ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com
Journal DOI: 10.21474/IJAR01

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

MY OPINION: ANRGE TO ENRICH YOUR KNOWLEDGE.

Shubhankar Mujumdar, Sudarshan Mane , Manish Kumar, Rohit More, Mr.Sagar Salunke.
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune.

Manuscript Info

Manuscript History:

Received: 15 March 2016
Final Accepted: 22 April 2016
Published Online: May 2016

Key words:
Navie Bayes , NLP, Opinion mining, Sentiment analysis, SQL alchemy.

Corresponding Author

Shubhankar Mujumdar.

Abstract

At present people are using various online platforms for discussion purpose. We can ask questions and get answers to our questions for other people using that platform. Some of these platforms are: Quora, face book pages, Google plus, etc. and these questions are from various fields like politics, science and technology, sports, entertainment, education, etc. The answers that we get on these platforms are mainly in the form of comments that include text data. To get a complete understanding of what is going on the user have to read all the comments and some of this comments can even have same content. It is wastage of time and energy and still the user doesn't get the answer that he is looking for. A way to provide better solution to these questions is to analyze the text data and generate a report of it. Report will be the summarized version of the entire conversation. Depending on few parameters report will be generated that can be referred by the user at any time. It will be a free platform developed in android. It will require a working internet connection. At present it will be used as a simple report generation platform but in future depending upon the user base it can extended for survey and research purpose by individuals and many different organization.

Introduction:-

In today's world there are a lot of applications for different purpose available in android. You think of something and you'll probably find an application for that purpose. Life is becoming very simple because of this. Still there are some areas where improvement is required. One of those areas is the online discussion platforms that provide a platform to its users where they can discuss some topic or can ask questions related to different fields. These fields can be sports, education, religion, politics, entertainment, profession, current affairs etc. The platforms that we are talking about are Quora, face book pages, Google plus, etc.. On these platforms we start with asking a question by posting it on the platform. It is categorized in the type of field to which the question belongs by using keyword generation algorithms and clustering algorithms. Once the categorization is done, other users can view it and if they like the question and have answer for that question they can reply to it. Answers are mainly in text format visible to everyone. Anyone having the same question can search for it and read the answers. Even though these platforms do a great job in providing answers to these questions but there are some problems with these platforms. On these platforms the answers that are post are very long in size and to understand what it says one has to read the entire answers. It consumes more time. Also the data redundancy is present in many answers that mean many replies can have same content and it means same. Also the numbers of replies are huge in some cases. To get complete understanding of what is going on one has to read the entire conversation and it is time consuming. My Opinion is such a platform that will generate analyzed report on the basis of replies received for question. Here AnRGE stands for Analysis and Report Generation Engine. There is a need to develop a platform where data redundancy is minimum and user get a more accurate answer to their question within minimum time period. The objectives of developing this platform are: minimize the data redundancy in comments, provide better answers to the questions, save our user from wasting his precious in reading the entire thread of replies by generating an analyzed report.

Literature survey:-

I. Rish[1] has proposed the Navie Bayes classifier.It tells all about navie bayes classifier and its effectiveness. G.Angulakshmi [2] discusses about opinion mining techniques, opinion summarization, sentiment analysis and tools used.ArtiBuche [3] proposed the work on how text is classified by Navie Bayes algorithm. Edward Loper[4]tells us about NLTK(Natural Language Toolkit) its working and uses. In S.G. Bhirud [5] feature based Opinion Mining is focused. Nidhi Mishra [6] has told about Classification of Opinion Mining Techniques and also discusses on various tasks of opinion mining and sentiment analysis have been discussed. G Jaganadh[7] focuses on opinion mining and sentiment analysis .In [8] concepts like feature ranking semantic analysis and semantic orientation are discussed. [9] Gives detailed explanation about semantic analysis.

C.J. Hutto [10] VADER, a simple rule-based model for general sentiment analysis, and compare its effectiveness to eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms.[11] tells about . Janyce M.Wiebe[12] has proposed about learning subjective adjectives from corpora. Gobinda G. Chowdhury [13] tells about natural language processing (NLP).

Methodology:-

Before going to methodology we will see some of the terms that are required to understand the main method easily and at last we will see the main architecture and working of the application.

