

Identification of the Drug Disease Using KNN Algorithm

¹Mrs .S. Nireesh Kumar, ²Mr. M. Imayavaramban, ³Mrs. M. Revathi,

¹Dhanalakshmi Srinivasan College of Engineering and Technology

²Dhanalakshmi Srinivasan College of Engineering and Technology

³Dhanalakshmi Srinivasan College of Engineering and Technology

ABSTRACT: *Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the first time in human history, this now have a huge volume of opinionated data recorded in digital form for analysis.*

I. INTRODUCTION

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features. There are several challenges in Sentiment analysis. The first is an opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". People can be contradictory in their statements.

Sentiment classification is classify documents (e.g., reviews) based on the overall sentiments expressed by opinion holders (authors), Positive, negative, and (possibly) neutral Since in this model an object 'O' itself is also a feature, then sentiment classification essentially determines the opinion expressed on 'O' in each document (e.g., review).

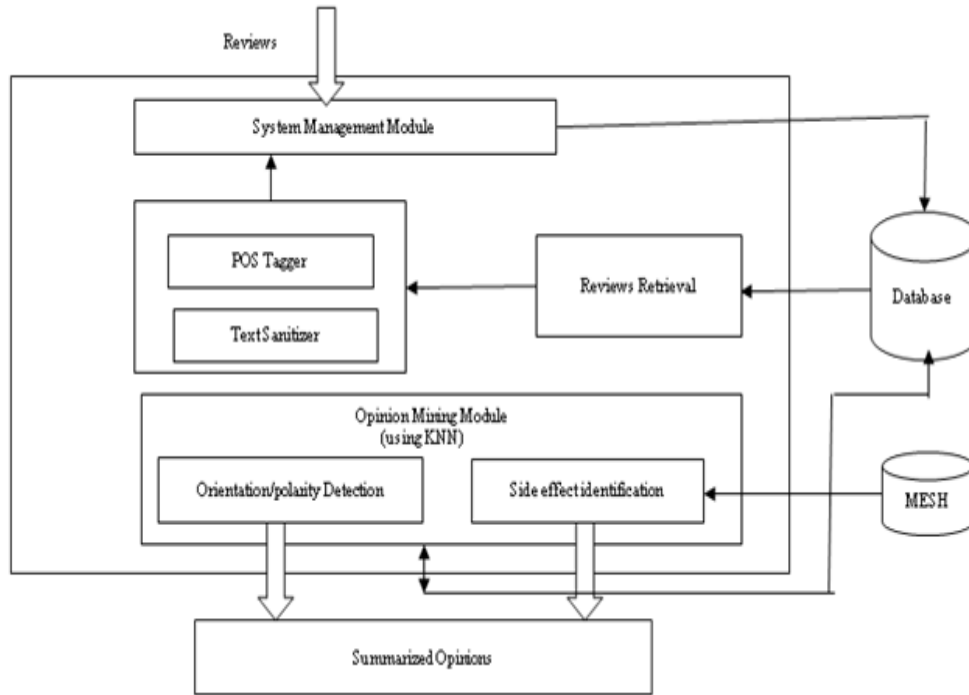
Sentiment classification of textual answers returned within questionnaires could serve other purposes as well. Other applications might include survey coding for the social or political sciences (such as when open questions inquire about the respondents' beliefs, social status, or political leanings) 2 or market research (such as when open questions deal with the respondents' perception of products, brands, or advertising campaigns). It focuses on converting the ontology based data to weighted data .It helps in advanced identification of the inter-social dynamics like ranking.

II. SYSTEM ANALYSIS

PROPOSED SYSTEM

1. It focuses on converting the ontology based data to weighted data.
2. It helps in advanced identification of the inter-social dynamics like ranking.
3. Further emphasizes on inclusion of formal language and medical lexical dictionaries for most prevalent diseases.

III. SYSTEM ARCHITECTURE:

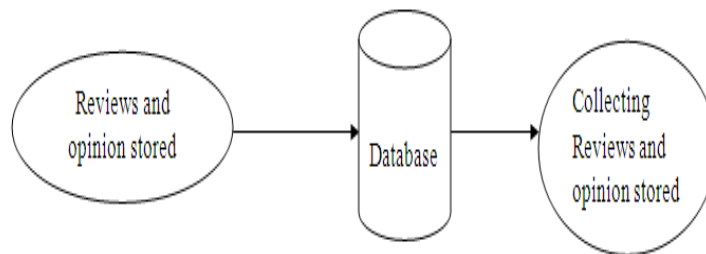


IV. MODULES

- Reviews Retrieval
- Text Sanitizer
- System Management Module
- Opinion Mining Module

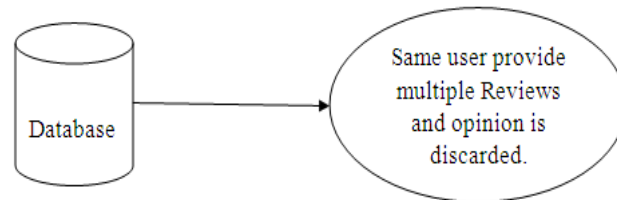
REVIEWS RETRIEVAL

Reviews/Opinions are fetched from the database. The reviews are collected in the database from several discussion forums pertaining to the medical domain. In addition to this, users are also allowed to make new posts. This allows the database to grow dynamically always processing updated entries. These entries include the user details and user opinions.



