The Wreck of the Titan Sentiment Analysis

B.M Njuguna

2023-02-16

Chapter One: Introduction

1.o Background Information

The **Wreck of the Titan** is a novel that was written by American writer **Morgan Robertson** (Robertson [1912]) . It was first published in 1898, and then later republished in 1912. The fictional story is about a ship called *Titan* which sinks in the North Atlantic Ocean, after hitting a iceberg, resulting to loss of most of its passengers and crew. This novel was written 14 yeas before the sinking of the RMS Titanic ship in 1912. The similarities between the fictional story and the actual event of the sinking of the ship in 1912 have led to much speculations about whether Robertson's may have inspired or even predicted the Titanic Disaster.

Robertson describes the ship as a large unsinkable passenger liner with many of the features similar to the titanic (Donnelly [2000]). The ship in the fictional story travels at high speed in a sea filled with icebergs which is similar to the Titanic ship. The book's description about the sinking of the ship have similarities with the actual description of the sinking of the ship. Also, in the book, there are challenges such as lack of enough life boats and confusion among the crew which are similar to the actual sinking of the ship. The books prescient warning about the dangers of the overconfidence in the face of natural disasters.

Sentiment analysis is a technique used to extract, identify and/or identify subjective information from text or novel (Mejova [2009]). It is a sub-field of Natural Language Processing that is also used to identify whether the sentiment of a text, novel or document is positive, negative or neutral. Sentiment analysis started in the 1950s with simple use of lexical (lexicons)[1] and heuristics database to identify and classify words as positive and negative. However, these early attempts were limited by their accuracy and effectiveness.

The field of sentiment analysis advanced significantly at around 2000 with the invention of more sophisticated machine learning algorithms and the availability of large amounts of labelled data for training and testing (Mäntylä et al. [2018]).

[1] Lexicons are dictionaries or lists of words and their meanings, often used in natural language processing and computational linguistics. They can also include additional information about words such as part-of-speech, pronunciation, and syntactic behavior. Lexicons can be created manually or automatically generated from large text corpora.

Objectives of the Study

1. The objective of this study was to examine the general sentiment of the novel *The wreck of the titanic.*

2. Examine the change of emotions across the trajectory of the novel.

3. Compare the results of different performances of the lexicons used.

Methodology

Basically, sentiment analysis method was used in this study. The study was done using *R statistical package.* Specifically, lexicons in R were used to obtain the sentiment of the novels.

First, The R package *gutenbergr* was used to download the novel in R. Then the R package *tidytext* was used to clean and put the text in the novel in the tidy format easier analysis. *tidytext* was also used to download the lexicons for the sentiment analysis. The three lexicons used for the analysis were as listed below;

- **AFINN**

- **Bing**

- **NRC**

All three lexicons are based on unigrams, i.e., single words. These lexicons contain many English words and the words are assigned scores for positive/negative sentiment, and also possibly emotions like joy, anger, sadness, and so forth. The **NRC** lexicon **categorizes words in a binary fashion ("yes"/"no") into categories of positive**, **negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust**. The **Bing** lexicon **categorizes words in a binary fashion into positive and negative categories**. The **AFINN** lexicon **assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment**.

Data Analysis

The novel was downloaded in R and the glimpse of the first part of the first page is shown below;

```
THE WRECK OF THE
      TITAN

   OR, FUTILITY
```

BY

MORGAN ROBERTSON

AUTOGRAPH EDITION
PUBLISHED BY
McCLURE'S MAGAZINE
AND
METROPOLITAN MAGAZINE

Copyright, 1898, by
M. F. MANSFIELD

Copyright, 1912, by
MORGAN ROBERTSON

CONTENTS

 THE WRECK OF THE TITAN;
 OR FUTILITY

THE WRECK OF THE TITAN

CHAPTER I

She was the largest craft afloat and the greatest of the works of men.
In her construction and maintenance were involved every science,
profession, and trade known to civilization. On her bridge were
officers, who, besides being the pick of the Royal Navy, had passed
rigid examinations in all studies that pertained to the winds, tides,
currents, and geography of the sea; they were not only seamen, but
scientists. The same professional standard applied to the personnel of
the engine-room, and the steward's department was equal to that of a
first-class hotel.