The Naive Bayes classifier [1] greatly simplifies learning by assuming that features are independent given class. Although independence is generally a poor assumption, in practice naive Bayes often competes well with more sophisticated classifiers. Our broad goal is to understand the data characteristics which affect the performance of naive Bayes.

NLTK [4] provides a simple, extensible, uniform framework for assignments, projects, and class demonstrations. It is well documented, easy to learn, and simple to use. NLTK allow computational linguistics classes to include more hands-on experience with using and building NLP components and systems.
NLTK is unique in its combination of three factors. First, it was deliberately designed as courseware and gives pedagogical goals primary status. Second, its target audience consists of both linguists and computer scientists, and it is accessible and challenging at many levels of prior computational skill. Finally, it is based on an object-oriented scripting language supporting rapid prototyping and literate programming.

SQLAlchemy is an open source SQL toolkit and object-relational mapper (ORM) for the Python programming language released under the MIT License. SQLAlchemy provides "a full suite of well known enterprise-level persistence patterns, designed for efficient and high-performing database access, adapted into a simple and Python domain language. SQLAlchemy's philosophy is that SQL databases behave less and less like object collections the more size and performance start to matter, while object collections behave less and less like tables and rows the more abstraction starts to matter. For this reason it has adopted the data mapper pattern rather than the active record pattern used by a number of other object-relational mappers. However, optional plugins allow users to develop using declarative syntax. SQLAlchemy is the Python SQL toolkit and Object Relational Mapper that gives application developers the full power and flexibility of SQL. SQL databases behave less like object collections the more size and performance start to matter; object collections behave less like tables and rows the more abstraction starts to matter. SQLAlchemy aims to accommodate both of these principles. SQLAlchemy considers the database to be a relational algebra engine, not just a collection of tables. Rows can be selected from not only tables but also joins and other select statements; any of these units can be composed into a larger structure. SQLAlchemy's expression language builds on this concept from its core.

Natural Language Processing (NLP) [12] is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform the desired tasks. The foundations of NLP lie in a number of disciplines, viz. computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc. Applications

of NLP include a number of fields of studies, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval

VADER,[10] a simple rule-based model for general sentiment analysis, and compare its effectiveness to eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms. Using a combination of qualitative and quantitative methods, we first construct and empirically validate a gold-standard list of lexical features (along with their associated sentiment intensity measures) which are specifically attuned to sentiment in micro blog like contexts. We then combine these lexical features with consideration for five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity.

Opinion mining [2] is a technique which is used to detect and extract subjective information in text documents In general; sentiment analysis tries to determine the sentiment of a writer about some aspect and also the overall contextual polarity of a document. The sentiment may be his or her judgment, mood or evaluation. A key problem in this area is sentiment classification, where a document is labeled as a positive or negative evaluation of a target object (film, book, product etc.)The evaluation of opinion can be done in two ways:

Direct opinion: - It gives positive opinion or negative opinion about the object directly. For example, "The picture quality of this camera is poor" expresses a direct opinion.

Comparison: - means to compare the object with some other similar objects. For example, "The picture quality of camera-y is better than that of Camera-x." expresses a comparison.

The work flow of opinion mining is as follows:

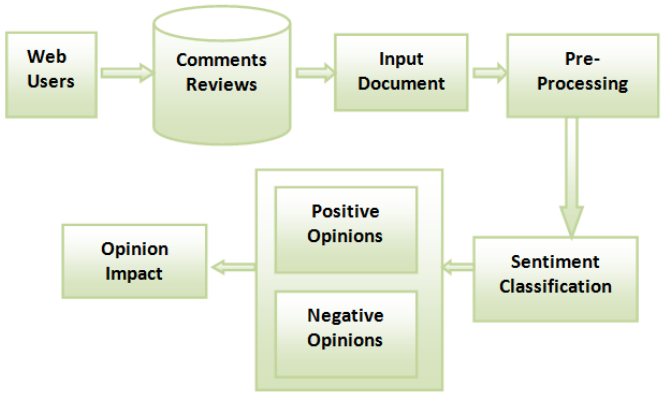


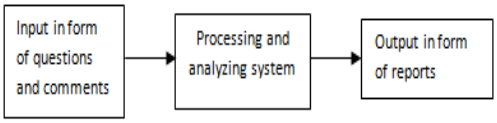
Figure 1:-

Flask is a micro web framework written in Python and based on the Werkzeug toolkit and Jinja2 template engine. It is BSD Licensed. Flask is called a micro framework because it does not presume or force a developer to use a particular tool or library. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Extensions are updated far more regularly than the core Flask program.

