IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906158 |

Application of Sentiment Analysis

Gavali Monish Shivaji, Brahmankar Yash Sanjay, Dhamani Tushar Babulal, Lagad Anjali Ganpat,

Mr. Bhalerao Rushikesh S.

Department of Information Technology, SVIT, Nashik, Maharashtra, India

Department of Information Technology, SVIT, Nashik, Maharashtra, India

Department of Information Technology, SVIT, Nashik, Maharashtra, India

Department of Information Technology, SVIT, Nashik, Maharashtra, India

HOD, Department of Information Technology, SVIT, Nashik, Maharashtra, India

ABSTRACT*:* Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gained much attention in recent years. In this paper, we aim to avoid the problem of sentiment polarity classification, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews. Experiments for both sentence-level categorization and review-level classification are performed with outcomes. At last, we also give insight into our future work on sentiment analysis.

I.INTRODUCTION

Sentiment is an opinion, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities. Internet is a resourceful place with respect to sentiment information. From a user's perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces, prompting data collection and analysis by researchers and developers. Hence, sentiment analysis seems having a strong fundament with the support of massive online data. Human conduct is incredibly affected by their abstract sentiments and convictions, for example, demeanor, feeling, assessment or slant. The choices we make can be affected by others' impressions of the world to an extensive degree, on the grounds that conveying the others' assessments is wired into every individual normally and portrays us as 'social creatures'. Nowadays, social communication channels like Twitter, Facebook, and YouTube have obtained so much popularity. Opinion mining is the other name of Sentiment Analysis which is under the category of machine learning and data mining. From the use of different social media, opinion mining or sentiment analysis techniques have to start with people's data for the analysis of a different kind of area like politic, economy or biology, etc. Massive amount of information related to distinct individual entities are recorded every day in digital forms. And hence such a fast growth of the field co-exist along with other social media related stuff such a forums discussion ,blogs, customer reviews ,Twitter and social network sites. Sentiment analysis includes classification of data into various classes like optimistic i.e. good sense or negative i.e. bad sense or neutral i.e. non-effective.

II.RESEARCH DESIGN AND METHODOLOGIES

1. Data collection:

Customer's usually express their sentiments on public forums or on media like the blogs, discussion boards, product reviews as well as on their private logs – Social media sites like Facebook and Twitter. Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, making the data huge and disorganized. Manual sentiment analysis of data is virtually impossible. Therefore, special programming languages like 'R' and 'Python' are used to process and analyze the data.

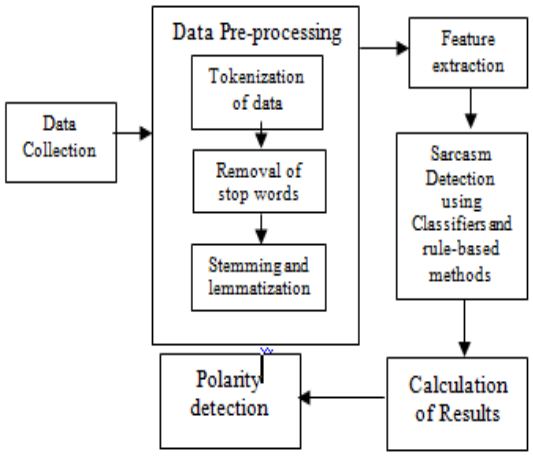
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906158 |

2. Data Preprocessing:

The preprocessing is done by avoiding the noise from the dataset. In other word, dataset is advanced according to prerequisite. The strategy utilized as a part of preprocessing is feature extraction. The following pre-processing steps are followed to scoring the sentence.



3. Text Preparation :

Text preparation is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

4. Tokenization:

Tokenization is the process of breaking a sequence of string into pieces. It may be words, keyword, symbols, keyword, sentences and other elements called tokens. Tokens can be a words, phrases or they may be whole sentences. In this process some characters like punctuation marks are removed and these tokens become the input to the other task such as parsing or text mining.

5. Feature Extraction:

After finishing the preprocessing, the next phase is to extract the feature from the preprocessed information. The main focus is to compute the score for sentence.

