Pimple Clear Kit, Acne Pimple Master Patch, and Clear Fit Master Patch (stickers that shrink pimples overnight) dominated this topic.

Other words mentioned were watsons (a drugstore/pharmacy), fair, advanced (most probably from the Advanced Snail 96 Mucin Power Essence—which helps heal damaged skin), and the brand Pixi.

Topic 4 is about cleansers.

Users pitted (or perhaps used in conjunction, as double cleansing with oil/micellar water, followed by a foaming cleanser is a popular regimen) COSRX's Low pH Good Morning Gel Cleanser and Hada Labo's Cleansing Oil (against/with) each other.

Other words in this topic are clear, water, and lotion—which can probably be alluded to COSRX's Oil-Free Ultra-Moisturizing Lotion. (Among beauty enthusiasts, moisturization is an important step to follow up cleansing).

Topic 5:

Topic 5 is the least clear cut of the five, but the top words comprise the product Hada Labo Face Wash.

It also mentions BHA, AHA, liquid, exfoliant, and toner, which can be alluded to COSRX's AHA/BHA Clarifying Toner and other AHA and BHA liquid products.

The brand Paula's Choice is also mentioned in this topic.