Corpora is a lightweight, fast and scalable corpus library able to store a collection of raw text documents with additional key-value headers. It uses Berkeley DB (bsddb3 module) for index managing what guarantee speed and bullet-proof. Text storage model is based on chunked flat, human readable text files. This architecture can easily scale up to millions documents, hundred of gigabytes collections. Corpora module provides four main features: create a new corpus, append documents to a corpus, random access to any document in a corpus using it's unique ``id``, sequential access to document collection (generator over collection).

The main methodology or implementation is as follows:
The discussion platforms that are in use in present times provide solution to various questions being asked in the form of comments. To understand the entire discussion the user has to read the complete discussion which is time consuming and doesn't even provide a satisfactory solution for the question. An android based application has been developed that makes use of NLP to perform opinion mining in order to extract information and use it to generate report to provide a better solution for the question

An android based application that makes use of data mining techniques to extract information and use it to generate report to provide a better solution for the question. The size of the input is not fixed but the type of input is related to current affairs. This means that the type of question posted on this platform is related to current affairs. Input validation is required to check whether the question posted is related to current affairs or not and if the question is not according to the bounds it will not be posted. The input is dependent on the bounds specified in the system.



I/O transition diagram

Figure 2:-

The problem that is being worked requires algorithms. So by using NLP and opinion mining technique to perform sentiment analysis on the comments posted by user to find out the positive or negative context of comments posted. Opinion mining is the computational study of people's opinions, appraisals and emotions toward entities, events and their attributes. It involves techniques from different disciplines like information retrieval, natural language processing and data mining .Opinions are so important those whenever one needs to make a decision, one wants to hear others' opinions. This is true for both individuals and organizations. If an individual wants to purchase a product, it is useful to see a summary of opinions of existing users so that he/she can make decision. This is better than reading a large number of reviews. He/she can also compare the summaries of opinions of different products, instead of reading a large number of reviews.

Using opinion mining a review can be evaluated at three different levels- at document level, sentence level and feature level. When review is evaluated at document level, whole review is classified into either positive or negative depending upon the opinion expressed in that review. When review is evaluated at sentence level, then each sentence in a review is classified into either positive or negative. Whereas feature level or feature based opinion mining gives summary which feature of product is liked or disliked by reviewer The major tasks of feature based opinion mining are.

1. To identify the products features in review,
2. To determine opinion expressed by the reviewer (positive, negative or neutral),
3. Summarize discovered information.

Motivation in this work is to develop an opinion search engine that will not only mine the opinions but will also extract useful information related to the item's features and use it to rate them as positive, neutral, or negative. This feature-based opinion mining will help the user focus on the features of the opinion/product he/she is interested in.

The main data source for the application is the input from the user which will be in simple text format. The data from the data set is pre-processed so as to set the data in the format which is acceptable to the data processing algorithms. For example, the tag [t] is inserted at the beginning of the title to indicate that the sentence following [t] is the title of the review. Moreover, the reviews' file which corresponds to a particular product is split into text files containing one review each.

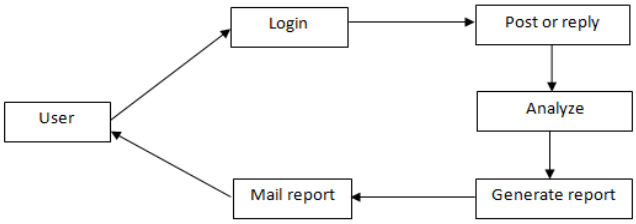
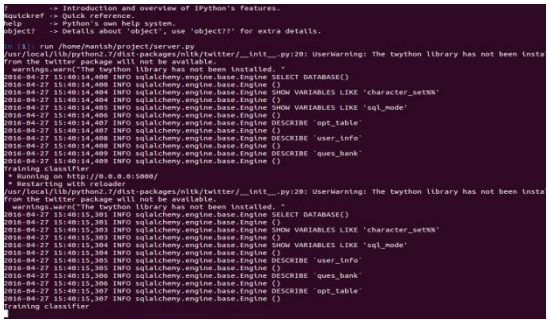