Two brass bands, two orchestras, and a theatrical company entertained
the passengers during waking hours; a corps of physicians attended to
the temporal, and a corps of chaplains to the spiritual, welfare of all
on board, while a well-drilled fire-company soothed the fears of nervous
ones and added to the general entertainment by daily practice with their
apparatus.

From her lofty bridge ran hidden telegraph lines to the bow, stern
engine-room, crow's-nest on the foremast, and to all parts of the ship
where work was done, each wire terminating in a marked dial with a
movable indicator, containing in its scope every order and answer
required in handling the massive hulk, either at the dock or at
sea--which eliminated, to a great extent, the hoarse, nerve-racking
shouts of officers and sailors.

From the bridge, engine-room, and a dozen places on her deck the
ninety-two doors of nineteen water-tight compartments could be closed in
half a minute by turning a lever. These doors would also close
automatically in the presence of water. With nine compartments flooded
the ship would still float, and as no known accident of the sea could
possibly fill this many, the steamship _Titan_ was considered
practically unsinkable.

Built of steel throughout, and for passenger traffic only, she carried
no combustible cargo to threaten her destruction by fire; and the
immunity from the demand for cargo space had enabled her designers to
discard the flat, kettle-bottom of cargo boats and give her the sharp
dead-rise--or slant from the keel--of a steam yacht, and this improved
her behavior in a seaway. She was eight hundred feet long, of seventy
thousand tons' displacement, seventy-five thousand horse-power, and on

Then the nrc, AFFIN and bing weren imported and the glimpse
of the first few rows of the different lexicons are as shown below. The
following is the nrc lexicons

word	sentiment
abacus	trust
abandon	fear
abandon	negative
abandon	sadness
abandoned	anger
abandoned	fear

The following is the AFINN lexicon. As mentioned above, the
AFINN lexicon scores the sentiments from -5 to 5, with the most
negative sentiment being -5 and the most positive being 5

word	value
abandon	-2
abandoned	-2
abandons	-2
abducted	-2
abduction	-2
abductions	-2

The lexicon bing lexicon is as shown below;

word	sentiment
2-faces	negative
abnormal	negative
abolish	negative
abominable	negative
abominably	negative
abominate	negative

