Quick Sentiment Analysis on AI sample

Paul Oldham

21/05/2018

Rough initial workings on sentiment analysis for AI tweets. Uses the Silge and Robinson *Text Mining with R: A Tidy Approach* from which the code is adapted.

Takes the 9175 AI seed tweets.

```
library(tidytext)
library(tidyverse)
# divide into words and remove stopwords
ai_tokens <- AI %>%
  unnest_tokens(., word, text) %>%
  anti_join(stop_words)

# count up and rank the words
ai_tokens %>%
  count(word, sort = TRUE)
```

```
## # A tibble: 36,286 x 2
##    word                       n
##    <chr>                  <int>
##  1 t.co                   11621
##  2 https                  11607
##  3 ai                     10431
##  4 btc                     1417
##  5 machinelearning         1347
##  6 iot                     1272
##  7 bigdata                 1266
##  8 artificialintelligence   897
##  9 tech                     887
## 10 amp                      774
## # ... with 36,276 more rows
```

```
# join to sentiments dictionary (in this case bing)
ai_sentiment <- ai_tokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)
ai_sentiment
```

```
## # A tibble: 1,157 x 4
##    word            negative positive sentiment
##    <chr>              <dbl>    <dbl>     <dbl>
##  1 abolish               2.       0.       -2.
##  2 absurd                2.       0.       -2.
##  3 absurdly              1.       0.       -1.
##  4 abundance             0.       2.        2.
##  5 abuse                 3.       0.       -3.
##  6 abuses                3.       0.       -3.
##  7 accessible            0.      10.       10.
```

```
##  8 accolade              0.       2.        2.
##  9 accomplished          0.       1.        1.
## 10 accomplishments       0.       1.        1.
## # ... with 1,147 more rows
```

```
# What are the top positive terms in the tweets

ai_sentiment %>%
  arrange(desc(sentiment))
```

```
## # A tibble: 1,157 x 4
##    word         negative positive sentiment
##    <chr>           <dbl>    <dbl>     <dbl>
##  1 intelligence       0.     664.      664.
##  2 innovation         0.     384.      384.
##  3 top                0.     189.      189.
##  4 agile              0.     159.      159.
##  5 smart              0.     151.      151.
##  6 free               0.     128.      128.
##  7 improve            0.      91.       91.
##  8 love               0.      71.       71.
##  9 creative           0.      70.       70.
## 10 intelligent        0.      70.       70.
## # ... with 1,147 more rows
```

```
# What are the top ranking negative terms
ai_sentiment %>%
  dplyr::filter(sentiment <= 0) %>%
  arrange(sentiment)
```

```
## # A tibble: 593 x 4
##    word       negative positive sentiment
##    <chr>         <dbl>    <dbl>     <dbl>
##  1 cloud          327.       0.     -327.
##  2 disruption      86.       0.      -86.
##  3 cancer          67.       0.      -67.
##  4 drones          58.       0.      -58.
##  5 fake            55.       0.      -55.
##  6 die             54.       0.      -54.
##  7 bias            53.       0.      -53.
##  8 melancholy      50.       0.      -50.
##  9 strain          50.       0.      -50.
## 10 risk            48.       0.      -48.
## # ... with 583 more rows
```