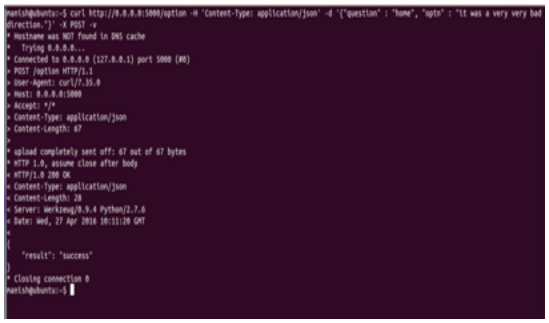
Figure 3:-

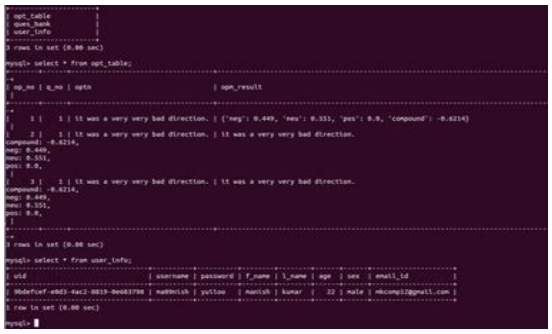
Figure (3) is architecture of the application .The user will login by giving his user name and password. It will go to database and see if the user name and password are matching. If both are matching the user will login successfully .If any one of the above is not matching the user will have to re login .The new user will have to open an account for using the application. After the user has logged in user can check for the notifications for the question that user has posted .Then user can post the question or may reply to the posted questions. This data will be stored in the database. Database is the one that will be with system. It contains all the details of the registered user and all the question posted by user and the answers. Administrator is the only person who has full rights to access the database and generate the report. Only he can make any changes to the database and provide report to the question completely deny the invalid question. The user will not have the permission to access the database of the application

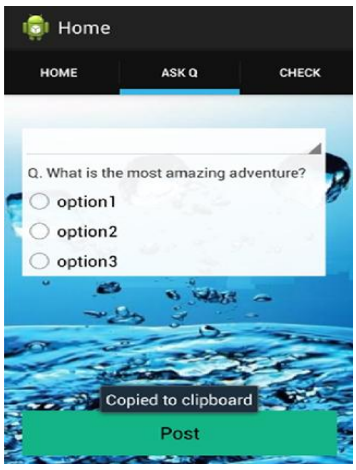
When the user has give reply to the questions or has posted any question then the question or the reply will be analyzed. The sentences of the question or reply will be broken then parsed, and then they will be processed by applying the techniques which are specified above. All this process will take place in the back end of the application. The system will analyze and provide a report on the analysis done in terms of report, bar charts, graphs, pie charts and will give a complete report on the various discussions and various questions which will be posted by user summarizing it and also using the opinion mining techniques of NLP, we will be able to perform sentiment analysis of the discussion. The user can also mail the report if he wants. Thus a better solution will be provided to the user, and also save the time of the other user who want to read the entire discussion with the help of the report which will be generated if user wants to generate.