Most Common Words

The most common words were identified and visualized as follows;
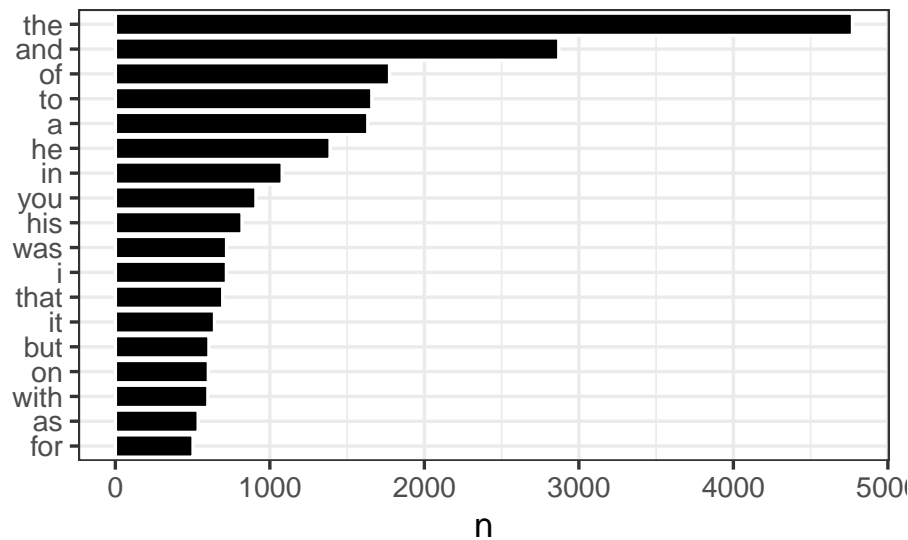
The word `the` was the most used word in the novel, followed by the word `and`. However these words are usually classified as stop words in sentiment analysis and hence showing the frequency of their usage in a novel do not have much interpretation. Thus, the stop words were removed and the following plot was obtained.
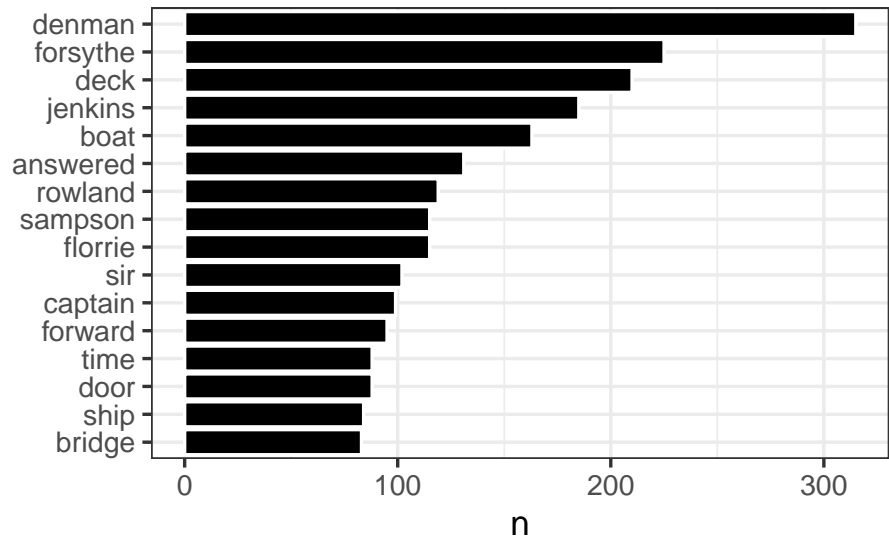

Thus, the most frequently used word in the novel was `Denman`. Denman is a character in the novel who was the owner of the ship. Other words that followed in the frequency of use were deck, jenkins, boat and answered.

Term Frequency

Term frequency is the number of times a word appear in a novel divided by the total number of words in the novel. Below shows some of the words and the number of times used in the novel. This number was inclusive of the stop words.

The word Denman has been used 315 times out of the total 25728 times. The plot below shows the term frequency of the words;
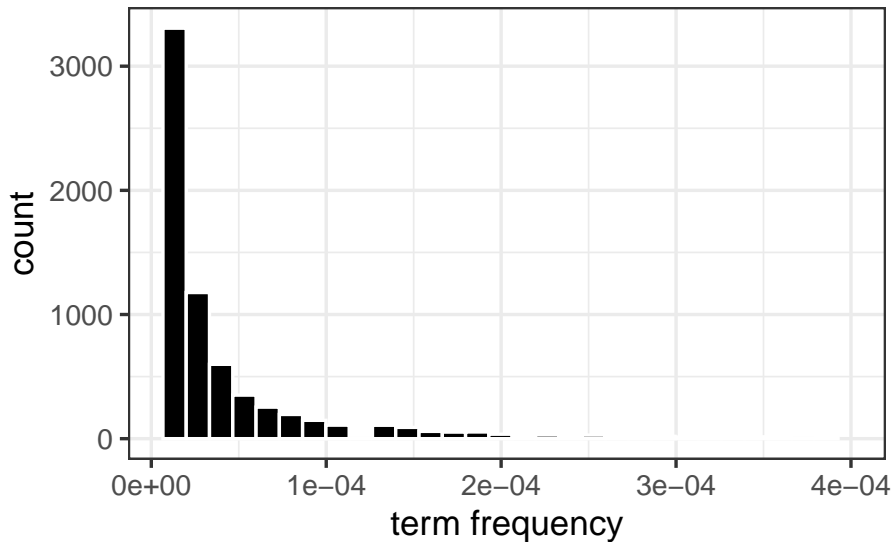
As can been seen from the above plot, there was a long tail to the right, due to the extremely frequent word. The above plot was similar to that of many novels. That is the many words occur rarely and the few words that occur frequently.

