INTERNATIONAL JOURNAL OF STATISTICS AND MEDICAL INFORMATICS

Vol. 11 No. 1/2019, pp: 1-

# 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.

Text mining in health care [11] helps us to uncover association between disease and prognostic variables, drug-drug reaction and adverse event reporting from electronic medical records and social media.

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

"Editor IJSMI (2018), Application of statistical tools in biomedical domain: An overview with help of software ISBN: 978-1986988551, www.ijsmi.com/book.php",

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**

*install.packages("XML")*

*install.packages("Sentiment Analysis)"*

**3.2 Define the packages in R environment**

library(XML)

library(SentimentAnalysis)

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

*url="http://... "*

*it<-readLines(url)*

*doc=htmlTreeParse(it,useInternalNodes = TRUE)*

*#retreive the content of paragraph <p> tag*

x2<-xpathSApply(doc, "//p", xmlValue)

**3.4 Sentiment score can be computed using analyzeSentiment from sentiment package**

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

*Output – review scores of 10 reviews*

| | WordCount | SentimentGI | NegativityGI | PositivityGI | SentimentHE | NegativityHE | PositivityHE | SentimentLM | NegativityLM | PositivityLM | RatioUncertaintyLM | SentimentQDAP | NegativityQDAP | PositivityQDAP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 112 | 0.12 | 0.13 | 0.24 | -0.02 | 0.02 | 0.00 | -0.07 | 0.08 | 0.01 | 0.00 | -0.04 | 0.13 | 0.10 |
| 2 | 29 | 0.31 | 0.07 | 0.38 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.00 | 0.17 | 0.03 | 0.21 |
| 3 | 13 | 0.38 | 0.08 | 0.46 | 0.08 | 0.00 | 0.08 | 0.08 | 0.00 | 0.08 | 0.00 | 0.31 | 0.00 | 0.31 |
| 4 | 28 | 0.32 | 0.00 | 0.32 | 0.04 | 0.00 | 0.04 | 0.11 | 0.00 | 0.11 | 0.00 | 0.29 | 0.00 | 0.29 |
| 5 | 47 | 0.15 | 0.04 | 0.19 | 0.00 | 0.00 | 0.00 | -0.06 | 0.06 | 0.00 | 0.06 | 0.13 | 0.06 | 0.19 |

| | WordCount | SentimentGI | NegativityGI | PositivityGI | SentimentHE | NegativityHE | PositivityHE | SentimentLM | NegativityLM | PositivityLM | RatioUncertaintyLM | SentimentQDAP | NegativityQDAP | PositivityQDAP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 88 | 0.32 | 0.07 | 0.39 | 0.00 | 0.01 | 0.01 | 0.01 | 0.03 | 0.05 | 0.00 | 0.18 | 0.05 | 0.23 |
| 7 | 36 | 0.19 | 0.14 | 0.33 | 0.03 | 0.00 | 0.03 | 0.00 | 0.08 | 0.08 | 0.03 | 0.11 | 0.08 | 0.19 |
| 8 | 25 | 0.24 | 0.04 | 0.28 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.04 | 0.12 |
| 9 | 25 | 0.40 | 0.04 | 0.44 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.04 | 0.00 | 0.32 | 0.04 | 0.36 |
| 10 | 55 | 0.05 | 0.13 | 0.18 | 0.00 | 0.00 | 0.00 | -0.02 | 0.05 | 0.04 | 0.04 | 0.11 | 0.05 | 0.16 |

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
- SentimentLM, NegativityLM, PositivityLM,RatioUncertaintyLM - Loughran-McDonald
- SentimentQDAP, NegativityQDAP, PositivityQDAP – Quantitatiev 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