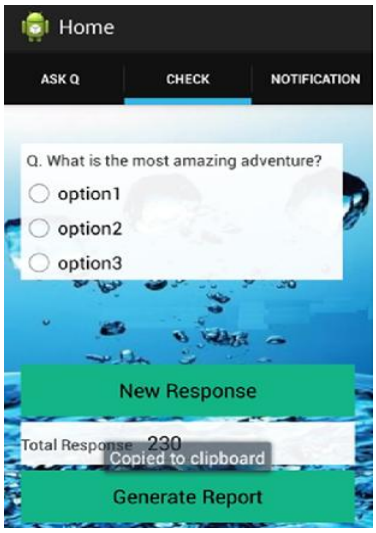
Results:-

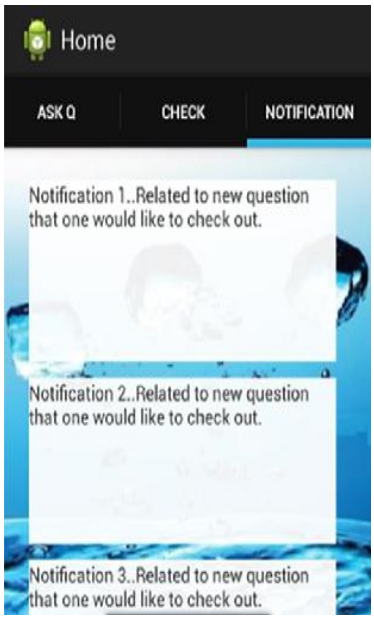







Home
HOME
ASK Q
CHECK
Q. What is the most amazing adventure?
option1
option2
option3
Copied to clipboard
Post





Discussions:-

Advantages:

- Provides better solution to user's questions.
- Generate a well analyzed report that includes pie charts, graphs, plain text report etc.
- Save users time.
- Reduce redundancy

Disadvantages:-

- For now the scope of the application is limited. Scope can be increased.

Applications:-

- An efficient online discussion platform.
- Survey system.
- Gives report on discussions in terms of graphs, text report etc.

Conclusion and future work:-

Thus platform called My Opinion will help to analyze the user comments and generate a report on the bases of analyses done. Opinion mining, sentiment analysis and natural language processing to process the comments to check whether someone commented positive content or negative content. Techniques to perform sentimental analysis and provide our user a better solution. Therefore, this is an open source platform where user can get proper response to the asked question. This is possible by using opinion mining technique under natural language processing. This technique works efficiently to identify the data objects, relationships between them and by analyzing it we can generate a report which will be in form of pie charts, bar graph and so on. Scope is limited to current affairs only level. We are working to develop a system that will take very less time to analyse the comments and then generate report on the basis of analysis done.

In future the scope can be increased to various topics and also voice recognization techniques can also be added in the application.

References:-

1. An empirical study of the naive Bayes classifier by I. Rish T.J. Watson Research Center.
2. An Analysis on Opinion Mining: Techniques and Tools by G.Angulakshmi, ,Dr.R.Manicka Chezian International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.
3. "Opinion Mining and Analysis: A Survey", International Journal on Natural Language Computing (IJNLC) Vol 2 No 3 June 2013Pg No 39-48. By ArtiBuche, Dr.M.B.Chandak, AkshayZadgoanakar.
4. NLTK: The Natural Language Toolkit byEdward Loper and Steven Bird Department of Computer and Information Science University of Pennsylvania, Philadelphia, PA 19104-6389, USA
5. A Peer Review of Feature Based Opinion Mining and Summarization by Padmapani P. Tribhuvan, S.G. Bhirud, Amrapali P. Tribhuvan International Journal of Computer Science and Information Technologies, Vol. 5 (1) ,2014, 247-250.
6. Classification of Opinion Mining Techniques by Nidhi Mishra,C K. Jha International Journal of Computer Applications , Volume 56—No.13,October 2012
7. Opinion Mining and Sentiment Analysis CSI Communication by G Jaganadh 2012
8. Feature-based Opinion Mining and Ranking by Magdalini Eirinaki San Jose State University, S. Pisal , Aruba Networks J. Singh, Google, Inc
9. Sentiment Analysis and Opinion Mining by Bing Liu, University Of Illinois at Chicago , AAAI-2011 Tutorial New book: Bing Liu. Sentiment Analysis and Opinion Mining
10. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text By C.J. Hutto Eric Gilbert Georgia Institute of Technology, Atlanta, GA 30032 cjhutto@gatech.edu gilbert@cc.gatech.edu .
11. Learning Subjective Adjectives from Corpora by Janyce M.Wiebe, Department of Computer Science, New Mexico State University, Las Cruces, NM 88003.
12. Natural Language Processing by Gobinda G. Chowdhury Dept. of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK , gobinda@dis.strath.ac.uk.