6. Stop words:

In computing, stop words are words which are filtered out before or after processing of naturallanguage data (text). Though "stop words" usually refers to the most common words in a language, there is no single universal list of stop words used by all naturallanguageprocessing tools, and indeed not all tools even use such a list. Some tools specifically avoid removing these stop words to support phrasesearch.

7. Bag-of-words Model :

Assuming their conditional independence, it takes an individual word from a sentence as a feature. Texts are arranged as an unordered collection of words. Each word is independent from the other. The challenge is the choice of words that are going to become feature. The sentence “This is a good movie” can be represented in feature vector as:
F0 = [`a`:1, `event`:1, `good`:1, `is`:1, `this`:1]

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906158 |

8. Sentiment Detection :

At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. Sentiment analysis is done at different levels using common computational techniques like Unigrams, lemmas, negation and so on.

9. Sentiment Classification :

Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.

Naive Bayes classifier is a popular supervised classifier, furnishes a way to express positive, negative and neutral feelings in the web text. Naive Bayes classifier utilizes conditional probability to classify words into their respective categories. The benefit of using Naive Bayes on text classification is that it needs small dataset for training. The raw data from web undergoes preprocessing, removal of numeric, foreign words, html tags and special symbols yielding the set of words. The tagging of words with labels of positive, negative and neutral tags is manually performed by human experts. This preprocessing produces word-category pairs for training set.

12. TF-IDF :

TF-IDF basically means Term Frequency-Inverse Document Frequency. It weighs and assigns importance to a keyword based on the number of times it appeared in a document. Weight(Wt) of term(t) in a document(d) is given by:
Wtd = TFtdlog[N/DFt]
TFtd : Number of occurrences of term t in document d
DFt : Number of documents containing the term t
N: Total number of documents

Project Preview :

```
import pandas as pd
import numpy as np
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn import naive_bayes
from sklearn.metrics import roc_auc_score,accuracy_score
import pickle
nltk.download("stopwords")
dataset = pd.read_csv('reviews.txt',sep = '\t', names =['Reviews','Comments'])
stopset = set(stopwords.words('english'))
vectorizer = TfidfVectorizer(use_idf = True,lowercase = True, strip_accents='ascii',stop_words=stopset)
X = vectorizer.fit_transform(dataset.Comments)
y = dataset.Reviews
pickle.dump(vectorizer, open('tranform.pkl', 'wb'))
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)
clf = naive_bayes.MultinomialNB()
clf.fit(X_train,y_train)
accuracy_score(y_test,clf.predict(X_test))*100
clf = naive_bayes.MultinomialNB()
clf.fit(X,y)
accuracy_score(y_test,clf.predict(X_test))*100
your_review = input("Give us one review of any movie = ")
movie_review_list = np.array([your_review])
movie_vector = vectorizer.transform(movie_review_list)
pred = clf.predict(movie_vector)
#pred = clf.predict(your_review)
if pred:
    print("Positive Review")
else:
    print("Negative Review")
```

III.CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process (Figure 2) has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906158 |

REFERENCES

1. Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367.. Association for Computational Linguistics, Stroudsburg, PA, USA
2. Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition.. Taylor and Francis Group, Boca.
3. Liu B, Hu M, Cheng J (2005) Opinion observer: Analyzing and comparing opinions on the web In: Proceedings of the 14th International Conference on World Wide Web, WWW '05, 342–351.. ACM, New York, NY, USA.
4. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining In: Proceedings of the Seventh conference on International Language Resources and Evaluation.. European Languages Resources Association, Valletta, Malta.
5. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04.. Association for Computational Linguistics, Stroudsburg, PA, USA.
6. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends InfRetr2(1-2): 1–135.
7. Turney PD (2002) Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, 417–424.. Association for Computational Linguistics, Stroudsburg, PA, USA.
8. Whitelaw C, Garg N, Argamon S (2005) Using appraisal groups for sentiment analysis In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM '05, 625–631.. ACM, New York, NY, USA.
9. Twitter (2014) Twitter apis. https://dev.twitter.com/start.
10. Liu B (2014) The science of detecting fake reviews. http://content26.com/blog/bing-liu-the-science-of-detecting-fake-reviews

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details