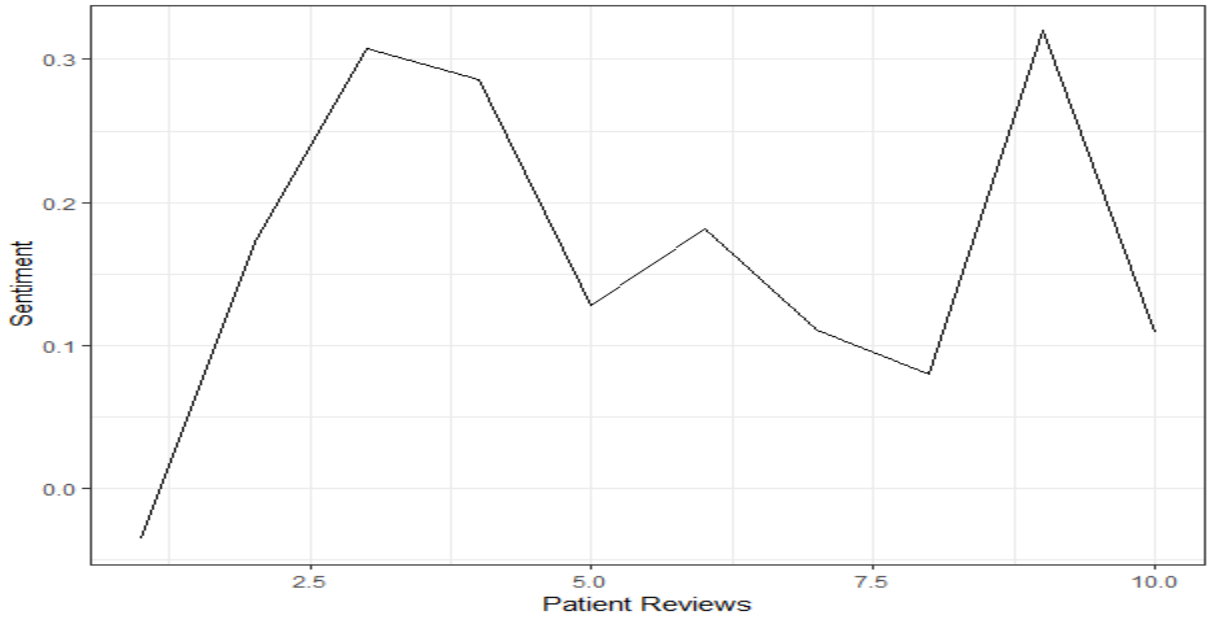
negative positive positive positive positive positive positive positive positive positive

Levels: negative neutral positive

Vol. 11 No. 1/2019, pp: 1-

**Summary**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| -0.03571 | 0.10960 | 0.15004 | 0.16598 | 0.25974 | 0.32000 |




**References**

1. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.
2. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
3. Saxena, D., Gupta, S., Joseph, J., & Mehra, R. (2019). SENTIMENT ANALYSIS. Journal Homepage: http://www. ijesm. co. in, 8(3).
4. Davcheva, E., Adam, M., & Benlian, A. (2019). User Dynamics in Mental Health Forums-A Sentiment Analysis Perspective(No. 110634). Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL).

5. Ramírez-Tinoco, F. J., Alor-Hernández, G., Sánchez-Cervantes, J. L., del Pilar Salas-Zárate, M., & Valencia-García, R. (2019). Use of Sentiment Analysis Techniques in Healthcare Domain. In *Current Trends in Semantic Web Technologies: Theory and Practice* (pp. 189-212). Springer, Cham.
6. Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of sentiment analysis for capturing patient experience from free-text comments posted online. Journal of medical Internet research, 15(11), e239.
7. Khan, M. T., & Khalid, S. (2016). Sentiment analysis for health care. In *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 676-689). IGI Global.
8. Editor IJSMI(2018), Application of statistical tools in biomedical domain, IJSMI, ISBN: 978-1986988551
9. Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.
10. Manning, C. D., Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press.
11. Popowich, F. (2005). Using text mining and natural language processing for health care claims processing. ACM SIGKDD Explorations Newsletter, 7(1), 59-66.
12. Spyns, P. (1996). Natural language processing in medicine: an overview. Methods of information in medicine, 35(04/05), 285-301.
13. Edgcomb, J. B., & Zima, B. (2019). Machine Learning, Natural Language Processing, and the Electronic Health Record: Innovations in Mental Health Services Research. Psychiatric Services, appi-ps.
14. https://cran.r-project.org/
15. https://www.rstudio.com/
16. https://cran.r-project.org/web/packages/XML/XML.pdf
17. https://cran.r-project.org/web/packages/twitteR/twitteR.pdf
18. Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).
19. https://cran.r-project.org/web/packages/SentimentAnalysis/SentimentAnalysis.pdf