Sentiment analysis in health care – an overview with the help of R software package

1. Introduction

With increase in online media which provides various platforms including the social media for patient to share, discuss and express their experiences related to quality of care received from the health care providers, about the healthcare professional they interacted with them, healthcare facilities they utilized. This has generated vast amount of information in the form of unstructured data which can be useful for decision making for various stakeholders in the healthcare sector. There is a need to build an analytical tool which can help us to analyze the sentiment present in the information generated from the above online sources. This paper provides an overview of sentiment analysis and builds a model to analyze the sentiments through help of R statistical software.

Keywords: Sentiment analysis, text mining, natural language processing, health care, R package

2. Sentiment Analysis

Sentiment analysis [1, 2, 3] helps us to extract the hidden sentiments from the data which is in the form texts using text mining and natural language processing techniques. Sentiment Analysis in health care [4, 5, 6, 7] field helps us to analyze the sentiment from different data sources such websites, blogs and social media which are expressed in the free text.

Text mining [8, 9] helps us to retrieve, preprocess, and classify the text based on the text mining techniques whereas the natural language processing [10] techniques helps to analyze the interaction between human languages the computer systems.


NLP is used in the health care [12, 13] field to analyze the case reports, electronic medical records, medical coding, and robotic assisted surgery.

It is assumed that the readers have basic knowledge of text mining and Natural language processing techniques and can refer the following book to get more information on the text mining and natural language processing techniques.

The paper uses R statistical software package[14] which is an open source package along with its GUI R-Studio[15] which can be downloaded from the below links :

a. https://cran.r-project.org
b. https://www.rstudio.com/

3. **Carrying out sentiment analysis using R package**

To start with the sentiment analysis we need to first extract information from online sources such as websites, blogs, social media websites like twitter. This can be done using htmlTreeParse function of XML package and for that we need install “XML” package[16] and get the retrieve the webpage content using readlines() function

We can also extract the same type of information from the twitter using the twitteR [17,18] r package and we need to create the developer account (app.twitter.com) to access the twitter application to retrieve the tweets. Once the account is created, we need to create an application from the twitter developer page which will give us the following keys to access

1. Consumer Key
2. Consumer Secret

Let us consider the patient feedback on an x hospital on the social media and extract the patient sentiment from the data set (only for the study purpose masking the patient details and hospital details).

In this example we will be using the patient feedback obtained from the NHS website UK and stored as text form for carrying out the sentiment analysis.

To carry out sentiment analysis we can to install the following built in packages through r studio environment:

1. XML[16]
2. Sentiment Analysis[19]
3.1 Install packages

```r
install.packages("XML")
install.packages("Sentiment Analysis")
```

3.2 Define the packages in R environment

```r
library(XML)
library(SentimentAnalysis)
```

3.3 Read the webpage and parse the HTML tags using XML package, readlines and htmlTreeParse and xpathSApply function to the contents in the paragraph tag <p>.

```r
url = "http://..."
it <- readLines(url)
doc = htmlTreeParse(it, useInternalNodes = TRUE)
# retrieve the content of paragraph <p> tag
x2 <- xpathSApply(doc, "/p", xmlValue)
```

3.4 Sentiment score can be computed using analyzeSentiment from sentiment package

```r
X3 <- analyzeSentiment(x2, language = "english", removeStopwords = TRUE, stemming = TRUE)
```

**Output – review scores of 10 reviews**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1 1</td>
<td>0.1 3</td>
<td>0.24 -0.02</td>
<td>0.0 2</td>
<td>0.00 -0.07</td>
<td>0.08 0.01</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 -0.04</td>
<td>0.13 0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.3 1</td>
<td>0.0 7</td>
<td>0.38 0.00</td>
<td>0.0 0</td>
<td>0.00 0.00</td>
<td>0.03 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.17 0.03</td>
<td>0.21 0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3 8</td>
<td>0.0 8</td>
<td>0.46 0.08</td>
<td>0.0 0</td>
<td>0.00 0.08</td>
<td>0.08 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.31 0.00</td>
<td>0.31 0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.3 2</td>
<td>0.0 0</td>
<td>0.32 0.04</td>
<td>0.0 0</td>
<td>0.00 0.11</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.00 0.00</td>
<td>0.29 0.00</td>
<td>0.29 0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1 5</td>
<td>0.0 4</td>
<td>0.19 0.00</td>
<td>0.0 0</td>
<td>0.00 0.06</td>
<td>0.06 0.00</td>
<td>0.00 0</td>
<td>0.00 0</td>
<td>0.00 0</td>
<td>0.13 0.06</td>
<td>0.06 0.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The above output uses the following dictionaries to calculate the sentiment scores negative and positive:

- `SentimentGI`, `NegativityGI`, `PositivityGI` - Harvard IV – General Inquirer Dictionary
- `SentimentHE`, `NegativityHE`, `PositivityHE` - Henry’s Financial dictionary
- `SentimentQDAP`, `NegativityQDAP`, `PositivityQDAP` – Quantitative Discourse Analysis Package

### 3.5 Sentiment direction, summary and plots can be created using converttoDirection, summary and plot functions

`convertToDirection(x3$SentimentQDAP)`

`summary(x3$SentimentLM)`

`plotSentiment(x3$SentimentLM, xlab="Patient reviews")`

#### Output

negative positive positive positive positive positive positive positive positive positive

Levels: negative neutral positive
Summary

Min. 1st Qu. Median Mean 3rd Qu. Max.

-0.03571 0.10960 0.15004 0.16598 0.25974 0.32000

References


14. https://cran.r-project.org/

15. https://www.rstudio.com/


17. https://cran.r-project.org/web/packages/twitteR/twitteR.pdf


19. https://cran.r-project.org/web/packages/SentimentAnalysis/SentimentAnalysis.pdf