Sentiment Analysis with Inner Join

Since the data was in the tidy format. Inner joi was used to classify the general mood of the novel across the trajectory of the novel. The nrc lexicon was used in this analysis.

word	sentiment
sir	positive
sir	trust
captain	positive
forward	positive
time	anticipation
good	anticipation

According to the nrc lexicons, words such as `forwad`, `captian` were classified as positive, while words such as `time` and `good` were classi-

fied as of *anticipation* mood. The table below shows the count of the specific categories of words in the novel as classified by the nrc lexicon

sentiment	total
anger	346
anticipation	259
disgust	216
fear	407
joy	203
negative	778
positive	612
sadness	334
surprise	169
trust	347

The above data was represented in a bar graph as follows;
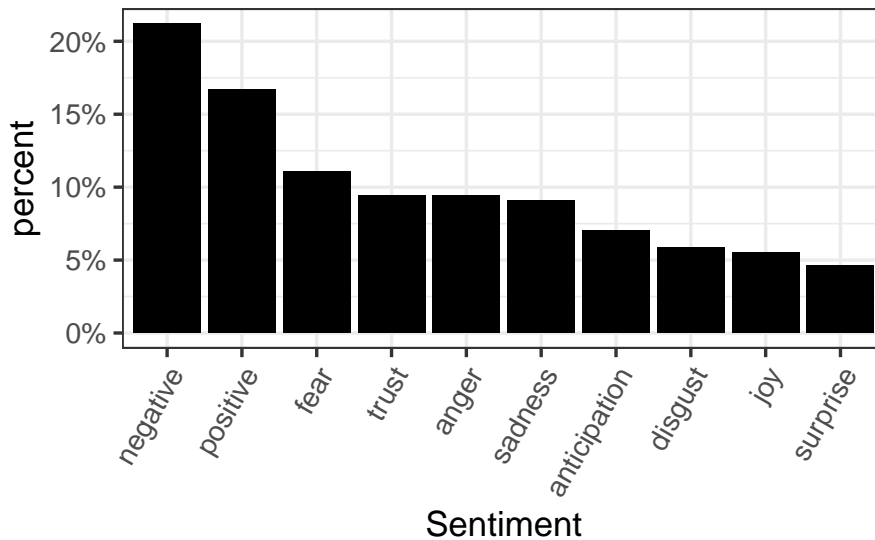
According to the nrc lexicon. most words were negative. The category of words which came second were the positive words while the fear-words came third. The tragic nature of the story may have resulted in the use of negative words-the ship was *unsinkable* yet it sinks after hitting an iceberg. Also, the writinh style at the time at which this novel was written could have attributed to many words being classified as negative.

The bing lexicon was also used to compare its results with the results obtained from the nrc lexicon.

| word	sentiment
well	positive
good	positive
right	positive
like	positive
work	positive
enough	positive