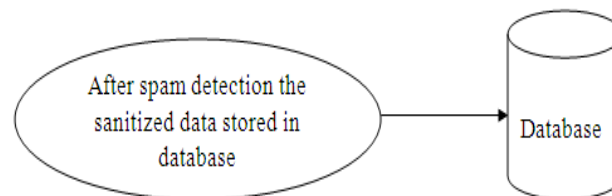
TEXT SANITIZATION

Text Sanitization is performed on the fetched reviews from database. It includes removal of SPAMS from the locally collected data. Spam reviews are identified by means of the following logic: Reviews posted in multiples by the same user are discarded. Many reviewers posting the same comment are also discarded. Irrelevant comments are also omitted.



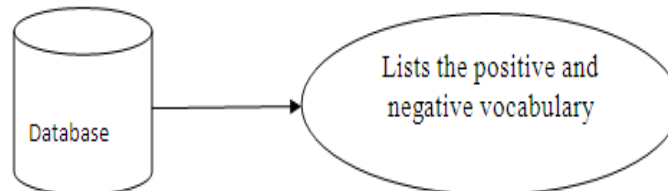
SYSTEM MANAGEMENT MODULE

System Management Module fetches these sanitized data and loads them back to the database. The mining and other process is going to be performed on these sanitized data only. This is an important component because it coordinates and manages the functions of other components of system.



OPINION MINING MODULE

The opinion mining module performs the core functions of the system. The sanitized opinions are fetched back from the database. User polarities are identified processing these entries. Text may be in long textual forms. Sentence level granular analysis is performed. A lexicon is integrated with the system that lists the positive and negative vocabulary using k nearest neighbor algorithm. Words are matched with the collected reviews and user polarities are detected using classification. Side effects listed by the users on drugs are also identified and uploaded in the database.



V. CONCLUSION

It converted a forum focused on ontology into weighted vectors to measure consumer thoughts on the drug Erlotinib using positive and negative terms alongside another list containing the side effects. Those methods were able to investigate positive and negative sentiment on cancer treatment using the drug by classification using the KNN. Most of the user data was clustered to the area of the map linked to positive sentiment, thus reflecting the general positive view of the users. Subsequent network based modeling of the forum yielded interesting insights on the underlying information exchange among users. Modules of strongly interacting users were identified using a multiscale community detection method.

FUTURE ENHANCEMENT

Future solutions will require more advanced detection of inter social dynamics and its effects on the members: such intersets of study may include rankings, ‘likes’ of posts, and friendships. Further emphasis on context posting will require formal language dictionaries that include medical terms for specific diseases, and informal language terms (‘slang’) to clarify posts. Finally, different platforms will allow up-to-date information on the status of the drug in case one social platform ceases to discuss the drug.

REFERENCES

- [1]. Akay A. Dragomir A and Erlandsson B. E. (2015), ‘A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin’,
- [2]. J. Biomed Health Inform, Vol: PP, Issue: 99. Angulakshmi G. and Dr.ManickaChezian R. (2014), ‘An Analysis on Opinion Mining: Techniques and Tools’, Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7.

- [3]. Danushka Bollegala and David Weir. (2013), 'Carroll Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus', IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 8.
- [4]. Das S. R. and Chen M. Y. (2007), 'Yahoo! for Amazon: Sentiment extraction from small talk on the Web', *Manage. Sci.*, vol. 53, pp. 1375–1388.
- [5]. Hsinchun Chen S. (2010), 'AI and Opinion Mining', Part 2. Published by the IEEE Computer Society.
- [6]. Janyce Wiebe and Ellen Riloff C. (2011), 'Finding Mutual Benefit between Subjectivity Analysis and Information Extraction', IEEE Transactions On Affective Computing, Vol. 2, No. 4.
- [7]. Jayashri Khairnar and Mayura Kinikar R. (2013), 'Machine Learning Algorithms for Opinion Mining and Sentiment Classification', International Journal of Scientific and Research Publications, Volume 3, Issue 6, ISSN 2250-3153.