The plot of the positive words versus the negative words was as shown below;
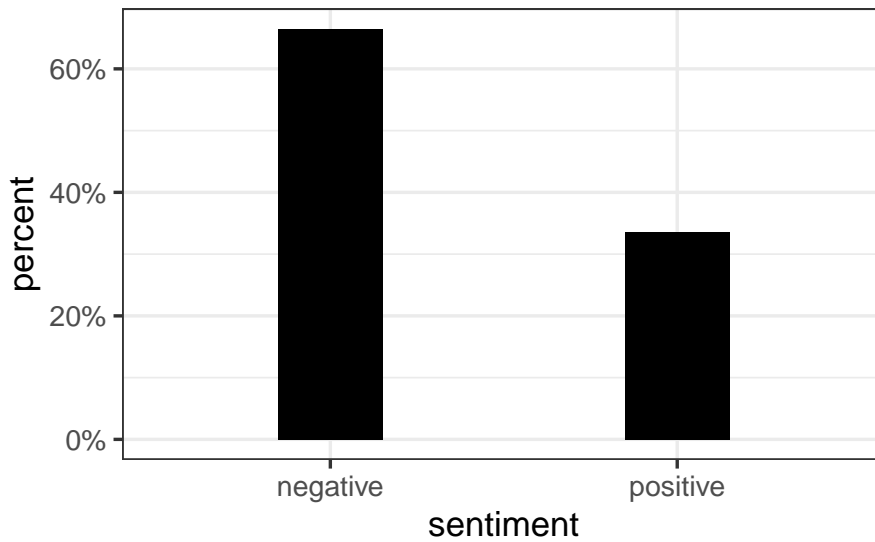
The results obtained from the bing lexicon are similar to the results obtained after using the nrc lexicon. The negative words dominate.

The Sentiments of the Novel in the Novel's Trajectory

To determine how the sentiments changed across the trajectory of the novel. the AFINN lexicon was used. The following graph was obtained;
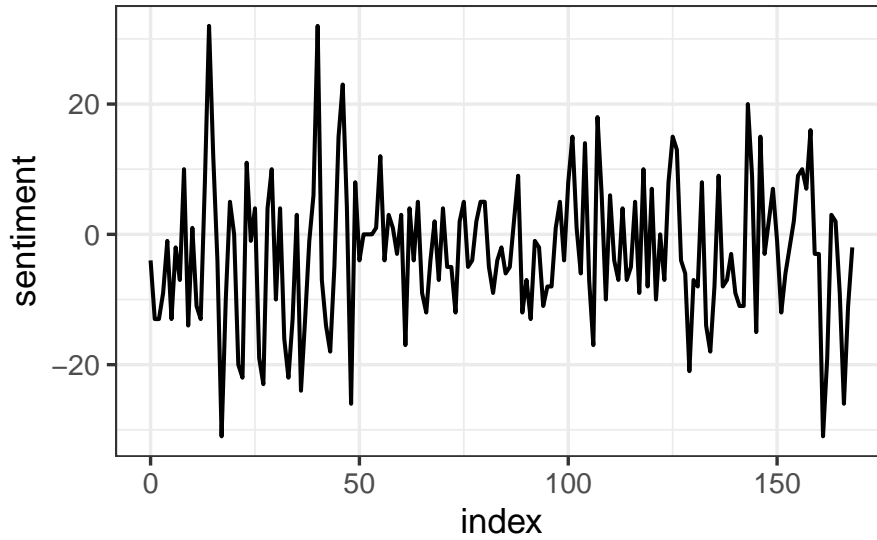
From the plot above, the most positive and negative words are many at the start of the novel. This may be because of the situation when boarding the ship,some people could have used negative words due to occurrences such as delay. They could have also used positive words due to the joy of boarding the popular ship. The score of the words reduce or stabilize towards the center of the novel. At the end of the novel, most words become negative. This could be attributed by the sinking of the ship towards the end of the novel. The cursing, despair could attribute to the negative words.

References

Fred Donnelly. The titanic's humanist paradox. *The Humanist*, 60(4): 36, 2000.

Mika V Mäntylä, Daniel Graziotin, and Miikka Kuutila. The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. *Computer Science Review*, 27:16–32, 2018.

Yelena Mejova. Sentiment analysis: An overview. *University of Iowa, Computer Science Department*, 2009.

Morgan Robertson. *The Wreck of the Titan, Or, Futility.* McClure's magazine and Metropolitan magazine, 1912.